How multiple prosodic boundaries of varying sizes influence syntactic parsing: Behavioral and ERP evidence

Efrat Pauker

School of Communication Sciences and Disorders

McGill University, Montréal, Canada

December 16th, 2013

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

© Efrat Pauker, 2013

Table of Contents

DEDICATION

To my parents, Lily and Yigal Pauker, whose passion for knowledge and education is a constant inspiration to me.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my PhD supervisor and mentor, Dr. Karsten Steinhauer. Karsten, you possess the vast knowledge and expertise of not one, but several accomplished scholars. Your depth of knowledge, research skills, teaching approach, profound methodological and theoretical understanding, meticulousness, and eye for aesthetics (in creating stimuli, experimental design, and even posters) has made me a better student and improved my research a great deal. The high standards you hold for your students as well as for yourself, encouraged me to improve, raise questions, constantly learn and adapt, and still realize that I have much to learn. Your support, optimism and cheerfulness, even after pulling all-nighters, when I was lost in my data, and at other challenging times, have been essential to my motivation and progress. Thank you for your sensitivity, sense of humor, generosity, and kindness.

I owe a very big thank you to Dr. Shari Baum, who has served at all the committees throughout the course of my studies - reading and commenting on my various papers at record time, even when it was not required of her. Thank you for your constant assistance in any matter, big or small, your patience, calming demeanor, sense of humor, kindness, respect, and much more. Thank you for lending an ear whenever I needed someone to talk to, and for always putting things in perspective.

I would like to thank my dissertation committee member, Dr. Michael Wagner, for his important contribution to the material recording and manipulation. Thank you for sharing your resources, for making the time to meet me and answer my frequent technical and theoretical questions, and for discussing the implications of the behavioral results.

I owe a debt of gratitude to a number of people who contributed and assisted to the completion of this project. I would like to thank Dr. John Drury for his help with quantifying the transitivity bias of the verbs, and for all the support and good advice along the way. I would like to thank Rachel Morasse for lending her voice to recording the stimuli over numerous sessions and for her assistance in the ERP data collection. I am extremely grateful for your hard work under the tight time constraints to get the sentences recorded with the intended prosody. I would like to thank Dr. Hyekyung Hwang, for her expert advice in selecting the pitch contour for the original recording and for her technical support. Thank you for making the time to attend some of my recording sessions (even on the weekends), for listening to the material at the various stages of its creation, and for providing valuable input. To Dr. Shari Baum, Dr. Meghan Clayards, Erin Vensel, Jason McDevitt and Eric Zentner - thank you for your "native intuition" in listening to my stimuli and for rating them. Dr. Meghan Clayards, thank you for your help with the Praat scripts, which greatly improved my material. I would like to thank Ajay Kaushik and Gautam Virdi for their help with the data analysis. To Dr. Nicolas Bourguignon, thank you for your technical assistance and for translating my abstract to French. A very big thank you is due to Dr. Tal Krasovsky, for creating the MATLAB scripts that speeded up my raw data analysis. Tsviki Tsadok, thank you so much for making time from your own thesis completion to assist me in various technical issues. To Dr. Inbal Itzhak, I owe you greatly for helping me set up my first experiment under time pressure, and for your constant assistance along the way. To Dr. Katy Carlson, Dr. Chuck Clifton, Dr. Lyn Frazier, Dr. Duane Watson, and Dr. Amy Schafer, who were kind and patient in answering my many questions about their theories. Your work has inspired my own. Finally, I would like to thank all the participants who took part in the studies.

I would like to thank the wonderful administrative staff at Beatty Hall: Karen Cavanagh, Antoinette Sommer, Lili Saran, and Miriam Daye, for helping me resolve any problem with a smile; for the many enjoyable conversations, cookies, and your mere presence. You made my years at the department so much better.

I would like to thank the former and present SCSD GPDs, Dr. Marc Pell, Dr. Linda Polka, for your help and advice, which greatly benefited my progress in the program. Thank you for your kindness and support.

I am extremely grateful to my friends and peers: Dr. Meytal Avgil-Tsadok, Dr. Chinar Dara, Dr. Henry Cheang, Dr. Nino Grillo, Revital Hacmon, Dr. Inbal Itzhak, Dr. Tal Krasovsky, Dr. Julie Mercier, Dr. Andrea Santi, Pnina Strasbourger, Dr. Erin White, and Erin Vensel. Thank you for making Montréal a second home to me, for understanding this arduous process better than anyone else ever could, for helping me push through the hard times, and for making the other times much more enjoyable. To the Israeli gang (and their respective children): Inbal, Tal, Zeev, Meytal, Tsviki, Revital, Hagay, Hadas, Yinon, Meirav, Eran, Shirit and Yair, you have become my extended family. Even when I couldn't come out and play, knowing you were there made a huge difference to me. Thanks for countless meals, hangouts, trips, and much more.

A special thank you is due to Dr. Inbal Itzhak, my friend, adopted big sister, and peer, who started this journey with me and has been a constant source of empathy, knowledge, support, fun, and encouragement. Thank you for always being there for me, rain or shine. You have been my rock in this ocean of trial and tribulations. Words cannot express my deep gratitude and the love I have for you. I am truly lucky for having as my friend. Future research would be required to investigate the physiology and anatomy of our shared lobe.

To my family, Lily, Yigal, Sharon and Nitay, you have been a constant source of love, support, and encouragement. Thank you for enduring the distance, for maintaining your faith in me and helping me restore my own, and for reminding me that when things are difficult it means I’m on my way up. My dearest Mom and Dad, thank you for your wisdom, for cheering me up no matter what, for your endless love, for your sensitivity and patience, and for helping me take things in perspective. To my younger sister, Sharon, who is much wiser than her years – I have learned so much from you. Thank you for all the phone calls, visits, and so much more. You are now on the same journey and I know you will be very successful. To my brother Nitay, thank you for your support, encouragement, and insight.

ABSTRACT

Prosodic boundaries (cued by pitch variations, final lengthening, pause) have been consistently demonstrated to have an immediate influence on parsing in a variety of syntactic structures cross-linguistically. For example, in sentences with temporary ambiguities such as Early and Late closure (EC/LC), which contain two potential boundary positions – the first (#1) compatible with EC and the second (#2) compatible with LC (e.g., *Whenever the bear was approaching* $_{\#1}$ *the people* $_{\#2}$ *(EC): …would run away*; *(LC): …the dogs would run away*), without the benefit of prosodic information, the preferred (or default) interpretation is LC, which consequently leads to processing difficulties (garden-path effects) in EC structures.

The majority of studies on spoken sentence processing has focused on the impact of a single boundary on the closure or attachment preference of a specific phrase or clause. However, more recently, several influential hypotheses have emerged that aim to account for the interplay between two boundaries in a sentence, specifically in terms of size and location; the most influential of these argue that processing is either (i) *local*, with large boundaries independently integrated, which serve as strategic cues to syntactic closure (Watson & Gibson, 2005), or (ii) *global*, with the relative difference between the magnitude of boundaries across an utterance modulating interpretation (Clifton, Carlson, & Frazier, 2002). Although differing in details, these hypotheses suggest that listeners process boundary information at the sentence level in a *categorical manner*. In contrast, there is some data to suggest that boundaries can differ in a *gradient quantitative manner*, and that listeners are sensitive to this range of boundary sizes.

The aims of the current dissertation were therefore to use behavioral and event-related potential (ERP) measures: (i) to contrast the predictions of the opposing theoretical accounts using temporary syntactic ambiguities, and (ii) to test whether gradient differences in boundary size impact listeners' parsing decisions in a gradient or categorical manner.

Using an innovative paradigm, I conducted two behavioral experiments (Study 1), and one ERP experiment (Study 2), where listeners were presented with highly controlled digitally-manipulated EC/LC sentences, each containing two prosodic boundaries (as in the example above), which differed only in terms of their relative sizes. The outcomes of the three experiments reveal an initial, profound bias of boundaries on syntactic preference, which was nearly impossible to override. In addition, subtle differences between prosodic boundaries are detected by the brain and affect the degree of processing difficulty. Finally, the effect of boundaries on parsing is far more intricate than previously assumed. These outcomes cannot be accommodated by a purely categorical account, and cast serious doubts on most current models of prosodic online processing. We present the extended Boundary Deletion Hypothesis (eBDH), an alternative account for prosodic phrasing, based on the results of all three experiments.

RÉSUMÉ

L'influence des frontières prosodiques (caractérisées par des variations d'intensité, d'allongement final et des pauses) sur l'analyse de structures syntaxiques a été démontrée de façon systématique dans plusieurs langues. Un exemple se retrouve dans les phrases à ambiguïtés temporaires, lesquelles contiennent deux positions potentielles : la première (#1) étant compatible avec une clôture précoce (CP), la deuxième (#2) avec une clôture tardive (par exemple, *Aussitôt que l'ours s'approcha* $_{\#1}$ *les gens* $_{\#2}$ *(CP) :… se sauvèrent ; (CT) : … les chiens se sauvèrent*). En l'absence d'information prosodique, l'interprétation privilégiée (ou par défaut) correspond à une structure en CT, ce qui a pour effet de provoquer des difficultés de traitement (effets « cul de sac ») des structures CP.

Une majorité d'études de phrases parlées se sont concentrées sur l'impact d'une seule frontière sur les préférences de clôture ou de jonction d'un syntagme ou d'une proposition spécifique. Toutefois, plusieurs théories importantes ont récemment vu le jour et tentent d'expliquer l'interaction s'établissant entre deux frontières dans une même phrase, prenant leur taille et leur position comme principaux facteurs. Selon les approches les plus importantes sur la question, le traitement est soit (i) *local*, les frontières plus importantes étant intégrées indépendamment et servant d'indices stratégiques à la clôture syntaxique (Watson & Gibson, 2005), soit (ii) *global*, la différence relative entre la magnitude des frontières à travers la phrase modulant l'interprétation (Clifton, Carlson & Frazier, 2002). Quoique différentes dans le détail, ces théories suggèrent que les personnes traitent l'information liée à la frontière au niveau de la phrase de *manière catégorielle*. Cependant, plusieurs données suggèrent que les frontières peuvent différer *de manière graduelle et quantitative*, et que les personnes sont sensibles à la taille des frontières.

Les buts de la présente thèse étaient dès lors d'utiliser des mesures comportementales et électrophysiologiques (Potentiels évoqués ou PÉs) afin de (i) contraster les prédictions propres à chacune de ces théories au moyens d'ambiguïtés syntaxiques temporaires, et (ii) de tester si les différences graduelles de tailles des frontières ont un impact graduel ou catégoriel sur les décisions d'analyse des participants.

Nous avons conduit deux expériences comportementales (Étude 1) et une expérience PÉs (Étude 2) au moyen d'un paradigme novateur, au cours duquel des phrases CP/CT manipulées numériquement et rigoureusement contrôlées étaient présentées aux participants. Ces phrases contenaient deux frontières prosodiques (comme dans l'exemple donné plus haut) différant exclusivement au niveau de leurs tailles relatives. Les trois expériences révèlent que les frontières induisent un biais initial, profond et impossible à surpasser sur les préférences syntaxiques. En outre, des différences subtiles entre les frontières prosodiques sont détectées par le cerveau et affecte le degré de difficulté du traitement. Enfin, l'effet des frontières prosodiques sur l'analyse syntaxique est bien plus sophistiqué qu'assumé précédemment. Ces résultats échappent à une hypothèse strictement catégorielle et mettent sérieusement en doute la plupart des modèles actuels de traitement prosodique en temps réel. Sur la base des présentes données, nous présentons l'Hypothèse étendue de Délétion de Frontières (HeDF), offrant une explication alternative du phrasé prosodique.

FUNDING

The studies presented in this thesis were funded by grants awarded to Dr. Karsten Steinhauer by the Canada Research Chair program and the Canada Foundation for Innovation (CRC/CFI; Project # 201876) and the Canadian Institutes of Health Research (CIHR; # MOP-74575), and by a McGill University Faculty of Medicine Internal Studentship awarded to Efrat Pauker.

STATEMENT OF ORIGINALITY

The studies presented in this thesis investigated the influence of multiple prosodic boundaries of various magnitudes on the interpretation of temporarily ambiguous sentences in English. Study 1, comprised of two behavioral experiments, demonstrates influence of two prosodic boundaries of 3 and 4 different sizes (respectively) on comprehension. Study 2 is the first ERP experiment to explore the online effect of multiple prosodic boundaries on processing. These studies employ an innovative paradigm and present novel findings regarding the dynamic contribution of prosodic boundaries to language comprehension as well as shed new light on the organization of phonological units in the brain. At the time of the official submission of the dissertation to McGill University, previous analyses of the data in Study 1 (Chapter 2) have been presented in posters at the CUNY Conference on Human Sentence Processing in March 2011, at the Experimental and Theoretical Advances in Prosody (ETAP) conference in September 2011, and at the Society for the Neurobiology of Language (SNL) conference in November 2013. A previous analysis of the data from Study 2 (Chapter 3) has been presented in a poster at the Society for the Neurobiology of Language (SNL) conference in November 2013.

CONTRIBUTION OF AUTHORS

The current dissertation is manuscript-based, containing two behavioral studies and one ERP study. The entire research project was conceptualized, designed, carried out, analyzed, and interpreted by the first author (Efrat Pauker), with continuous assistance from her PhD supervisor, Dr. Karsten Steinhauer (the third author in Study 1 and second author in Study 2). Dr. Michael Wagner (the second author in Study 1) contributed to the design and stimuli creation, and provided assistance with regard to the theoretical background and data interpretation of Study 1. The first author was also fully responsible for all manuscripts writing while receiving editorial suggestions and comments from the second and third authors.

CHAPTER 1: General Introduction

Language comprehension is an extremely complex task, which requires rapid processing of the incoming input and timely integration of various types of information to activate the relevant linguistic representations. For centuries, researchers have been intrigued by the ease and effortless manner by which listeners carry out this task, while facing inconsistent, distorted, incomplete, or ambiguous information. Earlier models of linguistic processing have deconstructed language into its components, assuming certain aspects (e.g., syntax) may be more significant to communication than others (e.g., prosody). However, in the past few decades, a number of studies have revealed that factors that were initially neglected may be key to our understanding of real-time language processing. The present investigations aim to contribute to the research on prosody-syntax mapping, and in particular to the existing debate on the manner in which prosodic phrasing guides syntactic parsing. Prosodic phrasing has generated much interest in recent decades because it is at the foundation of language processing. From the earliest stages of language perception, to the most intricate of constructions, phrasing is used to make sense of the conveyed message by dividing the speech stream into meaningful units, helping to reduce the load on working memory capacity. To this end, three studies were designed, using a novel methodology, to examine the effect of multiple prosodic boundaries on the comprehension of ambiguous syntactic structures. Another important goal was to test the predominant view of prosodic boundary representation, which has been crucial to conceptualizing and implementing boundary types in research.

One important aspect of the prosody-syntax relationship is that prosodic and syntactic boundaries often coincide. This allows listeners to detect syntactic breaks via prosodic information, and provides a tool for researchers to investigate this relationship by violating or manipulating this internal knowledge. However, the organization of prosodic and syntactic structures is not isomorphic (Shattuck-Hufnagel & Turk, 1996). That is, while researchers have theorized an internal structure for each domain, consisting of hierarchically organized units, in spoken language, these constituents do not always overlap. Instead, syntactic constituents can be prosodically grouped in a variety of ways, depending on multiple linguistic and other factors (e.g., constituent length, speech rate). As a result, it is not always possible to predict the syntactic structure from prosodic phrasing. However, many studies have shown that major prosodic and syntactic boundaries often coincide, thus opening a number of avenues for gaining a better understanding of the prosody-syntax mapping. One of the most effective approaches to achieve this goal has been the study of (temporarily or globally) ambiguous structures.

Although the study of prosodic phrasing has generated an extensive body of research, and despite the many advances in this field over the past years, many issues are controversial. One example concerns the status of boundaries varying in size. The prevalent view in phonological theory is that prosodic phrases constitute the top levels of a hierarchical prosodic grammar and that prosodic boundaries are perceived and processed in a categorical manner that largely ignores acoustic differences within each category. This framework is also reflected in the most widely used annotation system for speech prosody (Tones and Break Indices, or ToBI). More recently, however, researchers have criticized both a number of inconsistencies of the annotation system as well as the notion of categorical boundary perception more generally, and have instead proposed a gradient view of boundary processing. A second debate is about how multiple conflicting boundaries of an utterance are integrated in real-time, locally when they occur or globally when all information is available. A last issue concerns the neurocognitive mechanisms underlying prosodic phrasing. Event-related brain potentials (ERPs) offer a tool to investigate all these issues at once.

This dissertation presents two behavioral experiments and one ERP study that have been designed to test the predictions of various current hypotheses of prosodic processing with state-of-the art techniques and to advance our understanding of the prosody-syntax interface. The ERP study also aimed to shed light on some characteristics of a brain response in ERPs that is viewed as a direct reflection of prosodic phrasing. The following sections will provide a general overview of the relevant hypotheses and research techniques in order to develop the specific hypotheses for the three studies.

1.1. Prosodic phrasing and syntactic ambiguity

Historically, the study of language comprehension has focused on elements thought to be properly in the linguistic domain such as semantic and syntactic structure, in isolation from the acoustic-phonetic details of the speech that carried that structure to the listener's ear. While in the past the acoustic substrate of language input was felt to consist of "random noise", the details of which could be discarded after abstraction of linguistic meaning, it is now recognized that these acoustic details are fundamental to language comprehension. In particular, the acoustic correlates of prosody are at the core of any spoken utterance and the syntactic structure of any spoken sentence is necessarily defined by prosodic variations. Thus, any spoken sentence is necessarily defined by prosodic variations (Cutler, Dahan, & Van Donselaar, 1997). One of the best-studied aspects of prosody involves the way utterances are divided into meaningful processing units, which may have either a favorable or an adverse impact on comprehension. This mechanism is referred to as prosodic phrasing, whereby characteristic acoustic cues (e.g., duration, amplitude, fundamental frequency; see Cutler et al., 1997, for review) contribute to the chunking of utterances. For example, the sentence: '*You will be required to work twenty four hour shifts'* bears two different meanings depending on whether the words *twenty* and *four* are grouped together ([*You will be required to work* ***twenty four*** *hour shifts*]) or separately ([*You will be required to work* **twenty] [four** *hour shifts*]). In speech, utterances are segmented by prosodic boundaries, which have been described as perceptual breaks in the input, characterized with an array of acoustic features (e.g., final syllable lengthening, pitch variations, silence interval, pitch reset), whose configuration depends on the degree of separation between syntactic constituents (Cutler et al., 1997; Lehiste, 1973; Streeter, 1978; Swerts, 1997; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). For example, many studies have shown that boundary size is highly correlated with the acoustic dimensions of duration and pitch - that is, the amount of lengthening occurring on the pre-boundary word (hence, pre-boundary lengthening; Klatt, 1975; Lehiste, 1973; Shattuck-Hufnagel & Turk, 1996; Wightman et al., 1992; see Wagner & Watson, 2010, for review), and intonation variations (rising and falling Fø) on the pre-boundary word (Beach, 1991; Streeter, 1978; Wightman et al., 1992; see Cutler et al., 1997, for review).

Many studies have demonstrated that speakers produce a variety of patterns of prosodic phrasing to signal syntactic structure, which can in turn be used by listeners for structural disambiguation (Beach, 1991; Carroll & Slowiaczek, 1987; Clifton, Carlson, & Frazier, 2006; Kjelgaard & Speer, 1999; Kraljic & Brennan, 2005; Lehiste, 1973; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Pynte & Prieur, 1996; Schafer, Speer, Warren, & White, 2000, 2005; Snedeker & Trueswell, 2003; Warren, 1985; Warren, Grabe, & Nolan, 1995; see Cutler et al., 1997, for review). While certain aspects of prosodic phrasing are constrained by the morphosyntactic structure of the utterance, many are optional, including the location, size and number of bracketings, making it difficult (or even impossible at times) to predict the syntax from the prosody, and vice versa (Shattuck-Hufnagel & Turk, 1996). Several theories have proposed that while prosodic phrasing may not necessarily be produced with the listener in mind, speakers do produce certain prosodic cues in a premeditated manner, rather than arbitrarily. Thus, even though prosody is not uniform across speakers (or even across utterances of the same speaker), listeners are able employ their internal knowledge as speakers to associate certain prosodic productions with certain syntactic structures (Carlson, Clifton, & Frazier, 2001; Clifton, Carlson, & Frazier, 2002; Frazier, Clifton, & Carlson, 2004; Watson & Gibson, 2005). Another hypothesis has suggested that prosodic phrasing is so salient in certain structures that listeners can be easily misguided by an incorrect grouping of words, increasing the cost of processing difficulties, while the absence of any phrasing appears less challenging, because inserting a missing prosodic boundary may be easier than deleting a superfluous one (Steinhauer & Friederici, 2001).

1.2. The hierarchical structure of prosodic structure

In the literature, several categories of prosodic phrasing are assumed, which correspond to different levels of prosodic units. This assumption is rooted in theories of metrical phonology (Liberman, 1975; Liberman & Prince, 1977) and prosodic phonology, in particular the Strict Layer/Prosodic Hierarchy Hypothesis (Nespor & Vogel, 1986; Selkirk, 1980). Liberman and Prince (1977) proposed that, like music, language is organized into a rhythmic structure they termed a *metrical grid*. Each utterance is prosodically defined as a pattern of alternating strong (stressed) and weak (unstressed) syllables, which can be represented on a binary-branching metrical tree (Selkirk, 1986). As prominence relations beyond the word level were not initially defined, Selkirk (1986), and later Nespor and Vogel (1986) proposed that English utterances have a prosodic structure comprised of various constituents of different prosodic properties, which are hierarchically organized (e.g., syllable, foot, prosodic word, phonological phrase, intonation phrase, utterance)[1]. Each utterance is thus exhaustively phrased into smaller constituents, as in the example below (Selkirk, 1986; p. 384):

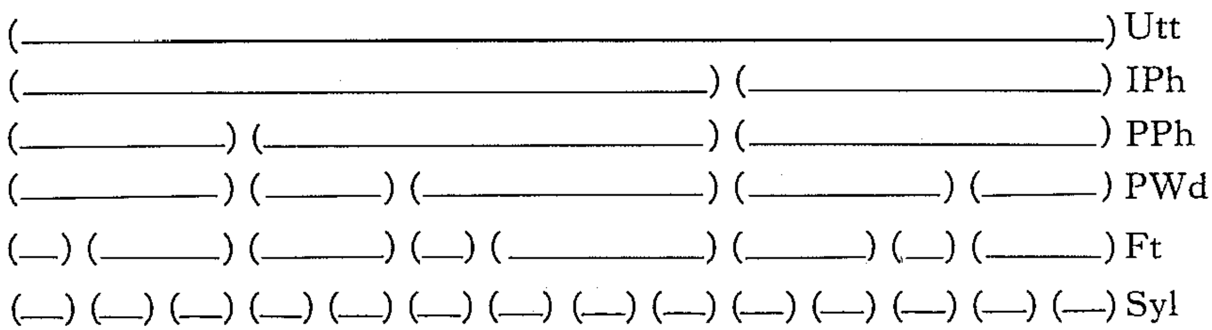


Figure 1. Phonological hierarchies model

Two prosodic constituents were posited at the phrase level: the larger intonational phrase (or IPh) and the smaller phonological phrase (or PPh), also referred to as intermediate phrase (or ip) in later accounts (here we will use the terms PPh and ip interchangeably). According to this view, the boundaries of these constituents (or their right edges) also conform to the prosodic hierarchy, so an IPh boundary is perceived as larger than a PPh boundary in the same utterance (Rossi, 1997).

Another theory of sentential prosody is an intonation theory (later dubbed the 'autosegmental-metrical theory' by R. D. Ladd, 1986) proposed by Pierrehumbert (1980), which represents the intonation of a sentence as a sequence of tonal events. These tonal events are *pitch*

[1] Other units have since been introduced (see Ladd, 1996 for review).

accents, marking stressed syllables, and *boundary (or edge) tones*, which are aligned with the boundaries (or right edges) of prosodic phrases (see R. D. Ladd, 2008 for review).

These two theories have been integrated to create the Tones and Break Indices (ToBI) transcription system of American English (Beckman & Elam, 1997; Beckman & Hirschberg, 1994; Beckman & Pierrehumbert, 1986; Shattuck-Hufnagel & Turk, 1996; Silverman et al., 1992), which describes the categorical distinctions between intonational and phonological phrases in greater detail. Although other annotation systems exist in various languages, ToBI is considered a standard for American English and is now one of the most popular and widely used prosodic annotation systems. As noted, ToBI draws clear distinctions between the IPh and ip categories, marking each with characteristic tonal events (*tones*) and a specific degree of disjuncture between phrases (*break indices*). Briefly, in each ip, prominent (stressed) words are marked with a *pitch accent* (*), which can be either high (H*) – representing pitch rise, low (L*) – representing pitch fall, or complex pitch (e.g., L*+H). The right edge of an ip is characterized by a high (H-) or low (L-) *phrase accent*. As each IPh contains at least one ip, it bears the same tonal pattern, but differs in that its right edge is characterized with an additional high (H%) or low (L%) *boundary tone*, located on the final word or syllable (depending on the constituents' length) of the phrase. The break indices are scaled from 0-4, where 4 is the strongest break, associated with IPh boundaries, and 3 is associated with ip boundaries (Beckman, Hirschberg, & Shattuck-Hufnagel, 2005; Pitrelli, Beckman, & Hirschberg, 1994; Shattuck-Hufnagel & Turk, 1996). Note that the transcribed tones and breaks are relative in each phrase and do not represent an absolute measure. An important distinction between the two boundary types is mainly the extent of durational changes associated with them. IPh boundaries are considered more prominent relative to ip boundaries, and are usually characterized with longer pre-final lengthening and pause durations (see Shattuck-Hufnagel & Turk, 1996 for review).

1.3. The effect of a single prosodic boundary

Prosodic phrasing has been of particular interest in the context of syntactic ambiguity resolution. Some of the main issues concern the type of structures that can be reliably disambiguated by phrasing, the time-course of prosodic and syntactic integration, whether prosodic phrasing can be predicted from the sentence's structure and vice versa (prosody-syntax mapping), and whether the size of the boundary affects comprehension (see Carlson, 2009, for review).

Early studies by Lehiste (1973) and Lehiste, Olive, and Streeter (1976), for example, found that global (or "standing") ambiguities, which are not resolved by lexical information downstream (Beach, 1991; See Cutler et al., 1997, for review), were successfully disambiguated by prosodic phrasing if they contained attachment ambiguities (see examples 2a, 2b; Lehiste, 1973), but not if they contained syntactic (word) category ambiguities (see example 1; Lehiste, 1973). That is, sentences with one or more surface bracketings have been found to be affected by prosody, whereas sentences with only one surface bracketings have not (see Schafer, 1997, for review). Similarly, prosodic grouping of local (or "temporary") ambiguities (see example 3a, 3b; Warren, 1985), which are resolved by lexical information downstream (Beach, 1991; See Cutler et al., 1997), differ in their surface bracketings by definition, and have been shown to be reliably disambiguated by prosodic phrasing.

1) [Visiting relatives] can be a nuisance
2a) [Steve or Sam] [and Bob] will come
2b) [Steve] [or Sam and Bob] will come
3a) [Before the king rides] [his horse takes ages to groom]
3b) [Before the king rides his horse] [Ted gives it a groom]

Initially, prosody was thought to only support, rather than direct syntactic parsing, mostly based on studies using stimulus disruption techniques (e.g., click localization, dichotic switch), where participants indicated the location of a presented disrupting stimulus within the sentence. Since participants often reported the interruption occurred at major syntactic boundary positions, researchers assumed syntactic structure determines sentence segmentation and can override prosodic information (see Cutler et al., 1997, for review). However, these techniques have been criticized for being heavily reliant on memory, rather than on online processing. Later techniques, such as the cross-modal naming paradigm (Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992) and sentence truncating (Beach, 1991), in which listeners hear the initial (short or long) parts of ambiguous sentences (e.g., [*Mary suspected*][*her boyfriend …*] / [*Mary suspected her boyfriend…*] … [*was lying*]/[*immediately*]) and chose the appropriate sentence completion, were thus designed to tap into the time-course of prosodic and syntactic integration. Using these paradigms, researchers have found that prosody can disambiguate sentences even before the disambiguating region is encountered, essentially directing the early stages of syntactic parsing.

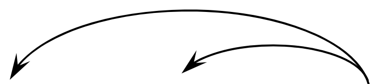
Since then, many studies have confirmed that prosodic boundaries can resolve syntactic ambiguities in different structures, including closure (Kjelgaard & Speer, 1999; Nagel, Shapiro, Tuller, & Nawy, 1996; Schafer et al., 2000; Speer, Kjelgaard, & Dobroth, 1996; Speer, Warren, & Schafer, 2003; Walker, Fongemie, & Daigle, 2001; Warren et al., 1995), and attachment (e.g., prepositional phrase) ambiguities (Grosjean, 1983; Pynte, 2006; Pynte & Prieur, 1996; Snedeker

& Trueswell, 2003). Although the majority of studies has focused primarily on the facilitating or interfering effect of large (IPh) boundaries on comprehension, more recently a growing amount of research has shown comparable effects using small (ip) boundaries, both in closure and attachment ambiguities (Carlson et al., 2001; Clifton et al., 2002; Hwang & Schafer, 2006; Kjelgaard & Speer, 1999; Millotte, René, Wales, & Christophe, 2008; Millotte, Wales, & Christophe, 2007; Schafer, 1997; Snedeker & Casserly, 2010). Nevertheless, the effect of ip boundaries on comprehension remains controversial, as some theorists argue only IPh boundaries can guide the parser towards a specific interpretation (Blodgett, 2004; Marcus & Hindle, 1990; Price et al., 1991; Watson & Gibson, 2004b, 2005).

1.4. Local vs. global approaches of prosodic processing

Thus far, research on prosodic phrasing at the sentence level has mainly focused on the presence vs. the absence of a single (mostly large) prosodic boundary. More recently, researchers have begun to investigate the effect an additional boundary would have on syntactic ambiguity processing, as well as the interplay between boundaries of various strengths. Two opposing positions have emerged in this line of research. The first approach suggests that prosodic boundaries are treated **locally**, as cues to syntactic closure, and are thus used strategically to avoid ambiguities. The second approach suggests the parser performs a more **global** analysis to determine the syntactic preference, by considering the relative strength of the boundaries in a sentence. We discuss these approaches in the following paragraphs.

A number of hypotheses favoring either local or global views have been put forward, whose distinct predictions have typically been tested using structurally ambiguous sentences containing two competing boundary positions, each of which supports a different sentence interpretation. For example, the sentence final prepositional phrase in (4) can either attach high to the verb (*attacked*) or low to the second NP (*the paper*) – making the ambiguous phrase *with flawed data* either the instrument with which Amanda approached the paper, or the modifier (or adjective) describing the paper.



(4) Amanda attacked $_{\#1}$ the paper $_{\#2}$ <u>with flawed data</u>

The distinction between local and global approaches developed gradually through a number of stages. Originally, research on prosodic phrasing focused on the effect that a single prosodic boundary at a point of syntactic ambiguity (position #1 or #2) may have on interpreting such sentences (see section 1.3 above). This approach treats boundaries as having an immediate and local effect on parsing (Marcus & Hindle, 1990; for review see Cutler et al., 1997): when encountered at a certain position in the sentence, a boundary signals to the parser that the current constituent is complete; as a result, it creates preference for a specific syntactic structure. Researchers have often focused on the effect of a boundary directly preceding the ambiguous phrase, whose presence or absence would often change the attachment preference (for review see Snedeker & Casserly, 2010).

More recently, theorists have claimed that the fact that these previous studies have consistently employed only one boundary per ambiguous sentence could be the reason that the effect of a prosodic boundary on parsing has been considered to be strictly local (see Frazier, Carlson, & Clifton, 2006). To test this claim, researchers examined the interaction between two boundaries of different sizes in ambiguous sentences such as (4). Their findings suggest that when boundary #1 is smaller than boundary #2, listeners show a higher rate of high attachment preference, but when boundary #1 is larger than boundary #2, listeners show a lower rate of high attachment preference (for review, see Snedeker & Casserly, 2010). These investigations led to a new account, suggesting that the global prosodic structure of the sentence – manifested by the distribution and relative magnitude among all boundaries within a given utterance – is pivotal to the prosody-syntax mapping (Carlson et al., 2001; Clifton et al., 2002; Schafer, 1997). That is, the strength of the boundary at position #1 relative to the strength of the boundary at position #2 reflects a specific prosodic contour associated with a specific syntactic interpretation.

In response to this proposal, proponents of the local approach have claimed that the introduction of multiple boundaries of different sizes into the sentence should not alter the parsing preferences found in earlier studies (with one IPh boundary). Based on the same findings described by the global hypotheses (weak - strong = more high attachment; strong - weak = less high attachment), it has been suggested that prosody-syntax mapping is performed locally, immediately upon encountering a strong prosodic boundary (Pynte & Prieur, 1996; Watson & Gibson, 2004a; for review see Snedeker & Casserly, 2010; 2005).

In the present study, we examine the predictions of the two most influential hypotheses from these opposing camps, and will now discuss them in the order in which they originated: (1) the *Informative Boundary Hypothesis (IBH)*, and (2) the *Anti-Attachment Hypothesis (AAH)*.

1.5. Theoretical accounts for the prosody-syntax interface

1.5.1. The Informative Boundary Hypothesis (IBH)

1.5.1.1. Theoretical background of the IBH

The IBH is based to a large extent on the hypotheses laid out in Schafer's (1997) dissertation. As this work is highly relevant to the IBH, we will first present it in some detail and then continue to discuss the IBH itself.

While early accounts of the effect of prosodic phrasing on syntactic ambiguity resolution (Price et al., 1991; Schafer et al., 2000; Speer et al., 1996) described the placement of prosodic boundaries as aligned to the right edges of major syntactic constituents, i.e., from the perspective of the speaker (see Schafer, 1997 for review), Schafer (1997) developed the Prosodic Visibility Hypothesis (PVH), that views the effect of prosodic phrasing on parsing as domain effects, i.e., from the perspective of the listener. The difference between these approaches is that the former views prosodic phrasing as local cues reflecting syntactic constraints, while the latter views syntactic parsing as driven by prosodic phrasing. Schafer argues that attaching an ambiguous constituent to a node within the same prosodic domain (or phrase) is more "visible" to the parser, and, therefore, favored over attachment across prosodic domains (delimited by prosodic boundaries). The three main principles of this hypothesis are (p.42):

(a) The phonological phrasing of an utterance determines the *visibility* of syntactic nodes.
(b) Nodes within the phonological phrase currently being processed are more visible than nodes outside of that phonological phrase; visibility is gradient across multiple phonological phrases.

(c) In the first analysis and reanalysis, attachment to a node with high visibility is less costly in terms of processing or attentional resources than attachment to a node with low visibility.

The PVH views the input to the parser not as mere word strings, but rather as prosodically packaged lexical expressions. These expressions are thus phrased into "prosodically-defined domains of material" (p.43), which determine whether a potential attachment site for a given node is more or less visible to the parser. As the PVH views attachment decisions as driven by the processing of syntactic content within these phonological domains, it also takes into account the global pattern of prosodic phrasing within the utterance (since boundary placement across the utterance has a crucial effect on parsing). Importantly, Schafer uses only ip boundaries to support her claims, arguing that they represent the level of internal packaging within an utterance.

Schafer argues that the PVH's predictions can be illustrated using previously studied materials, such as the closure ambiguities which were thus far studied using a single boundary (Speer et al., 1996; and should also apply to similar, later accounts such as Kjelgaard & Speer, 1999;), as in example (5) below. Schafer describes the visibility of the ambiguous NP *door* when the disambiguating region (*is/it's*) is encountered:

(5) a. (Whenever the guard checks the door)$_{ip}$ (it's locked)	(cooperating prosody)
b. (Whenever the guard checks)$_{ip}$ (the door is locked)	(cooperating prosody)
c. (Whenever the guard checks the door it's/is locked)	(baseline prosody)
*d. (whenever the guard checks)$_{ip}$ (the door it's locked)	(conflicting prosody)
*e. (whenever the guard checks the door)$_{ip}$ (is locked)	(conflicting prosody)

Note: Prosodic visibility is indicated by font size (larger visibility = larger font, and vice versa).

For example, in the cooperating prosody conditions (5a) and (5b), the ambiguous NP (*door*) is highly visible to the VP in the same phonological phrase (in (a) – the subordinate verb *checks*, and in (b) – the matrix verb *is*), and is less visible to the VP which is separated by a prosodic boundary. These prosodic patterns correctly lead to the assignment of a late closure structure in (a) and an early closure structure in (b). In the baseline condition (4c), the ambiguous NP is highly visible to both VPs, in which case a late closure should be the favored attachment preference, resulting in some processing difficulties when early closure is the intended structure. In conflicting prosody conditions (5d) and (5e), the low visibility of the correct VPs to the ambiguous NP results in the initial assignment of an incorrect closure structure, creating a syntactic violation, and an increased processing load as a result.

Schafer (1997) tested whether the PVH can also be applied to other types of attachment ambiguities, such as the PP attachment illustrated in (6). According to previous psycholinguistic accounts, without the benefit of prosodic information, the preferred (or default) attachment for such structures is high attachment of the PP to the verb (Frazier, 1987; Rayner, Carlson, & Frazier, 1983). However, placing boundaries after the VP or after the NP should affect the visibility, and, therefore, also the availability of certain attachment sites to the ambiguous constituent:

(6) a. ((The bus driver angered the rider)$_{ip}$ (with a mean look)$_{ip}$)$_{IPh}$ (higher % VP attachments)
b. ((The bus driver angered) $_{ip}$ (the rider with a mean look)$_{ip}$)$_{IPh}$ (higher % NP attachments)
c. ((The bus driver angered the rider with a mean look)$_{ip}$)$_{IPh}$ (higher % VP attachments)
d. ((The bus driver) $_{ip}$ (angered) $_{ip}$ (the rider)$_{ip}$ (with a mean look)$_{ip}$)$_{IPh}$ (intermediate % VP attachments)

Notes: (1) PVH predictions for each condition are indicated in brackets.
(2) Prosodic visibility is indicated by font size (larger visibility = larger font, and vice versa).

In line with the PVH's predictions, Schafer found graded attachment responses to her material, based on the "prosodic distance" of the PP from its potential attachment sites. When both attachment sites were equally available to the PP ((6a) and (6c)), most responses favored the default high attachment (61.5% and 59.9%, respectively); when the NP was most visible to the PP in (6b), significantly fewer responses favored high attachment (44.3%); and when the NP was less visible to the PP, but more visible than the VP (that was separated by more boundaries compared to the NP), as in (6d), an intermediate number of high attachment responses was found (52.6%).

Schafer does not discuss how a specific location or size of a prosodic boundary affects ambiguity resolution (other than the distance between an ambiguous constituent and a potential attachment site), since boundaries are perceived as a tool for distinguishing between prosodically-defined domains of lexical material and not as distinct entities. Nevertheless, she does suggest another hypothesis – the Interpretive Domain Hypothesis (IDH) – that argues that the phonological differences between ip and IPh boundaries suggest they may also have distinct effects on sentence processing. According to this hypothesis, any semantic and pragmatic integration that has not already taken place during processing will be performed when an IPh boundary is encountered. An IPh, therefore, serves to "wrap up" the current domain of the currently constructed phrase (the IDH does not apply to intermediate phrases). However, the predictions of the hypothesis are rather vague, since it was tested on very subtle semantically ambiguous adjectives, when an IPh intervened between a context-sensitive intersective adjective[2] and a head noun, rather than on traditional syntactic ambiguities. The results of this study

[2] Adjectives whose interpretation "results from the intersection of the meaning of the adjective and the meaning of the noun it modifies"; (Schafer, 1997. p.88). For example, *tall basketball player* can mean (i) a person who is relatively tall for a basketball player, or (ii) someone who is both tall and a basketball player.

(acceptability judgment task) – claimed to support the hypothesis – show differences in RTs between sentences containing an IPh vs. an ip between the adjective and its respective noun (supposedly reflecting processing difficulty), but can also be explained by the delay of material arrival, as IPh boundaries increase the sentence's length.

The PVH seems to have some potential weaknesses. First, in any structural ambiguity containing two boundaries, where the ambiguous region is located at sentence final position, a constituent appearing earlier in the sentence will always be less visible than a later-appearing constituent; this would result in higher visibility of the constituent closest to the ambiguous phrase (or the default preference in case both attachment sites are equidistant) when any two boundaries are involved. Second, it might be the case that IPh boundaries have an advantage over ip boundaries in terms of syntactic closure, but the hypothesis does not provide specific predictions in that regard. Third, the hypothesis has only received weak support from a behavioral task (acceptability judgment), where RTs were not controlled for, raising the possibility that other linguistic processes may have influenced parsing by the time subjects made their choice.

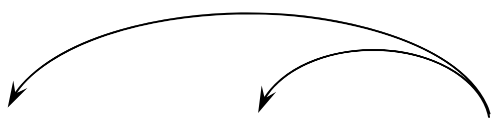
The importance of the PVH is that it was the first hypothesis to directly discuss the effect of multiple prosodic boundaries on ambiguity resolution. Its main weakness is that it exclusively focuses on the effect of intermediate boundaries. While another account considering the role of IPh boundaries was also outlined by Schafer, the two hypotheses were never merged into a single, more comprehensive framework accounting for both size and number of prosodic boundaries in a sentence. The Informative Boundary Hypothesis is the first account to reconcile these two factors.

1.5.1.2. The IBH – main principles and empirical evidence

The IBH (Carlson et al., 2001; Clifton et al., 2002) adopts Schafer's (1997) proposal that the prosodic phrasing throughout the sentence matters (as opposed to local prosodic boundaries) and integrates it with both the findings that speakers produce a variety of prosodic contours for a given structure (Schafer et al., 2000), and the existence of syntactic (and other) constraints on prosodic phrasing (Selkirk, 1984; Watson, Breen, & Gibson, 2006; Watson & Gibson, 2004b). As previously mentioned, many studies on spoken sentence processing have focused on the local impact of a single boundary on the closure of a specific phrase or clause. However, other studies have shown that a given (unambiguous) sentence can be produced with a variety of prosodic structures, including variability in such factors as the presence or absence of boundaries and their size (Kraljic & Brennan, 2005; Schafer et al., 2000; Snedeker & Trueswell, 2003; see also Frazier et al., 2006 for review). For example, Schafer et al. (2000) conducted a study employing a cooperative game task in which one subject gave specific, pre-learned instructions containing EC/LC ambiguities to another subject, regarding which piece to move on a game-board, as in example (7):

(7) When that moves $_{\#1}$ the square $_{\#2}$…
 a. … it should land in a good spot (LC)
 b. … will encounter a cookie (EC)

The speaker was thus required to disambiguate these structures so that the correct move would be made by the other player. The authors concluded that even within such a highly constrained discourse context, the prosodic structure was not always predictable from the syntactic structure (in line with the claim that the prosodic structure of an utterance is not necessarily isomorphic with syntactic structure; Nespor & Vogel, 1986; Selkirk, 1984; see Shattuck-Hufnagel & Turk, 1996). In addition, their findings suggested that listeners had to be sensitive to the global prosodic patterns within these sentences in order to correctly interpret local prosodic cues. For example, in (7), when the verb *moves* was followed by an ip boundary (boundary #1), it was interpreted as an early closure structure when *square* (boundary #2) was followed by a word boundary, but it was interpreted as a late closure structure when *square* (boundary #2) was followed by an IPh boundary. These insights were further investigated by Carlson et al. (2001), who manipulated the relative size of two phrase-level prosodic boundaries (i.e., category – ip, IPh) within sentences containing an attachment ambiguity (high vs. low) of a clausal adjunct, as in (8):



(8) Susie learned $_{\#1}$ that Bill telephoned $_{\#2}$ after John visited

Their findings revealed that the relative, rather than the absolute size, of the boundaries triggered specific attachment preferences; that is, more high attachment interpretations were made when boundary 2 was phonologically larger than boundary 1; by contrast, fewer high attachment interpretations were made when boundary 1 was phonologically larger than boundary 2; when both boundaries were of the same size, an average number of high attachments was observed.

Based on these studies, Clifton et al. (2002) proposed the Informative Boundary Hypothesis (IBH). The hypothesis suggests that given a structure such as (8), where the final ambiguous constituent (underlined) can attach either high (to modify the verb *learned*), or low (to modify the verb *telephoned*), the attachment preference will be influenced by the informativeness of the prosodic boundaries. That is, a boundary can be informative to a specific structure if its respective size is larger or smaller than expected at a specific location in the sentence (based on linguistic information, speech rate, etc.), or if it is relatively smaller or larger than a competing boundary at another location. As the material used to develop the hypothesis consists of ambiguous sentences where both potential boundary sites may contain a larger or a smaller boundary, the latter definition of informativeness holds; namely, a stronger boundary at (#1) supports low attachment while a stronger boundary at (#2) supports high attachment. When both boundaries are of the same size, they are no longer informative as to the attachment site of the ambiguous constituent. This implies that other constraints (e.g., acoustic, syntactic, lexical) become more effective in influencing the parsing decision. In other words, the default structural preference becomes the favored parse. Finally, it is also predicted that an irrelevant boundary, which does not precede an ambiguous constituent or follows a potential attachment site, will not be informative (e.g., a boundary after *Susie* in example 8).

The IBH follows the Rational Speaker hypothesis (Clifton et al., 2002, 2006), according to which listeners assume that speakers are rational and consistent in making prosodic choices. More specifically, speakers are expected to produce a prosodic pattern which reflects the intended syntax of the sentence, unless there is another reason that prevents them from doing so. However, there can be instances in which a boundary does not reflect a necessary syntactic boundary, but rather a prosodic necessity; for example, with increasing constituent length, a boundary might be necessary in order to produce the sentence more fluently. In turn, listeners are sensitive to the reason for using larger or smaller prosodic boundaries – namely whether a boundary was necessary due to the length of the phrase or in order to indicate the intended syntactic structure – and consider them accordingly (Clifton et al., 2006). This can imply that even when two boundaries are of the same size, under certain circumstances one might be more informative than the other. For example, in a sentence that contains two large (IPh) boundaries, where one boundary seems too large for that position (unmotivated by the length of the preceding constituent with regard to the meaning of that sentence) while the other boundary is plausible for its position, then the unnecessarily large boundary could be more informative (Frazier, personal communication). For instance, when instructed to choose between two visual paraphrases of ambiguous and unambiguous sentences, listeners found boundary (#1) in (9a) more informative than boundary (#1) in (9b). The results implied that listeners were aware of the purpose of prosodic breaks employment; boundary (#1), flanking the shorter NPs in (9a), was taken to reflect the syntactic structure of the sentence, while boundary (#1), flanking the larger NPs in (9b), was required mainly due to the constituents' length (Clifton et al., 2006).

(9) a. (Pat or Jay) $_{\#1}$ and (Lee) $_{\#2}$ convinced the bank president to extend the mortgage.
b. (Patricia Jones or Jacqueline Frazier) $_{\#1}$ and (Letitia Connolly) $_{\#2}$ convinced the bank president to extend the mortgage.

An advantage of the IBH over previous hypotheses is that it makes specific predictions for syntactic parsing under certain conditions and has been tested using several types of ambiguous structures, including adjuncts (see 8), relative clauses (10), possessives (11), and conjunctions (12):

(10) I met the daughter $_{\#1}$ of the colonel $_{\#2}$ who was on the balcony (relative clause)
(11) The daughter $_{\#1}$ of the Pharaoh's $_{\#2}$ son (possessive)
(12) Johnny $_{\#1}$ and Sharon's $_{\#2}$ in-laws (conjunction/coordination)

However, a potential weakness of the support for the hypothesis 's predictions may lie in the choice of materials and methodology used to test it. That is, the IBH has mostly been concerned with globally ambiguous structures (8-12). In each of these sentences, two possible parses are available, while the surface structure remains the same. This presents a clear advantage in the sense that the same sentence can be manipulated only prosodically to obtain two different meanings, as also noted by other researchers (Schafer, 1997; Snedeker & Casserly, 2010). However, from Carlson et al.'s (2001) and Clifton et al.'s (2002) results, it appears that the default attachment preferences (low-attachment for relative clause ambiguities – Late closure principle; Fodor, 1998, 2002; Frazier, 1987; Frazier & Rayner, 1982b; and high-attachment preference for PP ambiguities - Minimal attachment principle; Frazier, 1987; Frazier & Clifton, 1996) were only modulated, but not reversed, in the presence of prosodic contours that supported the contrasting parses. In fact, high attachment proportions were always below 50% for sentences like (8), (10) and (11), when two competing informative boundaries (e.g., IPh, ip) were present, while the strongest effect on comprehension was triggered by a single boundary (e.g., 0, ip/IPh). That is, prosody does have an effect on these materials, but it seems to be quite limited, while other (lexical and syntactic) factors constrain the preference more strongly.

By comparison, sentences like (12) were more dramatically influenced by the phrasing manipulation. Interestingly, conjunction ambiguities rely on specific groupings in order to be disambiguated and are thus highly responsive to prosodic phrasing (Clifton et al., 2006; Lehiste, 1973; Streeter, 1978; Wagner, 2005; see Wagner & Watson, 2010 for review). On the other hand, adjunct ambiguities (e.g., PP and relative clauses), according to the Construal theory (Frazier & Clifton, 1996, 1997), have non-primary relations (as opposed to obligatory arguments) to the thematic domains in the sentence. When two potential attachment sites are available, a variety of linguistic factors (e.g., semantic, prosodic, pragmatic) may influence the attachment preference. This could explain why prosody may have a weaker influence on these structures.

Some predictions of the IBH received mixed support. For example, the IBH makes the same predictions for structures 10 through 12, namely that the preference for high attachment should decrease with increasing size of the first boundary: (0,ip) > (ip, ip) > (IPh, ip). Whereas this pattern was found in (12), structures (10) and (11) only showed a difference between the smallest and the largest boundary, whereas neither of them differed from the intermediate pattern (ip, ip). It is unclear whether the predictions were inaccurate, or whether the strong structural preference for low attachment in (10) and (11) prevented smaller prosodic manipulations from showing a significant impact.

Finally, it could also be that the specific task utilized contributed to this effect. That is, participants' attachment interpretations were determined by selecting one of two reworded interpretations of each sentence. Since the task was presented after sentence offset and required participants to read two optional answers before making their choice, the results might reflect offline parsing decisions that have little to do with prosodic processing in real-time and more to do with late syntactic and lexical reanalysis.

1.5.2. The Anti-Attachment Hypothesis (AAH)

Unlike the IBH, which suggests prosodic boundaries are interpreted globally (by phrasing sentences into processing units and by competing with one another), the AAH suggests prosodic boundaries are interpreted locally, by serving as cues to syntactic closure (Watson & Gibson, 2005).

The AAH draws from previous accounts of the relationship between prosodic boundaries and syntactic structure in production, as well as listeners' usage of prosodic phrasing as a signal for syntactic structure; in particular, it hinges on such studies arguing that (i) IPh boundaries are generally inserted at (or aligned with) the edges of syntactic boundaries (Cooper & Paccia-Cooper, 1980; Selkirk, 1995; Truckenbrodt, 1999) as well as on (ii) theories estimating that prosodic boundaries are more likely to occur either before or after large constituents (Cooper & Paccia-Cooper, 1980; Ferreira, 1991, 1993). Expanding on these accounts, in their Left/Right Boundary hypothesis, Watson and Gibson (2004b) make two claims; first, since speakers are more likely to produce IPh boundaries after completed (especially long) constituents, a boundary following a constituent should signal it is completed and therefore no further attachments should be made to its head; second, speakers also tend to produce IPh boundaries before a large upcoming constituent, when it is not an obligatory argument (although it can still be a modifier) of the currently processed lexical head. Against this background, the AAH proposes the following rule (Watson & Gibson, 2005, p. 285): "Listeners prefer not to attach an incoming word to a lexical head that is immediately followed by an intonational phrase boundary". That is, IPh boundaries serve as local cues not to attach an upcoming constituent to the pre-boundary word. Contrary to the view that prosodic phrasing groups constituents together (Frazier & Clifton, 1997; Kjelgaard & Speer, 1999; Pynte, 2006; Schafer, 1997; Speer et al., 1996), the AAH assumes that prosodic boundaries force separation (rather than grouping) of constituents. As a result of this rule, the AAH has the following implications (Watson & Gibson, 2005, p. 285):

(a) The presence of an intonational boundary after a lexical head that is the site of subsequent attachment increases processing difficulty.

(b) The presence of an intonational boundary after a lexical head that is not the site of subsequent attachment decreases processing difficulty.

In other words, listeners will prefer to interpret boundaries as cues that the preceding constituent is completed, and will assume that the immediately following material is unlikely to attach to it. Consider example (5), repeated below as (13):

(13) a. The bus driver angered the rider # <u>with a mean look</u>
b. The bus driver angered # the rider <u>with a mean look</u>

In (13a), the IPh boundary signals that the ambiguous PP (*with a mean look*) should not attach locally to the pre-boundary word (*rider*), thus directing listeners towards high-attachment to the verb (*angered*). On the other hand, an IPh boundary after the verb in (13b), signals that this constituent is completed and cannot receive further local attachments, thus directing listeners towards low attachment of the PP to the low noun (*rider*). A recent eye-tracking study by Lee and Watson (unpublished manuscript) tested whether intonational boundaries serve as local or global cues to syntactic structure (Experiment 1). They varied the presence of a boundary in early and late boundary positions within locally ambiguous structures such as (14):

(14) a. Click on the daughters of the gentleman who is/are sitting (baseline)
b. Click on the daughters $_{\#IPh}$ of the gentleman who is/are sitting (low-attachment)
c. Click on the daughters of the gentleman $_{\#IPh}$ who is/are sitting (high-attachment)
d. Click on the daughters $_{\#IPh}$ of the gentleman $_{\#IPh}$ who is/are sitting (baseline)

Note: AAH predictions for each condition are indicated in brackets.

The sentences were balanced with respect to the number aspect of the high (*daughters/father*) and low (*gentleman/girls)* nouns, as well as that of the auxiliary verb (*is/are*). In contrast to previously used material, these structures are fully disambiguated by the auxiliary verb (number agreement). Further, the investigators manipulated the boundary sizes at early and late attachment positions, creating no-boundary, early-boundary (IPh, 0), late-boundary (0, IPh), and two-boundary (IPh, IPh) conditions (fully illustrated in (14)), and measured subjects' error rates and target fixations. Importantly, following principles (a) and (b) above, when two similar prosodic boundaries appear in a sentence (14d), it is predicted that they negate each other's effects (one increases and the other decreases processing difficulty), and therefore the default (baseline) preference should be exhibited for that sentence. The findings showed a marginally significant main effect of Late Boundary ($p<.09$ by subjects; $p<.07$ by items) for the high-attachment versions of the sentences only, indicating that subjects made fewer errors when a late IPh boundary was present than when it was absent. However, no main effect of Early Boundary or any interactions were found for either high or low attachment sentences. To support the AAH's predictions, a comparable early boundary main effect should have been observed for the low-attachment sentences (fewer errors when the early boundary was present), which was not the case. The results also show that the overwhelming preference for the low-attachment over the high-attachment versions obliterated the effect of the prosodic information on the interpretation of the low-attachment sentences: it neither facilitated nor interfered with processing. On the other hand, the less-preferred high-attachment versions benefited (marginally) from the presence of a late IPh boundary. Finally, the expectation that two IPh boundaries would cancel each other's effects and result in a baseline (default) preference was not supported, as conditions (14a) and (14d) showed different error rates (49% and 41.7%, respectively). The fixation data, on the other hand, were more supportive of the AAH's assumption that boundaries serve as local cues for closure by showing significantly ($p<.05$) more fixations away from the pre-boundary word (*daughters*) when an early boundary was present than when it was absent. However, the authors failed to show a comparable main effect of Late Boundary at the late boundary position. They explained that the sentences were quickly disambiguated after the late boundary, which may have negated its influence (it could also be due to the early time-locking position that was chosen, namely the onset of the low-noun – *gentleman* – which contained durational differences due to the presence of the following late boundary in two of the conditions). Because all sentences were pooled together, it is impossible to evaluate whether the incongruence between the behavioral (only Late Boundary main effect) and eye-tracking data (only Early Boundary main effect) reflects a theoretical or a methodological error.

In contrast to the IBH, the AAH considers the influence of only a single, large (IPh) boundary on syntactic interpretations, as smaller (ip) boundaries are not expected to signal closure. It should be asked, however, whether the AAH is able to account for the influence of more than one boundary, or boundaries of different sizes, on sentence processing. In another eye-tracking study (experiment 3), Lee and Watson (unpublished manuscript) varied the size of two boundaries within globally ambiguous structures such as (15):

(15) a. Click on the candle $_{\#ip}$ below the triangle $_{\#ip}$ that's in the blue circle (baseline)
b. Click on the candle $_{\#ip}$ below the triangle $_{\#IPh}$ that's in the blue circle (high-attachment)
c. Click on the candle $_{\#IPh}$ below the triangle $_{\#ip}$ that's in the blue circle (low-attachment)
d. Click on the candle $_{\#IPh}$ below the triangle $_{\#IPh}$ that's in the blue circle (baseline)
Note: AAH Predictions are indicated in brackets.

Unlike in Experiment 1, where the sentences were lexically disambiguated, here listeners' responses were predicted to be more highly influenced by the prosodic manipulation. As in Experiment 1, Lee and Watson predicted that an IPh boundary following a constituent would block its attachment to the post- boundary constituent. For example, in (15b) the IPh boundary after the low noun (*triangle*) should prevent its attachment to the subsequent ambiguous clause (*that's in the blue circle*), while the smaller ip boundary after the high noun (*candle*) should not affect the attachment strength, which would result in a high attachment preference. The opposite is predicted for (15c), which should show a low-attachment preference. On the other hand, in the baseline condition (15a), both weak ip boundaries should not trigger the syntactic closure of their preceding constituents, which predicts a default (baseline) attachment preference. Similarly, in (15d), since both IPh boundaries lower the likelihood of attachment for their preceding constituents equally, no preference should be expected (compared to the baseline). However, a closer examination reveals these predictions do not always coincide with the AAH's principles. According to the AAH, an IPh boundary should block attachment *only locally* (that is, of the post-boundary word to the pre-boundary word), while more distant attachments should still be allowed. This logic explains the high-attachment prediction for (15b), where the IPh appears directly before the ambiguous phrase. In this case, the ambiguous phrase can only attach high, as the low-attachment site is blocked by the IPh. However, this logic does not explain the prediction for a low-attachment preference in (15c), where both attachment sites are not locally blocked for the ambiguous phrase. Because the IPh in (15c) appears after the high noun, it should block its attachment to the low noun only, but not to the ambiguous phrase (this is also true for condition (14b) in Experiment 1). This prediction is more in line with the IBH's global account, where the overall prosodic structure should support this interpretation due to the relative difference between early and late boundary sizes. A purely local closure account cannot make such a prediction. In fact, the authors claim that "theories arguing that boundaries serve as cues to local syntactic structure or serve as points of local processing are not inconsistent with the global prosodic structure influencing syntactic parsing" (Lee & Watson, Unpublished manuscript; p.39). Indeed, like the IBH, the AAH also predicts that an IPh boundary will affect the attachment preference more strongly in the presence of an ip boundary – in the same direction as predicted by the IBH. The AAH claims this is due to a local effect of the stronger boundary, while the IBH claims this effect is due to the global prosodic pattern of the utterance. Moreover, similar to the IBH, the AAH also claims that two boundaries of the same size should yield the same attachment preference (the IBH makes this prediction in terms of "informativeness").

As for the behavioral results, numerically, Lee and Watson seemed to confirm their prediction of baseline attachment preferences for similar size boundaries, as the ip pair (15a) received similar high-attachment responses as the IPh pair (15d) (12.5% vs. 13.7%, respectively[3]), indicating their similar effect on processing. Also, condition (15b), containing a late IPh boundary, received more high-attachment responses (16%) than both baseline conditions, and condition (15c), containing an early IPh boundary, received fewer high-attachment responses compared to all other conditions (10.6%). However, statistically, results showed a significant main effect of Late Boundary, reflecting an increased high-attachment preference in conditions with a late IPh boundary than conditions with a late ip boundary. No other significant effects were found. This finding suggests that regardless of the size of the early boundary, when the late boundary was an IPh, it received more high-attachment ratings. This outcome is in contradiction to the AAH, which would predict a main effect of Early Boundary as well, since it assumes ip and IPh boundaries should affect processing differently at each boundary position separately.

The fixation data show that the presence of an early IPh boundary triggered significantly more fixations *away* from the pre-boundary word (*candle*; Figure 9), as expected by the AAH. The presence of a late IPh boundary triggered significantly more fixations *towards* the high-noun word (*candle*; Figure 10). The authors claimed these findings serve as evidence for the local effect of prosodic boundaries, since each boundary was treated separately at the moment of its appearance, creating an expectation for a certain parse, even before the onset of the upcoming word. However, in order to confirm this assertion, the authors should have also reported fixation rates away from the low noun (*triangle*) when it was followed by an IPh, instead of only fixations to the high-noun. It is possible they assumed that at this point in the sentence, the listeners already made their preference by integrating the information from both boundaries,

---

[3] The low proportion of high-attachment throughout the study is due to an overall low-attachment preference for the relative clause attachment.

which would motivate such an analysis. If this were the case, condition (15d) should have exhibited fewer fixations towards the high-noun, as it should be perceived as a baseline condition (i.e., low-attachment preference). However, both conditions (15b) and (15d), containing a late IPh boundary, showed the same increased rate of high noun fixations compared to conditions (15a) and (15c), which contained a late ip boundary.

The results of the behavioral and eye-tracking data in both experiments are only partially consistent with the AAH's predictions. In particular, the expectation that two boundaries of the same size would cancel each other's effects and favor the default structure was not supported; nor was the expectation that an early boundary would only support a low-attachment preference (a prediction that goes against the AAH's principle of a local closure effect).

1.6. Closure Positive Shift – the electrophysiological marker of boundaries

Although the outcomes of the reviewed studies have provided a much needed initial insight into the relationship between competing boundaries, nearly all of them have been conducted using offline behavioral measures, which often involved participants' subjective judgments. As a result, some of the core predictions of both hypotheses have been either difficult or impossible to examine, including answers to questions like: (i) is boundary information processed in isolation or in an accumulating manner?; (ii) do different boundary sizes impact the brain differently?; (iii) at what point in the sentence is a boundary perceived as informative/closure-inducing?; (iv) how and when do prosodic boundaries interact with other types of information in the sentence (e.g., syntactic, lexical)? In order to examine listeners' linguistic behavior during sentence presentation, instead of after its offset, some studies have used the cross-modal naming task, which has been taken to approximate online parsing decisions

(Kjelgaard & Speer, 1999; Marslen-Wilson et al., 1992; Warren et al., 1995). However, as this measure requires listeners to judge ambiguous sentence fragments (i.e., the same word string with various prosodies) based on the plausibility of a following visual target word, and then use this word to complete the sentence (during which time it is displayed on the screen), it is neither natural, nor one that reflects the actual time-course of sentence processing. By contrast, Event-Related Potentials (ERPs), which continuously measure subjects' brain activity throughout the entire sentence without interruption, allow researchers to examine the interaction between syntax and prosody in an objective online manner. Several ERPs have been found to be correlated with specific linguistic events and their processing difficulties (in terms of amplitude, latency, and duration), including outright syntactic violations (P600; Osterhout & Holcomb, 1992, 1993; Osterhout, Holcomb, & Swinney, 1994) and lexical anomalies (N400; Kutas & Federmeier, 2011; Kutas & Hillyard, 1980, 1984), and their combination (e.g., a biphasic N400-P600 pattern when an NP is not assigned a theta role [e.g., agent, patient]; Friederici & Frisch, 2000). However, despite the growing influence of this technique, few ERP studies had targeted the neural correlates of prosodic phrasing, until the discovery of a new ERP component in a seminal study by Steinhauer, Alter, and Friederici (1999).

Steinhauer et al. (1999; see Steinhauer, 2003, for review) were the first to explore the influence of prosodic phrasing on syntactic disambiguation, using ERPs. They presented listeners with German sentence pairs containing a temporary attachment ambiguity, where the ambiguous NP (*Anna*) is either the indirect object of the first verb (*verspricht – "promises"*), as in (16a), or the direct object of the second verb (*entlasten – "support"*). Due to German word order, the sentences become lexically disambiguated when the second verb is encountered (i.e., *arbeiten – "work"* in (16a); *entlasten – "support"* in 16(b)), at which point it becomes clear that the intransitive verb *arbeiten* does not take a direct object, whereas the obligatory transitive verb *entlasten* does.

(16a) [Peter verspricht Anna zu arbeiten]$_{IPh1}$ [und das Büro zu putzen]
Peter promises Anna to work and to clean the office

(16b) [Peter verspricht]$_{IPh1}$ [Anna zu entlasten]$_{IPh2}$ [und das Büro zu putzen]
Peter promises to support Anna and to clean the office

Based on the Garden-Path model (Minimal Attachment principle; Frazier, 1987), when facing an ambiguity, the parser initially attempts to construct the simplest syntactic structure possible. Thus, the initial preference should be to parse Anna as the object of the first verb, because NP2 in (16b) is more deeply embedded than in (16a), making it structurally more complex. Since both sentences are lexically identical up to the second verb, in the absence of prosodic cues to phrasing (as indicated by the square brackets), sentence (16b) was predicted to induce processing difficulties (or a garden-path effect; Frazier, 1987; Frazier & Fodor, 1978).

Based on accounts of prosody-syntax mapping (Selkirk, 1984) and on controversial reports claiming that prosodic phrasing can override initial structural preferences before the disambiguating word is encountered (Beach, 1991; Marslen-Wilson et al., 1992; Warren et al., 1995), Steinhauer and colleagues predicted that a boundary after the first verb (e.g., IPh1 in (16b)) would be sufficient to prevent the garden-path effect in (16b). Accordingly, they theorized that a superfluous boundary after the first verb in sentence (16a) should initially lead the parser to construct structure (16b), eliciting a garden-path effect when the second verb is encountered. In other words, such manipulation was expected to reverse the garden-path effect, making the default structure more difficult to process. To this end, Steinhauer et al. (1999) introduced a third condition, (16c), which was created by cross-splicing the first part of (16b) with the second part of (16a), resulting in a prosodic-syntax mismatch condition that was lexically identical to (16a), but which contained the (incongruous) early boundary from (16b).

(16c) * [Peter verspricht]$_{IPh1}$ [Anna zu arbeiten]$_{IPh2}$ [und das Büro zu putzen]
Peter promises to work Anna and to clean the office

As predicted, the results confirmed that the prosodic phrasing in all conditions guided the initial parsing decision. The behavioral data showed above 80% acceptability rates for the well-formed conditions, and only 6% acceptability rates for the mismatch condition. Furthermore, the ERP data revealed that the prosodically-induced garden-path in (16c) elicited a biphasic N400-P600 pattern on the second verb, which, similar to previous reading studies using ERPs, was taken to reflect a violation of the verb argument's structure (Rösler, Pütz, Friederici, & Hahne, 1993). Most importantly, a large positive centro-posterior waveform, spanning about 500ms, was elicited relative to each IPh boundary in all three conditions. A follow-up examination showed that this brainwave was not dependent on the presence of particular acoustic cues, such as a pause interrupting the speech stream, but rather on the phonological markers of a boundary (e.g., pre-final lengthening). Therefore, it was taken to reflect the closure of major prosodic boundaries, and was termed the Closure Positive Shift (CPS) accordingly (Bögels, Schriefers, Vonk, & Chwilla, 2011). Thus, Steinhauer et al. (1999) revealed not only that prosodic phrasing immediately interacts with the parser and can prevent and induce syntactic violations, but also that these processes are reflected by a brain response in ERPs.

Following studies have since replicated the CPS in German (Isel, Alter, & Friederici, 2005; Pannekamp, Toepel, Alter, Hahne, & Friederici, 2005; Steinhauer & Friederici, 2001) and in numerous other languages, including in English (Itzhak, Pauker, Drury, Baum, & Steinhauer, 2010; Pauker, Itzhak, Baum, & Steinhauer, 2011; Steinhauer, Abada, Pauker, Itzhak, & Baum, 2010), Dutch (Bögels, Schriefers, Vonk, Chwilla, & Kerkhofs, 2010; Kerkhofs, Vonk, Schriefers, & Chwilla, 2007, 2008), Japanese (Mueller, Hahne, Fujii, & Friederici, 2005), Korean (Hwang & Steinhauer, 2011), and Chinese (Li & Yang, 2009). The CPS has also been reliably evoked by boundaries in a variety of conditions, including Jabberwocky and pseudoword sentences (Pannekamp, et al., 2005), hummed and low-pass filtered sentences (Pannekamp et al., 2005; Steinhauer & Friederici, 2001), comma rules in written sentences (Steinhauer & Friederici, 2001), and even in the absence of overt boundaries, due to an expectation of an obligatory boundary following a transitive verb (Itzhak et al., 2010). Thus, as hypothesized by Steinhauer et al. (1999), the CPS reflects the prosodic phrasing of the incoming input, whether it is acoustically perceived or mentally generated based on syntactic and lexical constraints.

While the evidence on the CPS has been slowly accumulating over the past decade, several aspects of the CPS have yet to be examined. First, it is still unknown if the CPS is an all or none effect, or whether it is affected by boundary size. If the latter is the case, then in what manner is it affected? According to the categorical approach, only two different magnitudes should be observed if the CPS taps into the phonological domain, meaning it should reflect the two phonological categories. On the other hand, if prosodic boundaries are perceived on a continuum, then the CPS might display more subtle distinctions between boundary sizes. To date, only one study in Chinese (Li & Yang, 2009) tested the CPS with ip boundaries, and found no differences between the prosodic categories (although the presented data was rather noisy).

Second, does the magnitude of the CPS at a boundary affect the magnitude of boundary-induced garden-path effects (N400, P600)? Previous ERP studies have shown that (a) the amplitude of ERP garden-path effects reflects the severity of the violation in a gradient manner (Osterhout et al., 1994) and that (b) the *type* of prosody-induced garden-paths can result in distinct ERP correlates (Pauker et al., 2011; Steinhauer et al., 1999). However, these studies have been mostly concerned with the location or presence of prosodic boundaries, not the manipulation of their size. Everything else being equal, does increasing boundary size also increase the size of the CPS as well as the size of the corresponding garden-path effects? Answering this question may add another crucial layer of understanding to the interplay of prosodic and syntactic information.

1.7. The present research project

The studies in this dissertation investigated the predictions of the Anti-Attachment and the Informative Boundary hypotheses regarding syntactic preferences in the context of early (EC) and late (LC) closure ambiguities, as well as their (common) underlying assumption that prosodic boundaries should exhibit only categorical effects. To this end, two behavioral and one ERP studies were designed. All experiments employed an acceptability judgment task on a scale, as previous studies (de Pijper & Sanderman, 1994; Wagner & Crivellaro, 2010) have demonstrated this scoring system better taps into listeners' natural sensitivity to perceived differences in boundary size (compared to a traditional binary scoring system). The experiments were also designed such that they can be tested both behaviorally and electrophysiologically. That is, the sentence structure, material recording (specifically with respect to pitch contour and range, and speech rate), and prosodic manipulation, were highly controlled. This also enabled a direct comparison between the results of the studies.

Study 1 (Chapter 2), was comprised of Experiments 1 and 2. Both experiments targeted the predictions of the two hypotheses behaviorally. In Experiment 1, nine prosodic conditions in both structures (18 overall) were presented, and in Experiment 2, sixteen prosodic conditions in both structures (32 overall) were presented. Because this was the first time such structures have been tested with two phrase-level boundaries across conditions, the conclusions from Experiment 1 concerning the type of manipulation required for EC or LC preference, were applied to Experiment 2. Study 2 (Chapter 3), was comprised of Experiment 3, which employed the same material used in Experiment 2. In this study, the online effect of prosodic boundaries was tested using ERPs, primarily targeting the categorical vs. gradient debate. As we were interested in evaluating the magnitude of the CPS in each condition, all 16 conditions were required (4 Early Boundary and 4 Late Boundary sizes). This did not allow us to directly compare the predictions of the hypotheses, as fewer trials were used in each condition (ERP). However, it did enable us to examine the processing difficulties triggered by these boundary sizes. All three experiments were thus designed to provide a comprehensive view of the issues at hand, using a novel paradigm and an objective measure.

This set of studies suggests that boundaries are indeed processed in a gradient rather than a categorical manner, and that the pattern of acceptability mirrored the parametric manipulations of both boundaries. Results also show that in structures like EC/LC, where boundaries are necessary for disambiguation and thus expected by listeners, the early boundary was highly influential in directing the parser, making late boundaries of the same size less effective in adjusting or changing this initial bias. Moreover, the use of a subtle scoring system required listeners to comprehend the sentences (which was not necessarily the case in previous studies), revealing a surprisingly differentiated influence of boundary location on structural preference.

Electrophysiologically, we found that subtle differences between prosodic boundaries are indeed detected by the brain and affect the degree of processing difficulty. The amplitudes of both CPS and garden-path components were, at least to some extent, gradually influenced by the boundary size. The results of the ERP study also reflected the behavioral results, suggesting less severe processing difficulties in EC compared to LC. While these outcomes did not conform with, or even contradicted, the predictions of the two hypotheses, they were largely in line with the Boundary Deletion Hypothesis, predicting long lasting impact of prosodic boundaries on the parser (Steinhauer & Friederici, 2001), which was recently successfully tested using similar material (Pauker et al., 2011). We thus extended this hypothesis, based on the new findings of Study 1, to account for an array of boundary sizes, and validated its predictions based on the behavioral and ERP online findings of Study 2.

CHAPTER 2: Study 1

Rethinking the role of prosodic boundaries in ambiguous sentence resolution

Efrat Pauker

Michael Wagner

Karsten Steinhauer

ABSTRACT

Prosodic phrasing is essential to language comprehension, and has been demonstrated to be effective in facilitating and interfering with the resolution of a variety of syntactically ambiguous structures. Recent studies explore the interaction between two prosodic boundaries on sentence processing, both in terms of size and location within the sentence. In this study, we conducted two experiments to examine the predictions of two of the most influential hypotheses in this field. The first assumes that large boundaries influence syntactic processing locally, and the second assumes that the relative magnitude of boundaries across an utterance modulates parsing. We also tested their shared assumption that prosodic boundaries should exhibit strictly categorical effects. We presented listeners with spoken English garden-path sentences (EC/LC), each of which contained two prosodic boundaries. Boundary size was varied digitally to simulate a continuum, rather than categorical distinctions. Contrary to the predictions of both hypotheses, results showed a general advantage for EC over LC and a gradient pattern of acceptability that mirrors the parametric manipulation of both boundaries. Importantly, the acceptability of each structure was exclusively affected by the size of the superfluous boundary (e.g., early boundary in LC). To account for these findings, we extended an existing hypothesis predicting prosodic interference effects increase processing difficulty.

2.1. Introduction

2.1.1. Prosodic phrasing and syntactic ambiguity

Human languages contain a variety of ambiguities ranging from lower-level phonetic feature uncertainties (e.g., /s/ vs. /f/) to higher-level structural misunderstandings. Among the various types of syntactic ambiguities are those which initially (mis)lead the parsing mechanism to commit to a specific structural interpretation that is later revealed to be incorrect as subsequent lexical information is encountered, termed "garden-path" sentences. The name garden-path implies that the parser is led down a wrong "processing path", and is later required to backtrack and re-evaluate the initially assumed structure, often at a cost of additional processing load (varying depending on the degree of commitment and subsequent violation). According to the garden-path model (Frazier, 1987; Frazier & Fodor, 1978a), the parser employs a set of strategies in order to reduce the load on working memory. One principle of that model that is highly relevant to the current study is Late Closure (LC), which states: "when possible attach incoming lexical items into the clause or phrase currently being processed" (Frazier & Rayner, 1982a; p.180). For example, in the sentence "*I convinced her children are noisy*", the parser is likely to initially treat the word *her* as the possessive pronoun of *children* and group them together, rather than assume they belong to different phrases. While such ambiguities are resolved relatively late in reading (when the word *are* is encountered, in this example), in spoken language, prosodic phrasing provides earlier cues, which, in many cases, prevent the ambiguity altogether. That is, by grouping the sentence prosodically (i.e., [*I convinced her*] [*children are noisy*]) the intended syntactic phrasing becomes obvious to the listener. Many studies have demonstrated that naïve speakers phrase ambiguous spoken utterances differently to convey different meanings to listeners[4] (Kraljic & Brennan, 2005; Schafer et al., 2000, 2005; Snedeker & Trueswell, 2003; Warren, 1985), and that listeners use prosodic phrasing to draw the intended meaning from spoken sentences (Kraljic & Brennan, 2005; Price et al., 1991; Snedeker & Trueswell, 2003)[5].

2.1.2. Prosodic boundaries: the right edges of phrase-level prosodic constituents

The key element of prosodic phrasing is the placement and magnitude prosodic boundaries. Prosodic boundaries are defined as perceptual breaks in the natural speech stream and are marked by some (if not all) of the following cues: Pre-boundary final syllable lengthening, fundamental frequency changes, pause, and a pitch reset (see Cutler et al., 1997; Swerts, 1997; Wightman et al., 1992).

To account for the mapping between prosodic phrasing and syntax, several highly-influential theories have proposed that prosody is characterized with a “strictly layered” hierarchical structure that, similar to syntax (e.g., morpheme < word < phrase < sentence), is comprised of prosodic constituents (e.g., foot < phonological word < intermediate/phonological phrase < intonational phrase) that are exhaustively parsed into units of the next smaller level (Nespor & Vogel, 1986; Selkirk, 1984; 1995; but see Ladd, 1986, for a recursive account). Two types of boundaries have been theorized at the phrase level, at the right edges of their

---

[4] Although speakers differentiate between the possible parses by producing contrasting prosodic patterns, their performance depends on material and task demands. When the task explicitly requests disambiguation, a higher consistency of performance (more reliable prosodic disambiguation) is found (Cooper, Paccia, & Lapointe, 1978) compared to studies in which context provides disambiguating information (Allbritton, McKoon, & Ratcliff, 1996; Snedeker & Trueswell, 2003); in which case, prosodic disambiguation is less reliable and more inconsistent.

[5] It is important to mention that not all syntactic ambiguities can be resolved – equally or at all – using prosodic phrasing (for examples see Lehiste, 1973; Price et al., 1991). We will later discuss this issue in the context of the current project.

corresponding constituents – the smaller intermediate (or phonological) boundary (ip), and the larger intonational phrase boundary (IPh). The majority of research focusing on prosodic phrasing in English has subscribed to this dichotomy, and to its implementation via the Tones and Break Indices (ToBI) annotation system (Beckman & Elam, 1997; Beckman & Hirschberg, 1994; Beckman & Pierrehumbert, 1986; Shattuck-Hufnagel & Turk, 1996; Silverman et al., 1992), which describes prosodic categories in terms of their corresponding intonation contour (*tones*) and degree of perceived disjuncture between them (*break indices*). Specifically, the smaller ip units are characterized with one or more *pitch accents* (*), which mark stressed syllables with a rising (H*), falling (L*), or combined (e.g., L+H*) Fø. Following the pitch accent comes the *phrase accent* (-), outlining the intonation contour from that point to the right edge of the phrase, with either a rising (H-) or a falling (L-) Fø (Pierrehumbert, 1980; for review see Shattuck-Hufnagel & Turk, 1996). Each IPh is comprised of at least one ip, such that the final ip's right edge coincides with that of IPh's, which is characterized with a *boundary tone* (%). Like the other tonal events, it is either high (H%) or low (L%). Finally, each category is labeled with a different break size from an inventory of 5 sizes in total (i.e., 0-4), the largest of which is assigned to the IPh, and the next smaller (3) is assigned to the ip. Thus, the phrase-level categories differ with respect to a single tonal event (e.g., H*L-H%) and a slightly different break index – a difference which can be made more subtle when both edge tones follow the same pattern (H*L-L%). This fact has been utilized by studies attempting to create boundaries of an ambiguous nature, usually in mid-sentence locations (e.g., Kjelgaard & Speer, 1999). Also, since these measures are relative rather than definitive and depend on the labeler's visual and auditory perception of the speech file (see Wightman, 2002), a considerable acoustic variability within and across categories can be found across studies (Carlson et al., 2001; Kjelgaard & Speer, 1999;

Lee & Watson, personal communication; Snedeker & Casserly, 2010). This, in combination with criticism that the ToBI labelling may not faithfully represent the auditory nature of the speech file, since defined categories are pre-assigned a specific break index, as well as a low annotation consistency of tone types (Hirst, 2005; Wightman, 2002), has challenged the strict categorical division between boundary types. We will return to this point below.

2.1.3. Intonational phrase (IPh) boundaries and syntactic ambiguity resolution

Since prosodic phrasing "may impose an organization on the linguistic input" (Carlson, Clifton, & Frazier, 2009, p.1067), it can also override and even reverse such syntactic parsing principles like late closure that do not take into consideration prosodic information. For example, when prosodic boundaries are aligned with syntactic boundaries, they can affect attachment preferences and drive the parsing preference. This is often the case for those structures in which (different) syntactic groupings are achieved using different prosodic phrasing (Carlson, 2009). For example, listeners interpret sentences like (1) more easily and reliably when they are presented with prosodic boundaries (i.e., which group *old* and *men* or *men* and *women* together) compared to sentences like (2), which do not seem to be facilitated by the insertion of boundaries (Lehiste, 1973):

(1) Old men and women stayed at home
(2) The shooting of the hunters was terrible

Many studies have focused on the effect of prosodic boundaries – in particular, the more prominent Iph boundary category – in resolving different types of structural ambiguities that would otherwise induce processing difficulties (i.e., garden-path effects), including prepositional phrase (PP) attachments (Pynte, 2006; Pynte & Prieur, 1996; Snedeker & Trueswell, 2003) and closure ambiguities (Beach, 1991; Kjelgaard & Speer, 1999; Marslen-Wilson et al., 1992; Nagel et al., 1996; Schafer et al., 2000; Speer et al., 1996; Speer et al., 2003; Walker et al., 2001; Warren et al., 1995). Among these, one of the best-studied structures is the late (LC) and early (EC) closure ambiguity (Frazier & Rayner, 1982). Consider the following sentence fragment in (3):

(3) While the man parked cars…
(3a) …bikes were waiting (LC)
(3b) …were waiting (EC)

Due to the optionally transitive property of the verb *parked*, it may or may not require an argument, making the noun phrase *cars* ambiguous: it can either be the direct object of the verb *parked* of the subordinate clause (late closure interpretation), as in (3a), or the subject of the matrix clause *cars were waiting* (early closure interpretation), as in (3b). In reading, LC is the preferred closure interpretation (Frazier & Rayner, 1982), resulting in a garden-path effect when the disambiguating verb (*were*) is encountered in (3b). However, auditory studies using the same structures (Kjelgaard & Speer, 1999; Marslen-Wilson et al., 1992; Pauker et al., 2011; Speer et al., 1996; Steinhauer et al., 1999; Warren et al., 1995) revealed that the presence of a prosodic boundary before the ambiguous NP (the less-preferred EC structure according to the LC principle) prevented the ambiguity in (3b), even before the disambiguating region was encountered (Beach, 1991). These findings suggest that prosody can impose a syntactic preference on the parser.

An influential study by Kjelgaard and Speer (1999) revealed that prosodic phrasing not only overrides the LC preference, but also (at least partially) reverses it. They presented subjects with EC and LC sentence pairs in three conditions, which differed in terms of their phrasing properties: (i) *cooperating*, (ii) *baseline* and (iii) *conflicting* (e.g., *Whenever the guard checks #1 the door #2 is/ it's locked*). In the cooperating condition, the syntactic and prosodic boundaries coincided, creating well-formed sentences. In the baseline condition, the sentences were produced with weaker boundaries and relatively flat (or neutralized) intonation and were designed to be highly compatible with both EC and LC versions based on listeners' judgments. The conflicting conditions were created by digitally cross-splicing the cooperating structures such that the prosodic boundaries were aligned with the syntactic boundaries of the other structure (i.e., early boundary in LC and late boundary in EC), creating prosody-syntax violations. By measuring naming and response times as well as error rates, the investigators estimated the degree of processing difficulty for each condition. Their results showed that both EC and LC structures were facilitated equally by the cooperating boundaries and, importantly, that the conflicting boundaries led to garden-path effects in both structures. That is, prosodic phrasing was recognized at the very early stages of processing and guided syntactic parsing accordingly (Kjelgaard & Speer, 1999).

Nevertheless, Kjelgaard and Speer also found that LC sentences *in general* were more easily processed compared to their EC counterparts in both baseline and conflicting conditions, and explained these results as reflecting the structural preferences of late closure (as seen in reading). However, a similar study failed to observe a comparable advantage for the LC interpretation using the same prosodic manipulations but different EC/LC structures, like the ones illustrated in sentence (3) above (Walker et al., 2001). In addition, a later reading study in German (Steinhauer & Friederici, 2001) demonstrated that the effect of even subvocally-generated prosodic boundaries (as indexed by commas) on the processing of such structures is virtually impossible to ignore (and recover from). Their findings, which showed greater processing difficulties in LC structures with a superfluous comma compared to EC structures without a required comma (based on German punctuation rules), led to the formulation of the Boundary Deletion Hypothesis (i.e., BDH). According to the BDH, the mental deletion of an incompatible superfluous boundary should present greater difficulty to the processor compared to the retroactive mental insertion of a missing boundary. This hypothesis has since received support from studies in the auditory modality, showing that the garden-path effect elicited in EC sentences without any prosodic boundaries (a "classical" GP structure; e.g., *whenever a bear was approaching the people came running*) was significantly milder compared to the strong garden-path effect elicited in LC sentences containing an additional boundary (e.g., *whenever a bear was approaching # the people # the dogs came running*) (Pauker et al., 2011; Steinhauer et al., 1999). Follow-up studies have also demonstrated that the presence of prosodic boundaries completely overrides other competing lexical and structural preferences that would otherwise bias the parser (Itzhak et al., 2010). A recent study in Dutch employed coordination ambiguities (NP versus clause coordination) and also confirmed the predictions of the BDH, finding that "superfluous prosodic breaks lead to more severe processing problems than missing prosodic breaks" (Bögels, Schriefers, Vonk, Chwilla, & Kerkhofs, 2013).

Based on this evidence, Pauker et al. (2011) suggested another explanation to the LC advantage in the Kjelgaard and Speer (1999) study, namely, that unlike the EC sentences, the LC sentences did not undergo the assumed "classical" revision whereby the early conflicting boundary was mentally deleted and a later boundary was created. If that had been the case, both EC and LC garden-path sentences should have been equally difficult to process, given the similar facilitating effect prosodic boundaries had on their well-formed versions, and given the symmetry of required syntactic reanalyses (Gorrell, 1995). Rather, since in all LC sentences the ambiguous NP (*the door*) was the antecedent of the immediately following anaphoric pronoun (*it's*), the violation could have been quickly resolved by mentally inserting an additional boundary after the ambiguous NP to create a clause with a topicalized NP (*whenever the guard checks # the door – it's locked*). As for the LC advantage in the baseline conditions, the authors themselves hypothesized this could have been the effect of the transitively-biased initial verb (*checks*), observed in the preceding norming study.

2.1.4. Intermediate phrase (ip) boundaries and syntactic ambiguity resolution

To test whether the facilitation and interference effects found with IPh boundaries can also be demonstrated with ip boundaries, Kjelgaard and Speer (1999) replicated their earlier experiments with ip (they used the term 'phonological phrase' – PPh) instead of IPh boundaries (Experiment 4). In this experiment, they also aimed to validate whether the silent interval typical of IPh boundaries, rather than the mental representation of the phonological category itself, was responsible for their previous findings, by allowing additional processing time to resolve the ambiguity. Lastly, they argued that the ip is a level assumed to be more closely related to syntactic phrasing than the IPh (representing smaller constituents within the utterance), and therefore should present similar outcomes (although the prosody-syntax mapping is not exact; see Shattuck-Hufnagel & Turk, 1996 for review). They argued that "*because each prosodic constituent is exhaustively parsed into constituents at the next lowest level of the hierarchy ("strict layering," see Selkirk, 1995; Nespor & Vogel, 1986), whenever the right edge of an IPh occurs, it is immediately preceded by the right edge of a PPh. Thus, whenever we have associated a syntactic choice point with an IPh boundary and boundary tone, we have also associated it with a PPh boundary and phrase accent*" (p.179). This can be illustrated using examples (4a) and (4b) below, where each IPh boundary is immediately preceded by an ip boundary:

(4a) ((Whenever the guard checks)$_{Ip}$)$_{IPh}$ ((the door is locked)$_{Ip}$)$_{IPh}$
(4b) ((Whenever the guard checks the door)$_{Ip}$)$_{IPh}$ ((it's locked)$_{Ip}$)$_{IPh}$

They reasoned that the main differences between the categories are the longer phrase-final lengthening and pause duration in the IPh, but due to their close prosodic relatedness, both categories should have a similar impact on syntactic parsing decisions. Compared to the IPh boundaries, which were produced with a level 4 break (the maximum break size allowed by ToBI), the newly produced ip boundaries were intended to be compatible with levels 2 and 3 breaks and contained shorter phrase-final lengthening (based on the labeling of two listeners trained in phonetics). However, comparison between the average silent duration after each boundary type shows the differences were much larger than described, with 471 ms and 507 ms silent periods after the EC and LC IPh boundaries, respectively, compared to less than 1 ms following each of the ip boundaries. In fact, the silent period following the ip boundaries was shorter than that following the boundary positions in both the baseline (8-24 ms) and cooperating (19-27 ms) conditions in experiments 1-3, where a boundary was either supposed to remain ambiguous or completely absent (e.g., early boundary in the cooperating LC condition). Additionally, the selected pitch contour for the ip boundaries was less prominent than the one selected for the IPh boundaries, with more subtle rising pitch and phrase accents, compared to the more clearly defined rising and falling accents and low boundary tone for the IPh. As ip boundaries can technically also be produced with a high-low (H*-L-) pitch contour (but without a boundary tone), this manipulation seems to have been specifically chosen in order to increase the difference between the boundary types even further. Nevertheless, despite these considerable differences between the boundary types, results showed that ip boundaries influenced syntactic preferences similarly to IPh boundaries – both qualitatively and quantitatively – consistent with Kjelgaard and Speer's (1999) assumption that "*prosodic boundaries need not involve large pitch excursions, extensive phrase-final lengthening, or substantial silent durations to be effective in the resolution of temporary syntactic ambiguity*." (p.185).

It is important to mention that although no silent interval (i.e., a pause) followed the ip boundaries, the early and late closure sentences showed a significant difference in the duration of the pre-boundary words at the critical regions. That is, the duration of Verb1 (e.g., *checks*) was significantly longer in EC, and the duration of NP2 (e.g., *the door*) was significantly longer in LC. These differences were reversed in the conflicting conditions which were derived from the cooperating conditions using cross-splicing. These findings are important for two reasons: first, both levels of prosodic boundaries were found to guide syntactic parsing; second, durational differences in the form of pre-final lengthening – the common acoustic marker of both ip and IPh boundaries – was a sufficiently salient factor to trigger this effect.

In contrast to the more established view on the role of major intonational boundaries (IPh) in the comprehension of ambiguous structures, the effect of ip boundaries on syntactic parsing, which has been studied to a lesser extent, remains controversial. That is, some theorists hold that only IPh boundaries can modify syntactically ambiguous interpretation preference (Blodgett, 2004; Marcus & Hindle, 1990; Price et al., 1991; Watson & Gibson, 2004b, 2005).

However, a facilitative effect of ip boundaries on syntactic processing was found not only in studies using closure ambiguities (Hwang & Schafer, 2006), but also in PP attachments (Schafer, 1997; Snedeker & Casserly, 2010), adjunct attachments (Carlson, Clifton, & Frazier, 2001), conjunction and relative clause ambiguities (Clifton et al., 2002), and homophonic ambiguities (Millotte et al., 2008; Millotte et al., 2007). Similar to Kjelgaard and Speer (1999), in all of these studies, the ip boundaries were distinguished from the IPh boundaries in terms of durational measures: shorter pre-boundary word lengthening and no audible silence interval (or a very minimal one). Note that the intonational differences between the boundary types (e.g., H*-L– in an ip vs. H*-L–L% in an IPh) are also a reflection of the durational differences when the phrase accent (L–) and boundary tone (L%) go in the same direction, and are expressed in terms of longer pre-final lengthening. This issue of durational differences between ip and IPh boundaries is highly relevant to the current study and will be discussed later in the context of our experiments.

2.1.5. Local vs. global approaches of prosodic processing

Studies on prosodic phrasing have traditionally compared the effect of the presence vs. the absence of a single (mostly large) boundary on syntactic parsing. However, given both large and small boundary types have been found to affect comprehension (at least to some extent), recent investigations have begun to expand on previous findings by exploring the influence of two competing boundaries of different sizes (ip/IPh) on processing (Carlson et al., 2001; Clifton et al., 2002; Frazier, Carlson, et al., 2006; Schafer, 1997; Snedeker & Casserly, 2010; Watson, 2002; Watson & Gibson, 2004a, 2005). Two opposing frameworks have been developed thus far; the first argues that boundaries serve as strictly *local* cues to syntactic closure, and are thus processed independently of one another (Watson & Gibson, 2004a, 2005); the second argues that prosodic boundaries are processed relative to one another, such that the *global* prosodic structure of the entire sentence affects interpretation (Carlson et al., 2001; Clifton et al., 2002; Schafer, 1997). The Anti-Attachment Hypothesis (AAH) and the Informative Boundary Hypothesis (IBH) are currently the most influential hypotheses representing the local and global views, respectively. In the present study, we will examine the predictions of these hypotheses.

Thus far, both accounts have mainly focused on the interaction (or lack thereof) between two boundaries in the context of (global) attachment ambiguities containing two optional attachment sites (referred to as high [#1] or low [#2]), as in the following prepositional phrase (PP) ambiguity (Schafer, 1997):

(5) The engineer recorded $_{\#1}$ the musicians $_{\#2}$ <u>with lousy equipment</u>

The ambiguity in the above example emanates from whether the ambiguous PP (underlined) should attach high (#1) or low (#2), as it can either serve as an (optional) Instrument (VP-attachment), implying the recording was done using lousy equipment, or as a Modifier (NP-attachment), implying the musicians played lousy instruments (see Snedeker & Trueswell, 2003).

Unlike EC/LC sentences, such global ambiguities are typically not lexically disambiguated, unless they are preceded by disambiguating context (e.g., *how did the engineer record the musicians?* [#1]; *what kind of equipment did the musicians use for the recording?* [#2])[6]. However, since both meanings can be expressed by the same string of words, the proponents of both hypotheses have found these structures advantageous for prosodic manipulation (see Snedeker & Casserly, 2010, for further discussion). The underlying assumption is that any changes in listeners' sentence interpretation or judgments can be attributed exclusively to the prosodic manipulation (i.e., the different combinations of boundaries). To illustrate how each of the two frameworks explains these effects, we will first review their main assumptions and apply them to Example 5 above.

2.1.5.1. The Anti-Attachment Hypothesis (AAH)

The AAH (Watson, 2002; Watson & Gibson, 2004a, 2005) makes strong reference to the relationship between the prosody-syntax mapping in production and perception, and in particular to observations and hypotheses that large (IPh) prosodic boundaries generally coincide with large syntactic boundaries (Cooper & Paccia-Cooper, 1980; Selkirk, 1995; Truckenbrodt, 1999) and are more likely to appear before or after large constituents (Cooper & Paccia-Cooper, 1980;

[6] This limitation is often not inherent to the structural ambiguity (i.e., the PP could be chosen such that it makes sense only with one of the two attachment options: *The engineer recorded the band with the crazy name*), but is rather a choice of design, typically motivated by the task selected.

Ferreira, 1991, 1993). Since speakers tend to produce IPh boundaries after constituents to signal their completion, it is assumed that listeners utilize this information as a cue for syntactic closure, and expect no further attachments will be made to the pre-boundary word (Watson & Gibson, 2004b). Based on these accounts, Watson (2002, p. 82) originally proposed that:

> Listeners use intonational boundaries as cues to signal where not to make an attachment. Listeners prefer not to attach an incoming word to a lexical head that is immediately followed by an intonational boundary.

Thus, in contrast to other accounts (Frazier & Clifton, 1997; Kjelgaard & Speer, 1999; Pynte & Prieur, 1996; Schafer, 1997; Speer et al., 1996, to name a few), prosodic boundaries are perceived as separating, rather than grouping, elements. That is, boundaries are perceived by listeners as local cues that the pre-boundary constituent is complete, preventing further attachments to it. The AAH thus predicts that any attachments made to a word immediately followed by an IPh boundary are highly unlikely and should result in increased processing difficulty. On the other hand, ip boundaries are not predicted to trigger the semantic wrap-up effect associated with syntactic closure and should not present any processing difficulties when attachments are made to words preceding them (Watson, Gibson, personal communication). Importantly, the AAH also assumes incremental processing of boundaries; therefore, the presence of two competing IPh boundaries in the same sentence is expected to equally block each attachment site from attaching to the ambiguous phrase, effectively cancelling the effect of one another, and resulting in a baseline (or default) syntactic preference, but at a cost of increased processing demands (Watson & Gibson, 2005). Interestingly, the presence of two ip boundaries should yield the same outcome, as none of the boundaries directs the parser towards either interpretation, but at no additional processing cost (see also Lee & Watson, Unpublished manuscripts). Given these predictions, let us re-examine example (5), as repeated in (6) below:

(6a) The engineer recorded $_{\#ip}$ the musicians $_{\#ip}$ with lousy equipment
(6b) The engineer recorded $_{\#ip}$ the musicians $_{\#IPh}$ with lousy equipment
(6c) The engineer recorded $_{\#IPh}$ the musicians $_{\#ip}$ with lousy equipment
(6d) The engineer recorded $_{\#IPh}$ the musicians $_{\#IPh}$ with lousy equipment

Since PP attachment ambiguities have an intrinsic high-attachment bias (Minimal Attachment principle; Frazier, 1987; Frazier & Clifton, 1996), the baseline preference should be for a VP-attachment interpretation. According to the AAH, the IPh boundary after NP2 (*musicians*) in (6b) should prevent the subsequent ambiguous PP from attaching to it and instead favor high-attachment to the VP. Similarly, the IPh boundary after the VP in (6c) should block further attachments to this site, and instead support low-attachment to NP2. Both (6a) and (6d) should exhibit a high-attachment preference, although the difficulty associated with reducing the weights of both attachment sites by the IPh boundaries should result in lower high-attachment preference compared to the sentence containing two ip boundaries.

To our knowledge, only one study has directly tested the predictions of the AAH, using relative clause attachment ambiguities, where the overall preference is for low-attachment (Late Closure principle [Fodor, 1998, 2002]; e.g., *Click on the candle* $_{\#1}$ *below the triangle* $_{\#2}$ *that's in the blue circle*) (Lee & Watson, Unpublished manuscript).

However, the results of this study were inconclusive, as some of the analyses did not support the predictions of the AAH. For example, in contrast to their assumption that IPh but not ip boundaries should influence parsing, the difference between the ip and IPh at the early boundary location was not statistically significant (a finding not addressed by the authors). Note that a late IPh boundary, in such structures, supports the less-favored high-attachment interpretation. Given the overwhelming low-attachment ratings (90-80% across conditions), it is likely that the structural bias made the size of the early boundary type irrelevant/redundant, while a larger competing late boundary was more informative to the parser.

2.1.5.2. The Informative Boundary Hypothesis (IBH)

Whereas the AAH emphasizes that prosodic breaks are motivated by syntactic constraints in the speaker and should prevent listeners from making wrong parsing decisions, the IBH (Clifton et al., 2002) is rooted in a theoretical background that proposes prosodic "packages", or domains, facilitate correct syntactic parsing (Schafer, 1997). Specifically, Schafer (1997) claims that attachment of a node within the same prosodic domain should be easier (more "visible") than attachment of a node across prosodic domains. In addition, the visibility of an attachment site to the parser should be gradient across multiple prosodic domains. This proposal was the first to argue that the prosodic structure throughout the sentence, rather than isolated prosodic cues, should affect processing. Schafer (1997) originally defines prosodic domains in terms of ip boundaries. This is another important difference compared to the AAH, which dismisses the potential influence of smaller phrase-level boundaries. However, Schafer (1997) did not directly discuss the relative difference between the processing of ip and IPh boundaries – a gap which the IBH has sought to fill. Integrating this hypothesis with later findings showing that prosodic phrasing is realized in a variety of ways for a given structure (Schafer et al., 2000), and that syntactic preferences of adverbial adjunct ambiguities (*Susie learned $_{\#1}$ that Bill telephoned $_{\#2}$ <u>after John visited</u>*) are influenced by the relative difference between the early and late boundary sizes (Carlson et al., 2001), Clifton et al. (2002) proposed a more comprehensive hypothesis.

According to the IBH, it is not the absolute phonological size of the boundaries that matters, but rather the relative difference between them. That is, listeners attend to the overall difference between the boundaries across the sentence, and draw the intended meaning from this pattern, assuming the speaker makes rational and consistent prosodic choices that reflect the location of syntactic boundaries (the Rational Speaker Hypothesis; Clifton et al., 2002; see Frazier et al., 2006, for review). Based on this logic, boundaries are considered informative to the structure of the sentence if they differ in size (i.e., IPh vs. ip), whereas same-size boundaries are considered uninformative, in which case other linguistic constraints (e.g., lexical, syntactic) influence the parsing decision, resulting in the default (or baseline) syntactic preference. Let us examine the predictions of the IBH using example (5), repeated as (7) below:

(7a) The engineer recorded $_{\#ip}$ the musicians $_{\#ip}$ with lousy equipment
(7b) The engineer recorded $_{\#ip}$ the musicians $_{\#IPh}$ with lousy equipment
(7c) The engineer recorded $_{\#IPh}$ the musicians $_{\#ip}$ with lousy equipment
(7d) The engineer recorded $_{\#IPh}$ the musicians $_{\#IPh}$ with lousy equipment

According to the IBH, the larger late boundary in (7b) encourages high-attachment of the PP to the VP (VP-attachment), whereas the larger early boundary in (7c) encourages low-attachment of the PP to NP2 (NP-attachment). The equal-size boundaries in (7a) and (7d) are uninformative; therefore, the default high-attachment preference is expected.

Apart from adverbial adjuncts ambiguities, the IBH has been tested using a variety of global structural ambiguities, including PP (8), relative clauses (9), possessives (10), and conjunctions (11):

(8) Old men $_{\#1}$ and women $_{\#2}$ with very large houses
(9) I met the daughter $_{\#1}$ of the colonel $_{\#2}$ who was on the balcony
(10) The daughter $_{\#1}$ of the Pharaoh's $_{\#2}$ son
(11) Johnny $_{\#1}$ and Sharon's $_{\#2}$ in-laws

While their findings provided an overall support for the IBH, their results have been partly inconclusive or incompatible concerning its predictions. For example, in (8), the [IPh, IPh] pattern was statistically different from the [ip, ip], although both contain same-size boundaries, which should be equally uninformative. By contrast, no differences were found between [0, ip] and [ip, ip] patterns. Also, some of the strongest prosodic influences were found in the presence of a single boundary (e.g., [ip/IPh, 0]), rather than with two phrase-level boundaries of different phonological type. Finally, not all allowable permutations have been examined, thus preventing a direct comparison between the various patterns (e.g., [IPh, ip] vs. [ip, IPh]). Interestingly, while certain structures exhibited a very strong baseline preference (e.g., low-attachment for adjunct ambiguities; see Carlson et al., 2001), other structures, like (11), were more readily influenced by the prosodic manipulations. Importantly, unlike the other structures, here the disambiguation has been shown to be highly influenced by prosodic phrasing (Clifton et al., 2006; Lehiste, 1973; Streeter, 1978; Wagner, 2005; see Wagner & Watson, 2010 for review). The observation that the impact of prosody can vary considerably across syntactic ambiguities has been instrumental in designing the present study. We will address it in the following sections.

2.2. Weaknesses of the theories

2.2.1. Teasing the theories apart – predictions

Based on the core principles of both theories, it appears that the AAH and IBH each represents a contrasting view of prosodic boundary processing, which should be reflected by a different set of predictions. Nevertheless, based on the reviewed literature, both approaches make strikingly similar predictions (laid out in Table 1 below) about syntactic preferences of ambiguous structures, when two phrase-level boundaries (ip, IPh) are involved.

Table 1. Predictions made by AAH and IBH of syntactic preferences based on prosodic pattern

Prosodic pattern	*Prediction*	*AAH account*	*IBH account*
IPh > ip	Early closure/low attachment	IPh blocks competing parse	Relative boundary strength
ip < IPh	Late closure/high attachment	IPh blocks competing parse	Relative boundary strength
IPh = IPh	Baseline preference	2 IPh boundaries cancel each other out	Boundaries non-informative
ip = ip	Baseline preference	ip boundaries do not change syntactic bias	Boundaries non-informative

As seen in Table 1, when the main predictions are contrasted, no differences exist between the theories. As the current study focuses on the effect of two competing prosodic boundaries on comprehension, we will consider phrase-level boundaries only.

Note that the main difference between the IBH and AAH lies in the distinction between global and local prosodic processing. While the IBH argues that the listener evaluates the prosodic boundaries are informative for the sentence's structure based on their relative strength, the AAH argues that each boundary impacts the parser locally. This global vs. local dichotomy seems to imply that a global prosodic interpretation necessitates a "wait-and-see" strategy, whereby listeners make their decision of informativeness after all boundaries have unfolded, whereas local boundary integration is immediate. However, this would be a misconception, as the IBH in fact assumes incremental processing (Clifton, Carlson, & Frazier, personal communication). That is, each boundary is evaluated at the moment it unfolds, with each subsequent boundary being compared to the previous ones. Although the AAH also posits incremental processing, it does not specifically address the influence of more than one boundary on parsing (the same way the IBH does). Therefore, the locality assumption can be taken to suggest that multiple boundaries do not affect one another directly, but rather separately reduce or increase the likelihood of a local attachment. However, Watson and Gibson (2005) assume that two IPh boundaries should cancel each other's closure effects and result in a baseline preference. They also assume that the presence of a syntactically-compatible late boundary following a syntactically-incompatible early boundary (in the context of an unambiguous sentence; i.e., *John gave # the book # to Mary*) will improve the acceptability of the sentence (Watson & Gibson, 2004a). If processing were strictly local, then the acceptability of the sentence should not be ameliorated by the presence of another boundary. That is, if an obligatory attachment was prevented due to the presence of an early boundary (e.g., between a verb and its argument, as in: *John <u>gave</u> # <u>the book</u>...*), causing misunderstanding as a result, a later boundary should not invert it, as it carries information relevant only to a later attachment region (e.g., *...the book # to Mary*). However, if attachment decisions can be adjusted based on accumulating prosodic phrasing information, it implies that they are weighed against one another, in which case the AAH and the IBH make very similar assumptions about the nature of this process, with the IBH making more nuanced predictions, by considering the impact of smaller boundaries as well (as also supported by the literature; see Kjelgaard & Speer, 1999; Schafer, 1997, to name a few).

2.2.1.1. Phonological categories or a prosodic continuum?

One of the main reasons both theories make similar predictions for syntactic preference is their common reliance on the categorical view to determine boundary size – that is, the use of only two types of prosodic boundaries (ip and IPh) at the sentence level. Since the IBH argues that the relative difference between boundary sizes matters while the AAH argues for an effect of only IPh boundaries, if both theories assumed more than two phrase-level categories existed, teasing the theories apart would be simpler, as any two competing boundaries smaller than an IPh would be expected to affect parsing according to the IBH, but not according to the AAH.

The proponents of the categorical classification admit that despite being treated categorically, ip and IPh boundaries can be realized in different ways (Carlson et al., 2001). Nevertheless, they have also claimed that "syntactic structure can influence how an utterance is divided into ip's or IP's, but all boundaries of a given kind are equivalent and thus *continuous variation within a category plays no role in syntax-prosody interface*" (Snedeker & Casserly, 2010; p. 1239). To date, this line of reasoning hinges on the results of a single study by Carlson et al. (2001; Experiment 2), who tested whether the acoustic prominence rather than the phonological category of boundaries affected parsing preferences. They compared high attachment decisions for adjunct ambiguities (*Susie learned $_{\#1}$ that Bill telephoned $_{\#2}$ after John visited*), which were recorded with two different productions of ip boundaries. These ip boundaries differed only in terms of their durational properties (longer pre-final lengthening and pause duration), but not in their tonal properties, thus supposedly preserving the elements of the ip category. The so-called "long-ip" was placed in the late boundary position (#2, before the adjunct), where a relatively larger boundary is supposed to trigger high-attachment preference (see Table 1). The results showed no differences in high-attachment assignments (~35% in all conditions), suggesting that despite the (assumed) higher acoustic prominence of the long-ip relative to the "regular" ip, listeners still treated them as belonging to the same phonological category. However, an examination of the durational properties of both boundary types (the Fø values were identical) shows that the long-ip had a similar or shorter duration (both pre-final lengthening and silent interval) compared to almost all other ip boundaries used within Experiment 2 and across experiments (Experiments 3 and 4). The main exception was the "regular" ip boundary used as its comparison, which contained shorter pre-final lengthening (411 ms compared to a range of 496-549 ms) and break period relative to all ip boundaries at the same position (9 ms compared 269-373 ms). Therefore, it is highly likely that the long-ip boundary was treated as a regular ip boundary, not necessarily because it belonged to the same phonological category, but because it was acoustically similar to the other boundaries.

Other researchers have found evidence that challenges the purely categorical boundary perspective. In an important study, de Pijper and Sanderman (1994) asked naïve listeners to score boundary strength in naturally produced and delexicalized sentences on a scale from 1 to 10. They found that despite being untrained, listeners reliably recognized various levels of prosodic boundary strength, showing statistically significant agreement among them. They further analyzed the boundaries phonetically and compared the relationship of these acoustic properties to the perceived boundary sizes. They found that stronger boundaries were associated with tonal events (primarily pitch reset), but also pre-final lengthening and, to a lesser extent, pause duration.

A recent study by Millotte et al. (2008) indirectly provided evidence for within-category comprehension differences. The goal of the study was to test whether ip boundaries can constrain syntactic preferences in French. They presented listeners with sentences in which ambiguous homophones could either serve as a verb or as an adjective, as in (12):

(12a) Le petit chien mord la laisse qui le reticent (verb sentence)
[the little dog bites the leash that holds it back]

(12b) Le petit chien mort sera enterré demain (adjective sentence)
[the little dead dog will be buried tomorrow]

The sentences were phonetically indistinguishable until the word following the homophones. They were then recorded with cooperating prosodic boundaries, such that an ip boundary followed NP1 (*chien*) in (12a) and the adjective (*mort*) in (12b). Two versions of ip boundaries were recorded for each sentence: maximally and minimally informative, with the minimally informative ip having a flatter pitch contour as well as significantly shorter pre-final lengthening compared to the maximally informative ip. Participants listened to the beginnings of the sentences which were cut after the ambiguous homophones, and completed them in writing. Results showed not only a significant effect of ip boundaries on comprehension, but also that the boundary strength manipulation yielded significantly different syntactic preferences (verb/adjective), suggesting a graded effect of boundary type on syntactic processing. It might be claimed that the minimally-informative ip boundaries were created by altering both the tonal and durational properties, while Carlson et al. (2001) modified only the durational aspect of the boundaries. Nevertheless, both boundary sizes represented variations within the ip category, which should not, according to the categorical view, exhibit differentiated parsing decisions.

One of the most compelling findings for non-categorical boundary processing comes from a study by Wagner and Crivellaro (2010) who tested the parsing preferences of ambiguous sentences (e.g., *the tourist checked in the bags*), in which the boundaries were digitally manipulated in 6 increments. They found that subjects' responses were significantly influenced by the relative strength between the manipulated boundaries, suggesting that gradient quantitative, rather than categorical boundary size may be sufficient to explain the parsing decisions.

Finally, there exists a debate concerning even the theoretical foundation of prosodic constituency, which in turn affects the validity of the categorical entities used in ToBI. In her 1984 book (p.29), Selkirk raises the possibility that the definition of the ip category (or PhP, as defined here) is not as rigid as previously suggested:

> … language may exhibit more than one level of phonological phrase, in which case finer terminological distinctions can be made: PhP^1, PhP^2, . . . , PhP^n. With this terminology then, an intonational phrase is a special case of a phonological phrase, one that is associated with a characteristic tonal contour and that has an important function in representing the "information structure" of the sentence. The unit utterance, if it existed, would also be a phonological phrase in this sense.

Selkirk's (1984) suggestion that phrase-level prosodic constituents can be viewed as a continuum rather than as distinct pre-categorized levels touches on another theoretical question of how many levels exist in the prosodic hierarchy and how one determines their identity (for review see Wightman et al., 1992). While we do not attempt to answer this question here, it is highly relevant to the ToBI system, which was predominantly used to determine the strength of boundaries in studies that evaluated the IBH and AAH, as it hinges on the same theoretical background. In that sense, the evolution of ToBI also involved a debate as to how many levels should be considered and labeled, with the initial suggestion being seven levels (Price et al., 1991) eventually reduced to five (see Wightman, 2002). Furthermore, even when the number of labeled entities has been agreed upon, the way they are transcribed is also debatable. For example, in recent years, in order to maintain a high degree of agreement between the tonal events and break indices allegedly unique to each category, these two tiers have been linked together such that the characteristic features of each category can only be used together (e.g., break index of 4 and a boundary tone – marking an IPh). Wightman (2002; p.3) argued that "the restriction that certain break indices can be used only in combination with specific intonation labels is one of the most controversial aspects of the ToBI system. The linkage between the tiers further *de-emphasizes the perceptual experience of the listener*: The ToBI guidelines even suggest that, once either the tonal or phrasal labels have been produced by the listener, the redundant labels be inserted automatically to save time and increase inter-transcriber agreement". Such practice creates a circular argument for the identification of a specific category; that is, a category is determined based on a set of features that necessarily co-occur because they describe that category. If we consider Selkirk's (1984) claim, then perhaps more subtle differences exist within and between categories that might have been overlooked due to the nature of the labeling system. Crucial to the present study is the notion that while the categorical view is based on an ongoing debate, it should be regarded as a working hypothesis, not necessarily as an axiom.

2.2.1.2. Methodological concerns

As previously mentioned, all experiments used to test or validate the IBH and AAH employed the ToBI transcription system to determine boundary strength (or category). While ToBI provides general guidelines for the annotation of boundaries, it is not a precise measure. As a result, some acoustic properties of a given prosodic category also vary greatly across studies, such that the same values used to define the ip category in one study, were used to define the IPh category in another. For example, in Carlson et al. (2001), the ip boundaries in Experiment 2 contained a break ranging from 9-61ms, while in Experiments 3 and 4 (using the same sentences), they were considerably longer, ranging from 269-373 ms. By contrast, in Snedeker and Casserly (2010), the ip boundaries were characterized by a 30-100 ms break, while the IPh boundaries contained 100-300 ms breaks. That is, the definition of a category seems to be rather flexible and relative to each specific experiment, making the outcome non-uniform when comparing results across the board. One way to ensure prosodic uniformity across trials and experiments is to operationalize the creation of varying boundary sizes by using digital manipulation techniques (applied to an original speech file containing two naturally produced boundaries), as this type of manipulation allows for accurate control of the amount of acoustic variation (e.g., pre-final lengthening) assigned to each boundary.

Another concern regarding the outcomes of the studies examining the AAH and IBH lies in their exclusive use of a binary scoring system to determine syntactic preference, essentially requiring participants to make a forced choice between two given structures. One reason binary output is problematic is that it makes the rather simplistic assumption that preferring structure A to a certain degree X (e.g., 35% high-attachment) necessarily means accepting structure B to a (100%-X) degree (e.g., 65% low-attachment). In other words, it suggests that structural preferences are complementary *and* are, at the same time, a measure of structural acceptability. Only a handful of studies (Hwang & Schafer, 2006; Lee & Watson, Unpublished manuscript; Experiment 1) examined the ratings for each structure separately, and indeed reported more nuanced results. However, even then a binary paradigm was employed. Given the more subtle nature of the prosodic manipulations used in this line of research, and that more fine-grained ratings of prosodic phrasing have been reported using a scale (de Pijper & Sanderman, 1994; Sanderman & Collier, 1996; Wagner & Crivellaro, 2010), it is reasonable to assume that the outcome measures used thus far have the potential of masking a more intricate pattern of results that would otherwise allow a better understanding of the degree to which boundaries affect the comprehension of each construction, and a more accurate evaluation of the theories' predictions.

2.2.2. The current study

The current study aims to test the predictions of the IBH and AAH regarding the effect of two phrase-level prosodic boundaries on the resolution of syntactic ambiguities. To do so, we will: (1) employ an improved paradigm, which measures *within-structure parsing decisions*, (2) use *digital parametric manipulation* combined with a ToBI baseline (to ensure comparability by applying the same criteria as previous studies) to create the prosodic boundaries, in order to maintain uniformity across trials, and (3) *vary boundary strength in a gradient manner* in order to test the categorical aspect of prosodic boundaries. It is our assumption that previous outcomes were rather limited with respect to the degree to which prosody influenced syntactic preference because of the structures used (e.g., relative clauses have a strong low-attachment preference). While they did find significant effects, the small range of prosodic influence within these syntactic structures might have decreased the effect. If we intend to explore a range of boundary sizes, we must also ensure that the overall range of prosodic influence is greater. To achieve this, we will use structural ambiguities for which parsing decisions have been consistently shown to be heavily influenced by prosodic phrasing, namely EC/LC ambiguities. We will discuss the potential differences between these structures and previously used structures in the next section.

2.3. Experiment 1

Experiment 1 examines EC/LC sentence pairs containing 2 boundaries each, in 9 prosodic conditions (see Table 2), using an acceptability judgment task on a 7-point scale from least acceptable (i.e., a score of 0) to most acceptable (i.e., a score of 6). The task requires listeners to make quick, intuitive evaluations of the degree to which a sentence sounds acceptable to them. The rationale for using this measure was to allow subjects to express a finer degree of acceptability, which would usually be missed by forced-choice designs or a simple grammatical judgment paradigm, as they cannot distinguish between different rates of agreement (except, perhaps, in terms of response times, see Pauker et al., 2011). That is, 50% or 80% high-attachment or EC acceptance reflects only a ratio of general preference for one of two distinct options, but cannot reveal to what extent the structure itself was acceptable. With our procedure each sentence was given an individual score and, therefore, we were also able to (i) directly compare the scores for EC and LC sentences bearing the same prosodic pattern, and, crucially, (ii) evaluate scores within each structure (e.g., EC) separately. We believe this will allow us to reveal a more intricate pattern of results compared to the simple (high or low) attachment preference measures used thus far.

While global ambiguities have enabled an elegant design in which prosodic phrasing alone was manipulated while the lexical content remained the same – allowing an examination of its impact in isolation – it has been demonstrated that its impact is rather limited as it cannot override the effect of other, much stronger, lexical and syntactic factors in these structures (e.g., high/low attachment preferences). By comparison, although many factors influence EC/LC structure interpretation and compete with one another, including a general/initial LC preference (see sections 1.3 and 1.4), transitivity bias of the subordinate clause verb (e.g., *parked* in example [1]) (Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Pickering, Traxler, & Crocker, 2000), as well as aspect (past tense vs. progressive) (Frazier, Carminati, Cook, Majewski, & Rayner, 2006), prosodic phrasing seems to be the most salient, as it is able to override them (Itzhak et al., 2010). In global ambiguities, phrasing can only affect the ease of attachment to a specific location (high or low), but the final decision does not affect the grammaticality of the sentence. In EC/LC ambiguities, the location of prosodic boundaries in the utterance greatly influences parsing, as misleading phrasing can lead to ungrammatical structures (Hwang & Schafer, 2006; Kjelgaard & Speer, 1999; Schafer et al., 2000; Speer et al., 1996; Warren et al., 1995, among many others). Moreover, online (ERP) studies have demonstrated that prosodic phrasing strongly guides the syntactic preference of EC/LC sentences from a very early stage (Bögels et al., 2013; Itzhak et al., 2010; Pauker et al., 2011; Steinhauer et al., 1999). A recent ERP study that examined the combined influence of lexical (transitivity) bias and prosodic phrasing on EC/LC ambiguity comprehension revealed that transitivity bias was completely obliterated in the presence of prosodic boundaries favoring the competing parse, and appeared only in their absence (Itzhak et al., 2010). Finally, it has also been demonstrated that once a boundary has been established in EC/LC sentences, it is difficult (if not impossible) to mentally delete, in order to later repair the sentence (the Boundary Deletion Hypothesis; Pauker et al., 2011; Steinhauer & Friederici, 2001), whereas this would not be the case in global ambiguities because both parses are correct.

Another reason for favoring EC/LC structures, from a methodological point of view, is that they allow listeners a clear indication of whether their initial parsing preference was correct. In the previous studies using global ambiguities, a common technique to probe which parse was favored by listeners has been to wait until the sentence was fully heard and then present a forced-choice matching task, which could only be made after additional disambiguating aids [pictures/props in the visual world paradigm (Snedeker & Trueswell, 2003; Snedeker & Yuan, 2008), or the rephrasing of the sentences (Carlson et al., 2009; Clifton et al., 2002, 2006)] have been presented. On the other hand, prosodic manipulations (e.g., conflicting prosody) in EC/LC ambiguities immediately affect listeners' brain responses, who can then make their judgment upon sentence termination, without the need to indirectly examine how the sentence was interpreted, potentially introducing additional effects into the decision process. Since prosody has a strong impact on the acceptance of these structures, it should allow a larger range of influence (compared to the limited one seen in global ambiguities), which is necessary to examine a finer array of manipulations (a continuum of boundaries rather than two sizes).

As discussed earlier (section 1.6.3.1), when two boundaries are involved, both theories largely make the same predictions (see Table 1). One way to resolve this problem is by introducing a third type of boundary, whose size is smaller than an IPh but larger than an ip (i.e., mid-range boundary). According to the AAH, only IPh boundaries can trigger closure, while a boundary of a smaller size cannot. According to the IBH, the relative difference between boundary sizes should affect parsing. In this case, the predictions laid out in Table 1 could be extended to the ones made in Table 2 below.

As illustrated in Table 2, using a mid-range boundary can potentially reveal differences between the theories in conditions where an ip boundary follows a mid-range boundary and vice versa. In the first case [mid/ip], the IBH would predict EC preference because the first boundary is relatively larger than the second one, and in the second case [ip/mid], the IBH would predict LC preference because the second boundary is relatively larger than the first one. On the other hand, the AAH would predict the baseline preference in both cases, as both boundaries are smaller than an IPh and, therefore, should not trigger closure.

Table 2. Predictions made by AAH and IBH for syntactic preferences of prosodic patterns created using 3 boundary sizes (ip, mid-range, IPh), assuming the mid-range boundary size is distinct from the ip and IPh categories

Prosodic pattern	*AAH predictions*	*IBH predictions*
[IPh / ip]	Early closure	Early closure
[IPh / mid]	Early closure	Early closure
[IPh / IPh]	Baseline preference	Baseline preference
[mid / ip]	Baseline preference	Early closure
[mid / mid]	Baseline preference	Baseline preference
[mid / IPh]	Late closure	Late closure
[ip / ip]	Baseline preference	Baseline preference
[ip / mid]	Baseline preference	Late closure
[ip / IPh]	Late closure	Late closure

Note: mid = mid-range.

A potential problem with this suggestion, from the standpoint of the two theories, is that ip and IPh are the only categories allowed by ToBI to describe prosody at the phrase level. As noted by Carlson et al. (2001; p.67): "substantial variation in the physical signal is not sufficient to affect interpretation in the absence of variation in the phonological category of prosodic boundaries". It could be claimed that a mid-range boundary would only represent a variation within one of the categories and, therefore, yield indistinguishable effects. However, Carlson et al.'s (2001) claim was made in the context of their failure to find differences between long and short ip's (Experiment 2), which we have argued earlier (section 1.6.3.2) was probably due to a methodological shortcoming. By comparison, Millotte et al. (2008) also changed their physical signals substantially (minimal vs. maximally informative ip boundaries) and did find significant differences in subjects' responses, as did Wagner and Crivallero (2010). In light of these findings, using highly controlled materials as well as a task allowing for a more fine-grained outcome could potentially allow us to reveal differences between these boundary sizes.

Theoretically, by expanding the categorical boundary strength approach, both hypotheses should still be able to account for more than two categories of prosodic boundaries, if the endpoints of the phrase-level categorical division (ip and IPh) are kept the same. Therefore, the basic weak and strong boundaries were designed to fit the ip and IPh boundaries' criteria based on ToBI and used by the majority of studies testing these theories (Beckman & Elam, 1997; Beckman & Hirschberg, 1994; Silverman et al., 1992), while the mid-range boundary was intended to be at a perceptually neutral position between them. However, since no strict rules exist for determining the categories, and their definition is relative to one another in each experiment, we suggest that given a defined and agreed-upon ip category (based on ToBI categorization), an IPh category can be created by extending the duration of the pre-final lengthening and pause interval. A boundary size in between these two endpoints would be dubbed "mid-range". Creating all boundary sizes of a given utterance using the same speech file has an advantage in that it allows us to control for any acoustic variability among the conditions and evaluate the effect of the prosodic manipulation alone. Although currently there is no means to define a point on the range (or the range itself) between the two major categories, if a continuum of boundary sizes does exist, then such a "mid-range" boundary should be treated differently from the other two categories, as illustrated in Table 2. If, however, no continuum exists, the "mid-range" boundary would represent a variant of one of the two categories (Table 3, columns 1 or 2). A third possibility is that using an ip boundary as a baseline for the creation of the two other boundaries might only yield variants within the ip category. If listeners are able to distinguish among these within-category variations, the predictions made by the IBH and AAH should differ (Table 3, columns 3 and 4); if not, both hypotheses would predict the baseline preference for all conditions of both structures. Note that by assuming listeners are oblivious to the difference between or within categories, both hypotheses make the same predictions. Importantly, even if they adopted the notion of continuum (either within or between categories), the IBH and AAH would still make an either-or prediction for syntactic preference, relative to the baseline. That is, currently, they both assume similar acceptance ratings for conditions that direct listeners to favor one of the two parsing options, such that conditions [IPh/ip] and [IPh/mid], for example, should be similarly accepted as EC, as opposed to differ in a gradient manner. Thus, the relative difference between the ip and mid-range boundaries in this example would not be reflected in the acceptability rating.

Table 3. Predictions made by AAH and IBH for syntactic preferences based on prosodic pattern of 3 boundary sizes (ip, mid-range, IPh), as a function of "mid-range" boundary size.

Prosodic pattern	*1. Mid-range = ip (AAH + IBH)*	*2. Mid-range = IPh (AAH + IBH)*	*3. All boundaries = ip (IBH - continuum)*	*4. All boundaries = ip (AAH - continuum)*
[IPh / ip]	Early closure	Early closure	Early closure	Baseline preference
[IPh / mid]	Early closure	Baseline preference	Early closure	Baseline preference
[IPh / IPh]	Baseline preference	Baseline preference	Baseline preference	Baseline preference
[mid / ip]	Baseline preference	Early closure	Early closure	Baseline preference
[mid / mid]	Baseline preference	Baseline preference	Baseline preference	Baseline preference
[mid / IPh]	Late closure	Baseline preference	Late closure	Baseline preference
[ip / ip]	Baseline preference	Baseline preference	Baseline preference	Baseline preference
[ip / mid]	Baseline preference	Late closure	Late closure	Baseline preference
[ip / IPh]	Late closure	Late closure	Late closure	Baseline preference

Note: mid = mid-range.

Note that the baseline preference in Tables 2 and 3 is not specified. An overall expectation that LC should be the baseline preference (in spoken language) comes from the findings of Kjelgaard and Speer (1999), in particular, who found EC sentences were more difficult to process in the baseline condition, supposedly reflecting a violation of a default LC expectancy, similar to reading (see section 1.3). However, two interesting observations regarding the acoustic waveforms and durational measures of the boundaries in their study may shed more light on this preference. First, despite attempting to conceal boundary location from listeners, it appears that the baseline conditions contained two ip boundaries, each with an (L-) phrase accent. Second, these boundaries contained the same pitch contour as the boundaries in the cooperating conditions, and differed only in terms of duration (shorter pre-final lengthening and silence interval). Thus, it is reasonable to assume that listeners did detect these boundaries, although they were more subtle than the ones used in the cooperating condition. Importantly, the difference in the duration of the pre-final lengthening (a factor that has been consistently shown to signal boundary location; Klatt, 1975; Lehiste, 1973) between the early and late boundaries in both EC and LC baseline conditions mirrored that of the cooperating LC condition. That is, similar to the LC condition, which was characterized with longer pre-final lengthening on the NP (compared to the EC conditions, in which longer pre-final lengthening were found on the verb), in the baseline conditions, the pre-final lengthening of the early boundary was shorter than the pre-final lengthening of the late boundary. Kjelgaard and Speer (1999) claim that the baseline conditions were designed to be perceptually acceptable as both EC and LC, but the boundaries are clearly uneven. As a result, it can be argued that the fact EC was more difficult to process compared to LC in the baseline conditions can be attributed to this durational pattern. We suspect that using quantitatively similar boundaries in both positions (either ip or IPh) may result in a different pattern.

Finally, an issue that should be taken into account is the difference in the location of prosodic boundaries between EC/LC and global ambiguities. In global ambiguities, listeners heard both boundaries before the ambiguous phrase unfolded. In EC/LC, the boundaries flank the ambiguous NP. Since the ambiguity here revolves around whether the verb takes an argument or not, the presence of the early boundary may bias listeners to prefer an EC interpretation. On the other hand, based on the IBH and AAH, any relative difference between earlier- and later-appearing boundaries should also be taken into account, as no structural restrictions are made (on the contrary, Schafer [1997] attempts to explain EC/LC preference using the prosodic visibility hypothesis, which is the basis of the IBH; see section 1.6.1.1).

2.4. Methods

2.4.1. Participants

Twelve undergraduate students from McGill University (7 women, age range = 18–25 years) were recruited by advertisement and paid for their participation. All were right-handed (Edinburgh Handedness Inventory; Oldfield, 1971) native speakers of English with no known history of hearing impairment or brain injury. Prior to their participation, each subject signed a written informed consent.

2.4.2. Materials

Eighteen experimental conditions were created in the following steps: (i) Sentence structure design, (ii) selection of of pitch contour for the prosodic boundaries, (iii) sentence recording, and (iv) prosodic boundaries manipulation.

2.4.3. Sentence structure design

First, 45 EC and LC pairs in English were created based on sentences used in our previous experiment (Pauker et al., 2011). All sentences were constructed using the same syntactic skeleton to ensure syntactic uniformity:

EC/LC onset (shared content): [Whenever] [the noun1] [was verb1-ing #] [the noun2 #]...
(EC – disambiguating point): …[*would verb2*]
(LC– disambiguating point): …[*the noun3*] [*would verb2*]

The first VP in each sentence was optionally transitive (e.g., *was approaching*), thus compatible with both EC and LC structures. The transitivity count for all verbs (British National

Corpus; see Lapata, Keller, Schulte, & Walde, 2001) revealed an overall transitive bias. To avoid having structural preferences made due to this bias, we used the progressive aspect of the verb (rather than the simple aspect [e.g., *approach*]), which has previously been shown to reduce the transitivity bias while keeping the garden-path effects intact (Frazier, Carminati, et al., 2006).

The second VP was either presented without added content (e.g, *would <u>wait</u>*), or, if necessary, was followed by a short semantically and pragmatically-appropriate continuation (e.g., *would <u>close</u> early*). The second verb was always preceded by the modal verb *would*, which was irrelevant to the creation of the EC sentences. As we intended to derive the EC versions of the sentences from the original LC sentences by splicing out NP3, it was necessary to control for acoustic variability that could affect this procedure. While NP3 followed a boundary and began with a fricative (*the*), which allowed its removal without distorting the preceding word, it was also necessary to ensure the following word would not be coarticulated with it. The word *would* helped to separate NP3 from the second VP, as it was easier to identify and segment from the preceding and following words.

2.4.4. Boundaries/recording

Each sentence was designed to contain two boundaries (i.e., B1 and B2 in Table 7) – one at each closure site: an early boundary following VP1 (e.g., *was approaching*) and a late boundary following NP2 (e.g., *the people*). Each boundary was assigned the same prosodic contour. In order to ensure the prosodic structure remained highly acceptable to listeners in both EC and LC versions, while also maintaining the ambiguity of the sentences until the disambiguation region, we chose a H*-L- contour for our ip boundaries, similar to one used by Kjelgaard and Speer (1999) to create their “baseline” condition. This particular contour was instrumental in later creating natural-sounding IPh-compatible boundaries characterized with a H*L-L% contour, also used by Kjelgaard and Speer (1999) in their "cooperating" condition. We will explain the boundary manipulation procedure in detail below.

The 45 EC and LC pairs were recorded by a female native English speaker in a sound-attenuating booth (Praat speech software Boersma & Weenink, 1996; 44.1 kHz sampling rate, 16-bit amplitude resolution [Logitech H390 USB Headset]). Each sentence was produced with two ip boundaries – after VP1 and NP2 – based on the pre-determined prosodic contour. To reliably mark juncture, each boundary was also followed by a pitch reset, such that the word immediately following the boundary started at the same level of $F_ø$ and did not present a downstep in $F_ø$ (Swerts, 1997), which has been claimed to indicate that the upcoming words are nested within the same phrase (D. R. Ladd, 1988; also see Wagner, 2005). The sentences were then evaluated by a ToBI expert. Any sentence that did not contain the intended prosodic contour and/or pitch reset, was re-recorded. This process was repeated until all sentences fit the prosodic criteria.

2.4.5. Prosodic manipulations

2.4.5.1. Step 1 – preparation

To create the EC/LC sentence pairs such that each would be identical before the disambiguating region, we spliced out NP3 from the LC versions to create the corresponding EC versions. In cases where the coarticulation between *would* and the spliced NP3 was still audible, we either replaced it with *would* spliced out from the corresponding EC version, or with VP2 in its entirety (would + VP2) from the EC version. Following this procedure, five native English speakers listened to the sentences and rated them for naturalness and for any audible acoustic artifacts (e.g., energy spikes, uneven loudness, noise). Based on their judgments, the sentences were digitally modified using Adobe Audition version 1.5, or cross-spliced again, and then re-evaluated and fixed again if necessary. Next, using Praat speech software, all major constituents: *Whenever*, *NP1*, *VP1*, *NP2*, *NP3* (in LC), *Would*, and *VP2* (see Table 4), were annotated in one tier. In a second tier, we marked the desired onset and offset of the pre-final lengthening sites in VP1 and NP2 (onset – the vowel of the last syllable's rime; offset – word offset), as well as the onset location for the pause interval, based on spectrogram and sound file examination.

Table 4. Sample stimuli for the nine experimental conditions

Condition	Sentence	Disambiguating region	
		EC	LC
	Conj NP1 VP1 B1 NP2 B2	*Would VP2*	*NP3 Would VP2*
[1_1]	Whenever the bear was approaching / the people /	…would run away	…the dogs would run away
[1_2]	Whenever the bear was approaching / the people //	…would run away	…the dogs would run away
[1_3]	Whenever the bear was approaching / the people ///	…would run away	…the dogs would run away
[2_1]	Whenever the bear was approaching // the people /	…would run away	…the dogs would run away
[2_2]	Whenever the bear was approaching // the people //	…would run away	…the dogs would run away
[2_3]	Whenever the bear was approaching // the people ///	…would run away	…the dogs would run away
[3_1]	Whenever the bear was approaching /// the people /	…would run away	…the dogs would run away
[3_2]	Whenever the bear was approaching /// the people //	…would run away	…the dogs would run away
[3_3]	Whenever the bear was approaching /// the people///	…would run away	…the dogs would run away

Notes: 1) In each condition, the first number corresponds to the size of the early boundary and the second number corresponds to the size of the late boundary; 2) "/" = original /ip/ boundary; 3) "//" = manipulated mid-range boundary; 4) "///" = manipulated /IPh/-compatible boundary.

2.4.5.2. Step 2 – acoustic manipulation

Eight new conditions (for each syntax) were created by manipulating the duration of the annotated portions of VP1 and NP2 (the second tier only) using PSOLA resynthesis in Praat. The total amount of added duration at each boundary was partitioned between the lengthening of the pre-boundary word (pre-final lengthening) and the pause interval, at a ratio of 25% lengthening and 75% pause duration, which had been found to be the most natural-sounding according to three native English listeners. Then, several versions of IPh-compatible and mid-range boundaries were created and presented to native English speakers. Based on their judgments, the IPh-compatible boundary was created by adding a total of 320 ms to the original boundary (80 ms lengthening; 240 ms pause interval), and the mid-range boundary was created by adding a total of 80 ms to the original boundary (20 ms lengthening; 60 ms pause interval). The Praat script created all the possible permutations of boundary size combinations (3 × 3) for each sentence (see Table 6), resulting in 9 versions per structure, or 18 overall and a total of [18 × 45 (verbs) =] 810 sentences. The manipulated sentences were then individually perceptually evaluated to determine whether they were acoustically natural-sounding and acceptable. Any sentence that presented an acoustic anomaly was re-annotated and manipulated again. This process was repeated until all sentences met these standards. An example of the prosodic manipulation of boundary size is displayed in Figure 1.

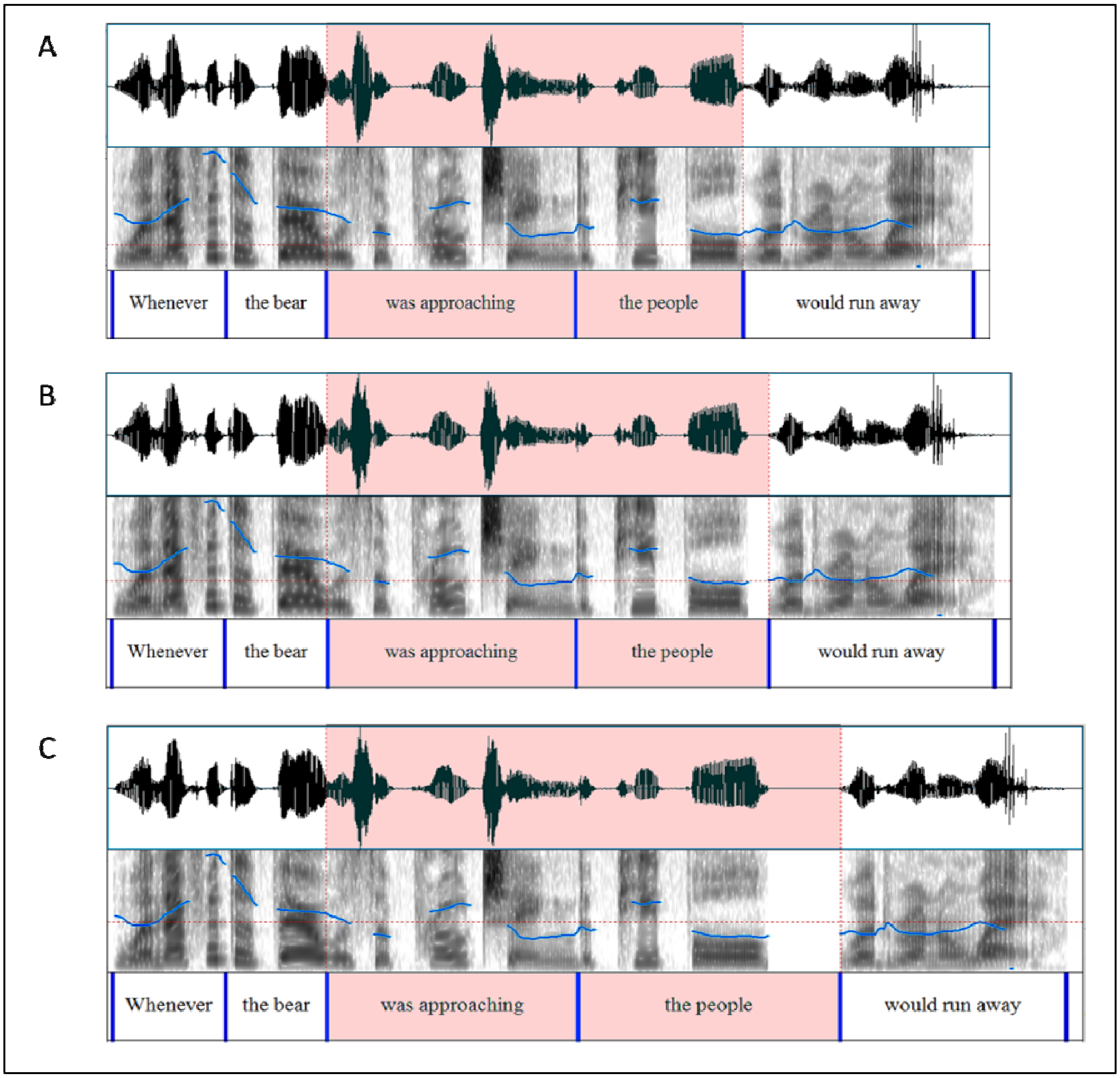


Figure 1. Waveforms and pitch contours of sample sentences in which boundary 2 was manipulated while boundary 1 remained the same. (A): condition [1_1], containing the two original ip boundaries. (B) condition [1_2], containing the original ip boundary followed by the manipulated mid-range boundary. (C): condition [1_3], containing the original ip boundary followed by the manipulated IPh-compatible boundary.

Mean duration and $F_ø$ measurements of selected regions of the original sentences (containing two ip boundaries) are presented in Tables 5 and 6, respectively. Note that both EC and LC versions of each sentence are identical prior to the disambiguating region. Since all prosodic manipulations were conducted on the shared portion of each EC/LC sentence pair, they are, in effect, indistinguishable.

Table 5. Mean durations, in ms, of the major constituents in the 42 EC/LC experimental sentences

Whenever	*NP1* (*the bear*)	*V1* (*was approaching*)	*NP2* (*the people*)		*NP3* (*the dogs*)	*would*	*V2* (*run away*)
350 (32)	*480 (84)*	*702 (60)*	*553 (59)*	***EC***		*155 (34)*	*802 (194)*
				LC	*512 (88)*	*135 (30)*	*795 (197)*

Note. Standard deviations are indicated in parentheses.

Table 6. Mean Fø Maxima and Minima, in Hz, of VP1 and NP2 in the 42 EC/LC experimental sentences

	V1 $F_ø$ targets			*NP2 $F_ø$ targets*		
	Mean	*Max*	*Min*	*Mean*	*Max*	*Min*
EC/LC	*243 (9)*	*313 (17)*	*204 (9)*	*246 (13)*	*310 (16)*	*201 (14)*

Note. Standard deviations are indicated in parentheses.

2.4.6. Pseudo-randomizing

To allow a feasible testing time and to avoid participants' fatigue, six experimental lists were created that contained a sample of sentences from the full list. First, 3 original lists, comprised of one third of the total number of 42 sentences (3 sentences out of the original 45 were designated for the practice block) were created. Each list contained 252 sentences (14 items per condition × 18 conditions) and was divided into 4 blocks of 63 sentences each. The lists were pseudo-randomized based on five criteria: (i) degree of predicted prosodic anomaly, (ii) syntactic structure, (iii) semantic content (e.g., animals, sports, professions), (iv) prosodic structure, and (v) sentence length. The randomization rules used to create the experimental lists are displayed in Table 7 below. To control for order of appearance, 3 additional lists (mirror images) were created by reversing both the block and the sentence order of each original list.

Table 7. Pseudo-randomization criteria and rules used to create the 3 original experimental lists

Criteria (starting with highest priority)	Variables	Randomizing rule
1. Degree of prosodic anomaly	(i) high; (ii) intermediate; (iii) low	No more than 3 repetitions of the same level in a row
2. Syntactic structure	(i) EC; (ii) LC	No more than 3 repetitions of the same structure in a row
3. Semantic structure	8 semantic fields (e.g., sports)	2 sentences of the same field separated by at least 2 sentence of other fields
4. Prosodic structure	9 prosodic conditions (3 × 3)	No more than 2 identical prosodic conditions in a row
5. Sentence length	(i) short; (ii) long	No more than 3 short or 3 long sentences in a row

2.4.7. Procedure

Participants sat in a comfortable chair approximately 80 cm in front of a computer monitor and listened to spoken sentences presented binaurally via insert-phones (Etymotic Research Inc., Elk Grove Village, IL). Subjects were instructed to press one of 7 marked keyboard keys, each representing a degree of acceptability on a continuum between “most acceptable” and “least acceptable”, to indicate the degree of acceptability of each presented sentence (acceptability judgment task). Each trial began when a visual cue (i.e., “+”) appeared on the screen 1500 ms prior to sentence presentation, remaining visible until the end of the sentence. Following sentence termination, a response prompt appeared on the screen (i.e., “Please rate!”). At the beginning of each session, participants were given a short practice block (i.e., 5 sentences), after which further clarifications were made by the experimenter, if necessary.

2.4.8. Data Analysis

Acceptability ratings were computed as the average acceptability score given to each condition separately. Data were subjected to repeated measures ANOVA with the factors *Syntax (2)* × *Early Boundary (3)* × *Late Boundary (3)*.

2.5. Results

2.5.1. Main effects

Recall that most previous studies did not allow independent acceptability rates for each structure, and, therefore, were able to present only the acceptability proportions for each structure. Based on this type of measure, the AAH and IBH have made the (implicit) assumption that an increase in one structure's preference necessarily leads to the proportional decrease of the other structure's preference (e.g., 60% high attachment preference = 40% low attachment preference). In principle, this logic should also hold for a paradigm allowing acceptability scores for each structure independently; namely, the theories would predict that the scoring pattern observed thus far would be maintained. Given that our design was completely symmetrical in both structures, the IBH and AAH should not predict main effects of Syntax, Early Boundary, or Late Boundary, because any differences in one condition/structure should be canceled out by the complementary condition/structure when the scores are summed up (e.g., 70% preference for EC when the early boundary is 4 means 30% preference for LC when the early boundary is 4. When averaged, the result is 50%). Even if a specific structure should be preferred over the other, the proportional counter-score for the other structure should prevent any main effect from emerging.

Contrary to this prediction, we found main effects of Syntax [$F(1,11) = 9.2$, $p<.02$], Early Boundary [$F(2,22) = 7.5$, $p<.02$] and Late Boundary [$F(2,22) = 5.5$, $p<.04$]. The main effect of Syntax reflected an overall EC preference. This effect is illustrated by the total average for each syntactic structure, presented in the Grand Means column in Table 8. Figure 2, showing the difference between the average acceptability scores of EC and LC (EC minus LC) in each condition, clearly demonstrates that EC was the preferred structure across all conditions, as each bar has a positive value. Interestingly, EC was preferred over LC even in those conditions whose pattern of prosodic phrasing was predicted to be most compatible with LC (i.e., [1_3] and [1_2]). In addition, when the early boundary is held constant and the late boundary increases in size, the difference between EC and LC gradually decreases (black arrows). For example, when the early boundary size is 3, the difference between the acceptability scores of EC and LC gradually diminish as the late boundary becomes larger, in the following manner: [3_1] > [3_2] > [3_3]. This pattern was consistent at all boundary levels. At the same time, when the late boundary size is held constant and the early boundary decreases in size, the difference between the acceptability scores of EC and LC gradually decreases (red arrow), in the following manner: [3_1] > [2_1] > [1_1]. However, based solely on this data, it is still unclear what drives this effect – for example, whether lower bars are the result of an increase in LC acceptability, a decrease in EC acceptability, or both.

Table 8. Averages and corresponding percentage of conditions acceptability in EC and LC

Condition	*3_1*	*3_2*	*3_3*	*2_1*	*2_2*	*2_3*	*1_1*	*1_2*	*1_3*	*Grand Mean*
EC	4.04	3.92	3.24	4.16	3.92	3.21	4.13	3.79	3.23	3.74
% EC (EC/6)	67	65	54	69	65	54	69	63	54	
LC	2.13	2.44	2.39	2.67	2.74	2.93	3.17	3.10	3.15	2.75
% LC (LC/6)	36	41	40	45	46	49	53	52	53	
EC+LC	6.17	6.36	5.63	6.83	6.66	6.14	7.30	6.89	6.38	
% EC EC/(EC+LC)	65	62	58	61	59	52	57	55	51	

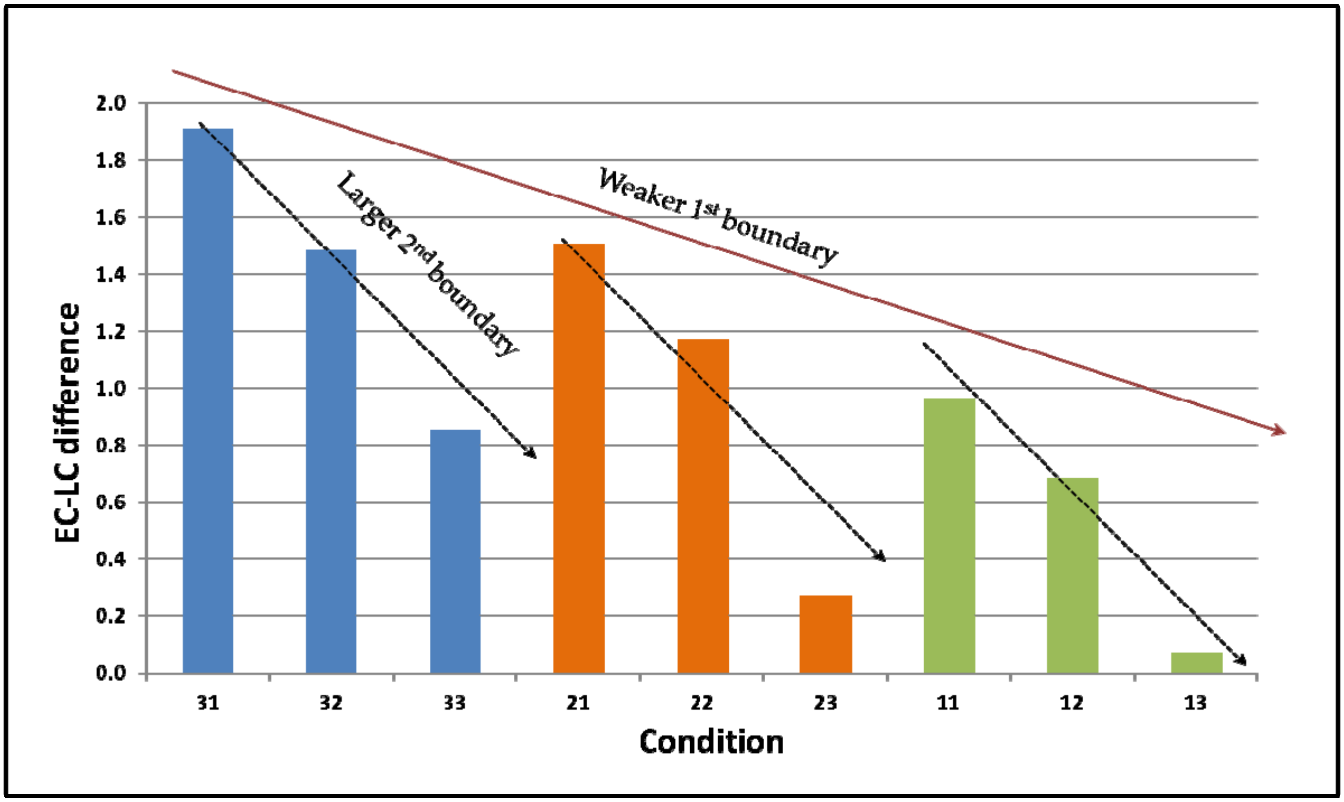


Figure 2. Difference in acceptability between EC and LC structures as a function of prosodic boundary size. X-axis: Displays the prosodic pattern for each condition (e.g., **3_1**: Boundary #1 = IPh, Boundary #2 = ip; **2_2**: Boundary #1 = intermediate, Boundary #2 = intermediate; **1_3**: Boundary #1 = ip, Boundary #2 = IPh). Y-axis: EC acceptability minus LC acceptability for each prosodic structure.

As discussed, in previous studies, in order to estimate the difference between prosodic patterns, the percentage of one structure's preference (e.g., high-attachment) would be computed and compared across conditions. In Table 8, we present the average scores and corresponding proportions for each condition in each structure (top four rows), as well as a computation more closely matching the one available in previous studies, in the two bottom rows (i.e., proportions of EC acceptability out of the total average score in each condition, without taking factor Syntax into account). Comparing the actual and estimated proportions for EC shows that the latter measure is less sensitive than the former as it does not accurately depict the actual preference for EC (or the corresponding preference for LC) within each condition. This suggests that in a study only allowing a binary choice, the differences between the resulting proportions (and the actual internal preferences) would entirely miss the actual acceptability of the sentences.

To illustrate the main effects of Early Boundary and Late Boundary, as computed by the ANOVA, we used the total acceptability scores for each condition (collapsed across EC and LC, as presented in the fifth row in Table 8. See also section 2.5.2 below, and corresponding Figures 4 and 5) to create a score matrix, presented in Table 9. Columns represent the effect of Early Boundary manipulation on acceptability, with Late Boundary held constant; rows represent the effect of Late Boundary manipulation on acceptability, with Early Boundary held constant. The emerging pattern of results shows that as boundary size decreases (in all directions: vertically, horizontally, diagonally), the score increases, for both Early Boundary and Late Boundary factors. This is illustrated in the table by the downward (blue) arrow (Early Boundary manipulation) and leftward (green) arrow (Late Boundary manipulation). Each cell in those directions contains a score larger than the previous one (the only exception is condition [3_1], which is slightly smaller than [3_2]).

Table 9. Total acceptability scores for each prosodic condition (EC+LC) as a function of boundary position.

		Late Boundary		
		1	2	3
Early Boundary	3	6.17	6.36	5.63
	2	6.83	6.66	6.14
	1	7.30	6.89	6.38

To further examine how manipulating each boundary location influenced acceptability, we conducted follow-up analyses for each factor. For the factor Early Boundary we found that boundary size 3 (IPh) acceptability was significantly smaller than both boundary sizes 1 ($p<.02$) and 2 ($p<.02$), whereas the mean acceptability of boundary size 1 (ip) was larger than that of boundary size 2 (mid-range), but showed only a trend ($p=.061$). Similarly, the pairwise comparisons between boundary levels for the factor Late Boundary revealed that acceptability for boundary size 3 was significantly smaller than for boundary sizes 1 ($p<.05$) and 2 ($p<.03$), while boundaries 1 and 2 did not differ significantly. As shown in Table 9, the direction of the effect for both Early Boundary and Late Boundary factors was negative: As boundary size increased, the overall acceptability decreased in a graded manner (1 > 2 > 3).

2.5.2. Interactions

Unlike previous studies, our design allowed us to test for potential interactions between syntax and boundary position, in order to examine the effect of boundary manipulations on the respective acceptability of each structure (EC and LC), separately. If designed similarly (even with only the two “official” phrase-level boundaries – ip and IPh), both theories would have expected significant two-way interactions between Syntax and Early Boundary and Syntax and Late Boundary. Specifically, they would have predicted significant main effects of both Early Boundary and Late Boundary for each syntax type – going in opposite directions – the reason being that the boundaries supporting each structure (Early Boundary in EC and Late Boundary in LC) should increase the acceptability for that structure while decreasing the acceptability of the competing structure, to the same extent. The hypotheses, therefore, would assume a completely symmetrical pattern of results, as illustrated in Figure 3. As this pattern should have been sufficient to explain the results, a 3-way interaction would not have been expected.

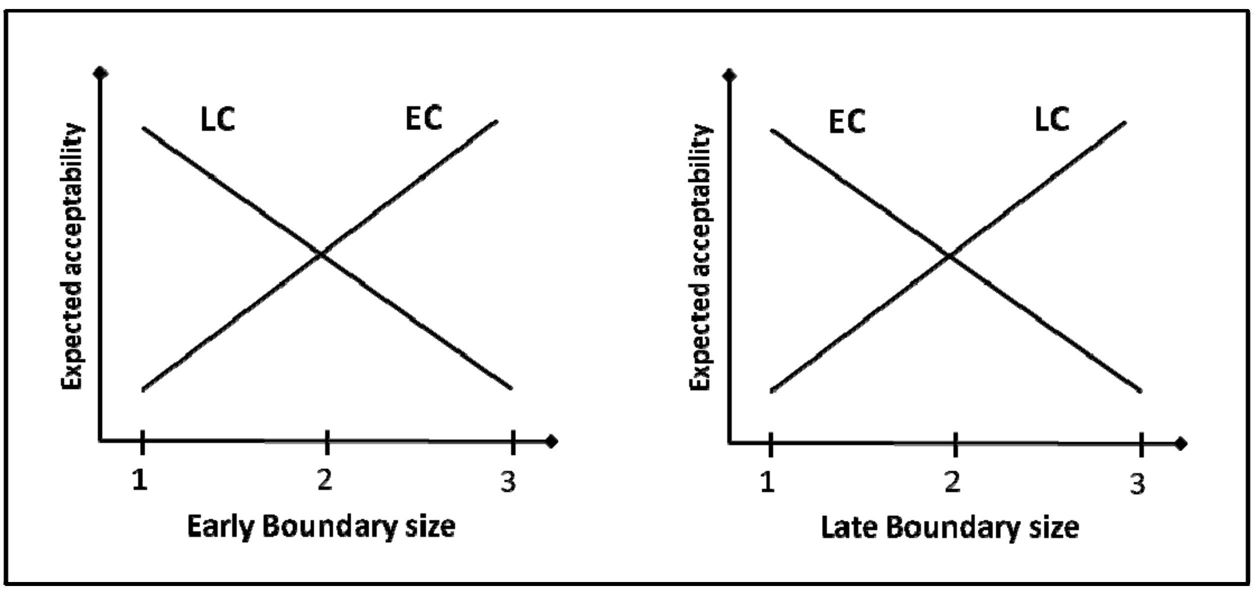


Figure 3. The expected acceptability pattern for EC and LC as a function of boundary 1 and boundary 2 size.

In line with these predictions, we found significant interactions of Syntax × Early Boundary [$F(2,22) = 16.14$, $p<.01$] and Syntax × Late Boundary [$F(2,22) = 12.57$, $p<.01$], as well as a non-significant Syntax × Early Boundary × Late Boundary [$F<1$]) interaction. However, the follow-up analyses revealed a completely unexpected pattern. In LC, we found a significant main effect of Early Boundary [$F(2,22) = 11.8$, $p<.01$], but not of Late Boundary ($p>.1$), and in EC, we found the opposite outcome: a significant main effect of Late Boundary [$F(2,22) = 9.77$, $p<.01$), but not of Early Boundary ($p>.7$). This unique pattern of scores is different from the prediction of both theories in several ways. First, the early boundary size affected only the acceptability of LC and the late boundary size affected only the acceptability of EC, and not as expected. Second, instead of mirroring the acceptability pattern within each structure (as illustrated in Figure 3), the size of the "supporting" boundary (early boundary in EC) in both structures had no effect on acceptability whatsoever. Third, the pairwise comparisons showed that the acceptability of each boundary size was different from the other in a manner reflecting the parametric prosodic manipulations in both structures similarly. Recall that according to both hypotheses, no differences should have been found between boundary size 2 (mid-range) and either boundary size 1 or 3 (based on their assumption it should belong to one of the two categories – ip or IPh). Instead, in LC, all boundary sizes were not only significantly different from one another [1 vs. 2 ($p<.03$), 1 vs. 3 ($p<.01$), 2 vs. 3 ($p<.01$)], they were also incrementally different: 1 > 2 > 3. That is, the acceptability for LC decreased with each increasing Early Boundary size. In EC, acceptability for boundary sizes 1 vs. 3 ($p<.01$) and 2 vs. 3 ($p<.01$) was significantly different, and acceptability for boundary sizes 1 vs. 2 ($p=.072$) showed a trend in the same direction. Here, too, we found a graded effect of acceptability, similar to the one found in LC, of decreasing acceptability with increasing Late Boundary size.

The direction of the effects is illustrated in Figure 4, and the acceptability scores within EC and LC are illustrated in Figure 5.

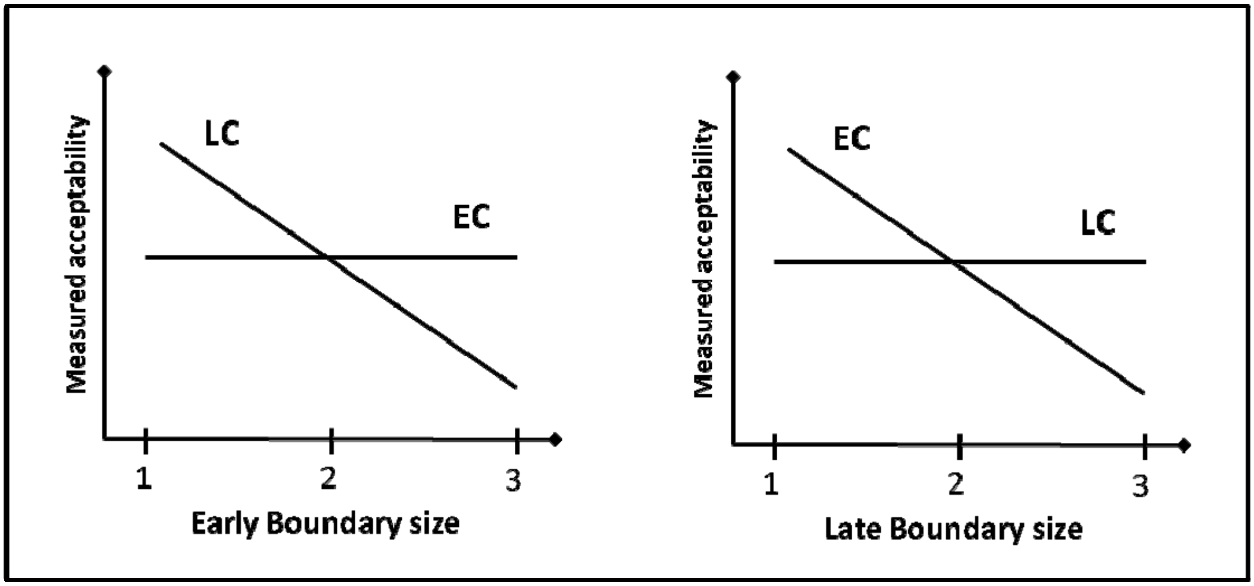


Figure 4. The measured acceptability pattern for EC and LC at early and late boundary locations.

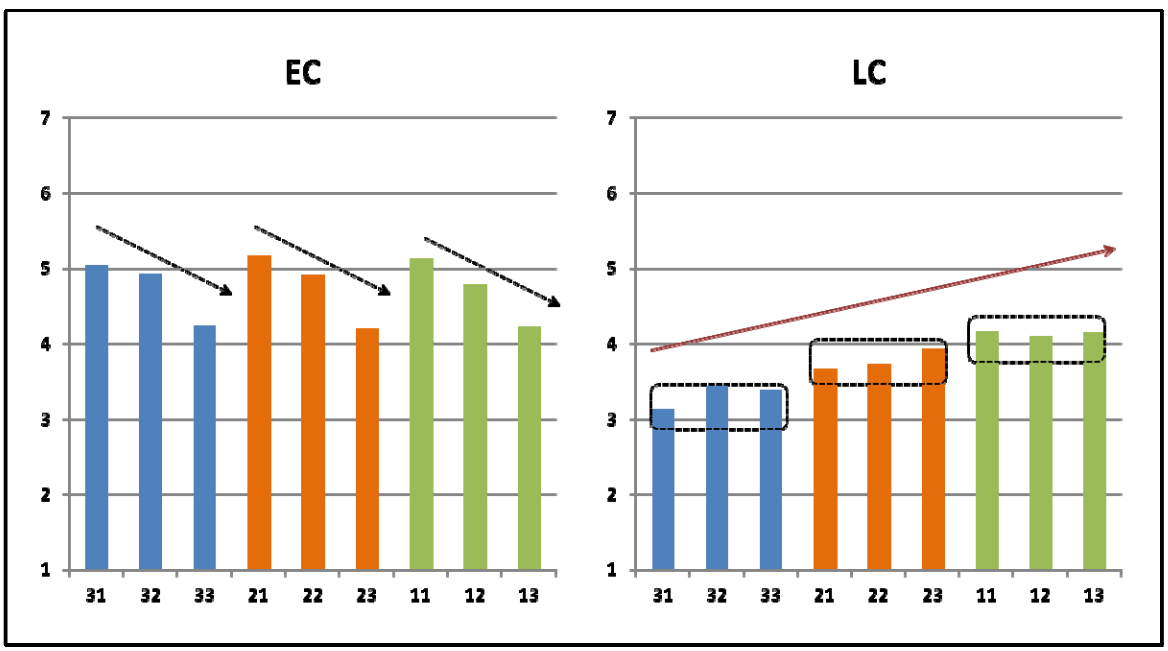


Figure 5. Acceptability scores per condition in EC and LC as a function of prosodic boundary size.

2.5.3. Same-size boundary conditions

According to both hypotheses, same-size boundary conditions (e.g., [1_1]) should receive the same acceptability rates. If this is the case, then no differences should emerge between the scores of conditions [1_1], [2_2] and [3_3]. To test this hypothesis, we ran a repeated measures ANOVA with within-subject factors *Syntax (2)* × *Prosody (3)* on the acceptability scores of these conditions and found a main effect of Syntax [$F(1,11) = 9.13$, $p<.02$], similar to the one found for the entire data set (reflecting the overall EC preference over LC) and a main effect of Prosody [$F(2,22) = 8.55$, $p<.02$]. The interaction between factors Syntax × Prosody was non-significant ($F<1$). Pairwise comparisons for the factor Prosody revealed that, similar to the overall pattern of boundary size manipulation acceptability, here too the smaller the boundary, the higher the acceptability. Condition [3_3] was significantly different than conditions [1_1] ($p<.02$) and [2_2] ($p<.01$) and conditions [1_1] vs. [2_2] showed a trend in the same direction ($p=.067$).

2.6. Discussion

In Experiment 1, we failed to obtain the acceptability preferences predicted by the IBH and AAH. Both theories assume that when one structure is favored, the other should be proportionally disfavored; this should have resulted in a complementary pattern of acceptability scores in EC and LC which, when averaged across structures or across conditions, should have canceled out any differences between them, thus preventing main effects of Syntax, Early Boundary and Late Boundary from emerging. Thanks to the syntax-independent scoring grid employed in this experiment, we found that the distribution of scores was not constant as the hypotheses would have anticipated.

2.6.1. General EC preference

The general preference for EC over LC, in all conditions, indicated that the smallest boundary size (ip), was already sufficient to trigger early syntactic closure and significantly lower the acceptability of the competing LC interpretation, even when subsequent, larger boundaries should have provided strong evidence in favor of the late closure structure and reverse the initial decision (i.e., condition [1_3], and arguably condition [1_2]). Neither the AAH nor the IBH can account for these findings. The IBH argues that the entire prosodic pattern should be taken into consideration, rather than one boundary being more heavily weighed, with the relatively larger boundary signalling informativeness. The AAH, on the other hand, does predict a local effect of boundaries, but limits it to IPh boundaries only, stressing that ip boundaries cannot trigger closure, which our data have demonstrated to be incorrect. Instead, the data indicate that the order of boundary appearance played a major role in the acceptability judgments. Although it currently does not discuss the relative influence of boundary location on interpretation, as it was not systematically tested using a superfluous boundary in both EC and LC sentences, the Boundary Deletion Hypothesis (BDH; Pauker et al., 2011; Steinhauer & Friederici, 2001; see also Bögels et al., 2013) can be extended to account for the impact of the early boundary on processing. According to the BDH, the mental deletion of a superfluous (incompatible) boundary should present greater difficulty to the processor compared to the retroactive mental insertion of a missing boundary. This assumption (in the auditory domain) is based on observations of increased processing difficulty (in the form of a larger garden path effects and lower acceptability scores) of LC structures containing two IPh boundaries, and a relatively lower processing difficulty (in the form of a weaker garden path effect and higher acceptability scores) of EC sentences containing no boundaries (see Pauker et al., 2011). Similar to previous studies (Pauker et al., 2011; Steinhauer & Friederici, 2001), in the current study we observed an overall difficulty accepting LC structures in the presence of a superfluous early boundary, in line with the current version of the BDH[7]. Unlike previous studies, here we were able to make a direct comparison to the effect of a superfluous boundary on EC acceptability. Results show that the late boundary did not lower the acceptability of EC to the same extent as the early boundary did LC, as the scores were always higher in EC compared to LC. One explanation for this outcome may lie in the differences between the structures used in previous studies and the ones used here. Recall that the IBH and AAH examined mainly global ambiguities, where the ambiguous clause (e.g., mostly an adjunct: PP, relative clause, adverbial clause) appears in sentence-final position, after both boundaries have unfolded, and is compatible with both attachment sites (albeit frequently having a structural bias towards one parse in particular). As a result, in these structures, the presence of the boundaries does not immediately strengthen or lower either interpretation, which remain acceptable before the ambiguous portion unfolds, allowing listeners to wait until both boundaries have been heard to make their decision. By comparison, in EC/LC structures, the boundaries flank the ambiguous constituent such that the early boundary appears before the ambiguous NP and the late boundary appears directly after the ambiguous NP. Because the ambiguity in EC/LC stems from whether the optionally-transitive verb takes an argument or not (that is, whether the first clause is complete and should be followed by the matrix clause, or whether more material is about to be integrated), a boundary at either an early or late position is not only expected in this context, but also required. Compared to global ambiguities, auditory studies have shown that both EC and LC structures are

---

[7] This is also in accordance with the Rational Speaker Hypothesis (Frazier, Carlson, & Clifton, 2006), which claims that listeners assume boundaries are produced with a consistent and rational intent, and, therefore, do not dismiss them.

highly acceptable in the presence of a single compatible boundary (Kjelgaard & Speer, 1999; Pauker et al., 2011; Walker et al., 2001). As each sentence in our experiment contained an early boundary – at least an ip – it seems to have served as a signal to the parser that an EC structure is currently being processed, thus fulfilling this expectation. Of course, not all sentences were expected to be EC, but by the time the late boundary appeared, the first boundary had already been integrated and consolidated with the initial interpretation, and could not be ignored or mentally deleted (in line with the BDH). In addition, the incompatible (superfluous) boundary in the EC sentences appeared directly before the disambiguating region, giving the parser less time to integrate it, which made its interfering effect easier to recover from. It seems, therefore, that at least in the context of our EC/LC sentences, the early boundary had a privileged status compared to the late boundary.

Another possibility for the advantage of EC over LC in this study, is that the progressive aspect of the optionally transitive verb (e.g., *approaching*) had overcompensated for the overall transitive bias it was used to reduce and made EC more highly preferred than LC (see Frazier et al., 2006). If this bias is indeed strong enough to completely override the effect of prosodic boundary information, lower scores for LC compared to EC should have also been obtained when a single boundary was present in the sentence. If this were the case, we should have observed an advantage of EC over LC at all Early Boundary levels, as EC should be immediately favored after the presentation of the first verb. However, results show that only boundaries 2 and 3 differed significantly between the structures ($p<.02$ and $p<.01$, respectively), while boundary one did not ($p>.07$). Given that this latter outcome does show a trend in the same direction, a more reliable outcome is needed. Also, the extent to which the aspect of the first verb contributes to the EC advantage can only be reliably ascertained by comparing the current results to ones obtained using the optionally-transitive verb in simple, rather than progressive aspect (e.g., *approach* vs. *approaching*).

Finally, it is possible that compared to previous studies using a single IPh boundary (Kjelgaard & Speer, 1999; Pauker et al., 2011; Steinhauer et al., 1999), the acoustic range between the smallest (ip) and largest (manipulated IPh) boundaries in the current study was not sufficiently large to allow an LC preference to emerge. By increasing the acoustic difference between the smallest and largest boundary sizes it might be possible to reduce, to some extent, the overall EC preference and promote LC preference for those conditions whose prosodic pattern is most compatible with that structure.

2.6.2. EC and LC acceptability patterns

Our data show that the main effects of Early and Late boundaries were not simply additive to the main effect of Syntax (i.e., they did not equally affect the two structures), and could not, on their own, provide insight into the distribution of scores within each syntactic structure. Although the majority of previous studies did not allow an examination of each structure separately, the general assumption of both hypotheses is that the acceptability pattern for each structure at each boundary position should be complementary (i.e., a mirror image of each other) (see Figure 3). That is, regardless of the predicted outcome (LC or EC preference), when one structure is favored, the other should be disfavored to the same extent. For this reason, previous studies have often reported the outcome by presenting acceptability proportions for only one structure (e.g., high-attachment proportions), which seems to suggest that any manipulation favoring structure A should equally disfavor structure B. Even investigations using EC/LC contrasts that did report preferences for EC and LC separately (Hwang & Schafer, 2006;

Kjelgaard & Speer, 1999; Pauker et al., 2011; Schafer et al., 2000; Speer et al., 1996; Walker et al., 2001) based them on a binary scoring paradigm (e.g., acceptable/unacceptable), which might not have been sufficiently sensitive. The current study, therefore, provides a first opportunity to examine more precisely any differences between scores given to the various conditions in each structure, as the scoring grid allowed us to map their degree of acceptability in a finer manner. Given previous assumptions, it was highly unpredictable that each structure would be selectively affected by a different boundary site, in a manner that seems to defy intuition: EC acceptability was incrementally lowered by the size of the incompatible (superfluous) late boundary, and LC acceptability was incrementally lowered by the size of the incompatible (superfluous) early boundary, while the magnitude of the compatible boundaries (i.e., early in EC and late in LC) had no impact on acceptability whatsoever (see Figures 4 and 5). That is, instead of supporting, and thereby increasing the acceptability of the correct structure, the boundaries here served to reject, or penalize the competing structure. While neither the IBH nor the AAH can account for these findings, the key to understanding this puzzle may be the BDH. The BDH assumes that in order to derive the correct meaning from the sentences, listeners had to ignore, or "mentally delete" the incongruent superfluous boundary that interferes with the correct phrasing of the sentence. Importantly, by using three levels of boundary sizes, it was also revealed that the reanalysis difficulty associated with boundary deletion depends on its relative size: the larger it is, the stronger the interference to the correct phrasing, and thus the lower the acceptability of that condition. For example, in LC sentence (13) below, the superfluous early boundary separates the VP (*was baking*) from its direct object (*the cake*).The larger the interfering boundary, the harder it is for the parser to process the sentence correctly, resulting in a lower acceptability for LC. The same holds for EC: the late boundary in (14) separates the subject (*the cake*) from its verb phrase (*would smell good*); as the inappropriate boundary increases, so does the processing difficulty, resulting in a lower acceptability for EC. Compared to the interfering boundaries, the late boundary in LC and the early boundary in EC are expected and congruent with each structure, and therefore do not have a negative effect on acceptability.

(13) Whenever the girl was baking # the cake # the house would smell good (LC)
(14) Whenever the girl was baking # the cake # would smell good (EC)

2.6.3. Two distinct effects

The main effect of Syntax and interactions of Syntax and Early and Late Boundary reflect two distinct effects taking place at different stages of processing. The interfering effect of the incompatible-boundaries on acceptability, as reflected by the interactions, could only take place once the sentences became unambiguous, as it depends on the confirmed structure of the sentence. In other words, in order to determine which of the boundaries served as a distractor to the intended structure, and the extent to which it interfered with parsing (based on its magnitude), one must first verify which structure is being presented. On the other hand, the early bias towards EC structure, reflected by the main effect of Syntax, occurred at an earlier stage of processing, before the sentences were disambiguated. Importantly, while both interactions represent similar (albeit opposite) acceptability patterns in EC and LC, the overall scores for EC are consistently higher than the scores for LC. If the original and extended assumptions of the BDH hold, then the presence of the early boundary created a strong early preference towards EC, which affected the overall ratings of the structures (in line with previous studies showing the immediate integration of prosodic boundaries and syntactic information (Itzhak et al., 2010;

Kerkhofs et al., 2007; Kjelgaard & Speer, 1999; Pauker et al., 2011; Steinhauer et al., 1999). Then, at the point of disambiguation, when the syntactic structure of the target sentence has been fully revealed, structural revisions requiring listeners to mentally delete the interfering boundary took place (in line with previous studies showing that garden-path effects in EC/LC mismatches occur relative to this point in time (Bögels et al., 2011; Bögels et al., 2013; Kerkhofs et al., 2007; Pauker et al., 2011; Steinhauer et al., 1999). The resulting score, therefore, reflected both effects – an interfering effect of the superfluous boundary in each structure, but higher overall scores for EC.

2.6.4. Categorical vs. continuous boundary perception

The categorical view of boundary processing, assumed by both theories, rejects the notion that within-category variations should yield differentiated acceptability results (see Carlson et al., 2001). Because the boundaries used in the current study were in part digitally manipulated and not conventionally produced, and because the ToBI guidelines themselves are not entirely unambiguous, we could not confidently determine whether intermediate boundary size 2 belongs to the ip or IPh category, or even whether boundary size 3 can be labeled a "classical" IPh. In other words, it could be claimed that (i) boundaries 1 and 2 are both ip's whereas boundary 3 is an IPh, or that (ii) boundaries 2 and 3 are both IPh's. It could even be claimed that (iii) *all* boundaries in our experiment represent variations within the ip category, given they were derived from it and were relatively acoustically similar to (although often shorter in duration than) ip boundaries used in previous studies (see Kjelgaard & Speer, 1999; Carlson et al., 2001). If the first or second assumptions were correct, then the acceptability of boundary 2 should have been the same as the acceptability of either boundary 1 or 3. If the third assumption were correct, then no differences in acceptability among conditions should have emerged at all, because all boundaries should have been treated as belonging to the same category. Contrary to these assumptions, we found that in LC, all boundary sizes yielded significantly different scores. Not only that, but the parametric manipulation was reflected in the scores given to each boundary size, with the largest boundary receiving the lowest rates and the smallest the highest rates (i.e., 1 > 2 > 3). This was also true in EC, where all boundaries showed the same pattern of acceptability. The only difference between structures was that boundaries 1 and 2 in EC were only numerically but not statistically different. This sole asymmetry between structures may indicate that given the strong early bias towards EC structure, the interference effect of the smaller late boundary sizes (1 and 2) was relatively weak (and similar) compared to that of boundary 3, while in LC each early boundary level strongly affected acceptability. Given the high similarity between the results of both structures and the fact that the difference was close to significance (p=.072), we assume that boundaries 1 and 2 were probably not perceived as categorically equivalent. However, a follow-up experiment would be necessary to examine whether these results can be replicated, and provide a stronger indication of whether boundary sizes 1 and 2 are indeed variants of the same categorical inventory or whether they have a gradient effect on processing.

2.6.5. Same-size boundary conditions

Regardless of whether EC or LC should have been the "default" structure, both the AAH and IBH claim that conditions with two same-size boundaries should be equally accepted, as the both boundaries cancel each other's effect (according to the AAH), or because neither boundary is informative (according to the IBH). Contrary to this prediction, we found that conditions with smaller boundary sizes received higher rates compared to conditions with larger boundary sizes, in both EC and LC similarly. Specifically, conditions [1_1] and [2_2] received significantly higher scores than condition [3_3], and conditions [1_1] was numerically larger than [2_2], although it only showed a trend. These findings challenge the interplay between prosodic boundaries as viewed by the theories. First, it seems that not only the relative, but the absolute size of boundaries matters, as assumed by the AAH but not by the IBH. Second, it suggests that listeners are sensitive to a range of boundary variations, as opposed to the strict categorical view assumed by both theories. However, given the difference between conditions [1_1] and [2_2] was not clear-cut, additional data are required to confirm this claim.

2.7. Experiment 2

Experiment 1 provided new insights on the inner-workings of multiple prosodic boundaries of various sizes on the comprehension of EC/LC ambiguities. By expanding the traditional binary answering grid to a scale and allowing individual scoring for each structure, we were able to observe fine-grained distinctions across conditions and syntactic structures. Our main findings revealed: (i) an effect of boundary-order appearance on structural preference, which made EC the overall preferred structure, although currently this outcome is somewhat confounded with the use of progressive tense in the critical optionally-transitive verb; (ii) in contrast to previous studies, the size of the interfering boundary in each structure affected the acceptability of the sentences, while the supporting boundary had no effect on it whatsoever; (iii) the differences across boundary levels within each syntactic structure suggest that, contrary to the categorical dichotomy between ip and IPh boundaries, listeners not only detected the rather subtle acoustic variations between boundary levels, but also based their judgments on them; (iv) finally, we found that conditions containing two equal-size boundaries were perceived and scored differently. However, it remains to be confirmed whether the early boundary blocks any option for LC preference, or whether the acoustic range between the boundary sizes used here was perceptually too small for such differences to emerge, and whether all same-size boundary conditions indeed differ from one another reliably. In addition, given the novelty of these findings, it is necessary to validate them through replication.

In Experiment 2, we aimed to replicate and extend the findings of Experiment 1 by introducing an additional prosodic boundary weaker than the original /ip/. We hypothesized that increasing the range between the smallest and largest boundary sizes would promote LC preference of (at least) the end-point condition most compatible with this structure (weakest early boundary, strongest late boundary). Second, given that the differences between the scores of all boundary levels in EC as well as the same-size boundary conditions were not conclusive, a fourth boundary level would be instrumental in providing more compelling evidence to either the categorical or the continuous view, by allowing 6 pairwise comparisons rather than the 3 possible with only three levels.

2.8. Methods

2.8.1. Participants

Twenty undergraduate students from McGill University (17 women, age range = 18–25 years) were recruited by advertisement and paid for their participation. All were right-handed (Edinburgh Handedness Inventory; Oldfield, 1971) native speakers of English with no known history of hearing impairment or brain injury. Prior to their participation, each subject gave written informed consent. One subject who did not complete the full testing session was excluded from analysis.

2.8.2. Materials

The experimental conditions were created anew using the original EC and LC sentences (containing two ip boundaries) used in Experiment 1. Based on the results of the first study, we learned that the original ip boundary was already strong enough to trigger closure, creating an across-the-board EC bias. To reduce this strong preference (in particular, in those conditions bearing distinctive LC prosodic pattern, e.g., [1_3]), an additional, weaker, boundary size was created, which was designed to simulate the absence of a boundary. This "no-boundary" condition was created by reducing the duration of the pre-boundary syllable rime by 70 ms (based on the ratings of 4 native English speakers who found it highly natural-sounding). Note that a genuine absence of a boundary was not permissible using these sentences, as the pitch contour and pitch reset indicating a boundary were present even when the pre-final lengthening duration was minimized. However, to maintain acoustic uniformity across trials, we chose not to record new sentences containing a single boundary.

The introduction of the new, smaller boundary rendered the prosodic manipulation more prone to speech rate differences. That is, we noticed that in some cases, there was some variability in speech rate that was not obvious in Experiment 1. We established a speech rate standard based on the majority of the sentences (i.e., 26), and categorized the remaining 19 for the degree to which they were perceived as slightly faster or slower than the standard. We then normalized their duration in Praat, using the first word *Whenever* – shared by all sentences – as a reference. Each normalized sentence was evaluated for speech rate and naturalness by 3 trained phonologists, and was manipulated again, if necessary. Duration measurements for the new EC/LC original pairs are presented in Table 10. Note that the overall differences compared to

Experiment 1 are quite subtle. No differences were found in $F_ø$ values between Experiments 1 and 2.

Given the additional smaller boundary size and normalized speech rate, it was also required that the magnitude of the manipulated boundaries be adjusted in order to optimize the range between the smallest and largest boundaries. Based on the judgments of 3 trained phonologists (2 of whom were native English speakers), the new IPh-compatible boundary was shortened by adding a total of 250 ms (compared to 320 ms in Experiment 1) to the original ip boundary (62.5 ms lengthening; 187.5 ms pause interval); the mid-range boundary was increased by adding a total of 100 ms (compared to 80 ms) to the original ip boundary (25 ms lengthening; 75 ms pause interval).

Thus, both perceptually and in terms of acoustic measures, the intermediate boundary was closer to the IPh boundary and more different from the original ip boundary, and was thus expected to be closer to the perceived mid-point between ip and IPh boundaries than in Experiment 1. Note also that the absolute distance between the smallest and the largest boundary in terms of durational measures was the same as in Experiment 1, i.e., 320 ms (0 to +320 ms in Exp.1, and -70 to +250 ms in Exp.2).

Table 10. Mean durations, in ms, of the major constituents in the 40 EC/LC experimental sentences

Whenever	*NP1* (*the bear*)	*V1* (*was approaching*)	*NP2* (*the people*)		*NP3* (*the dogs*)	*would*	*V2* (*run away*)
353 (26)	*487 (83)*	*713 (61)*	*560 (58)*	***EC***		*156 (32)*	*806 (190)*
				LC	*519 (89)*	*136 (32)*	*800 (190)*

Note. Standard deviations are indicated in parentheses.

The Praat script created all the possible permutations of boundary size combination (4 × 4) for each sentence (see Table 11), resulting in 16 versions per structure, or 32 overall and a total number of [32 × 45 (verbs) =] 1440 sentences. The manipulated sentences were then individually inspected to determine whether they were acoustically natural-sounding. Any sentence that presented acoustic anomaly was re-annotated and manipulated again. This process was repeated until all sentences met these standards.

Of the 1440 stimuli, 1280 (based on 40 of the 45 sentences) were selected for the experiment, while representative conditions derived from the remaining 5 sentences were later used during the practice block. Next, the 1280 experimental stimuli were evenly distributed across 4 original lists, each of which contained each of the 40 sentences in exactly four prosodic conditions, both in EC and LC. Thus, the 320 sentences of each list were comprised of 10 items in each of the 32 conditions. These lists were divided into 4 blocks of 80 stimuli each. Pseudo-randomization procedure was similar to that used in Experiment 1, with the following exceptions: (i) no more than 2 sentences of the same semantic field (e.g., sports) appeared consecutively, and (iii) no more than 5 short or 5 long sentences appeared consecutively. Last, to control for sequence effects, 4 additional lists (mirror images) were created by reversing both the block and the sentence order of each original list.

2.8.3. Procedure

Same as in Experiment 1.

2.8.4. Data Analysis

Acceptability ratings were computed as the average acceptability score given to each condition separately. Data were subjected to repeated measures ANOVA with the factors Syntax (2) × Early Boundary (4) × Late Boundary (4).

Table 11. Sample stimuli for the sixteen experimental conditions

Condition	Sentence	Disambiguating region	
		EC	**LC**
	Conj NP1 VP1 B1 NP2 B2	*Would VP2*	*NP3 Would VP2*
[1_1]	Whenever the bear was approaching / the people /	…would run away	…the dogs would run away
[1_2]	Whenever the bear was approaching / the people //	…would run away	…the dogs would run away
[1_3]	Whenever the bear was approaching / the people ///	…would run away	…the dogs would run away
[1_4]	Whenever the bear was approaching / the people ////	…would run away	…the dogs would run away
[2_1]	Whenever the bear was approaching // the people /	…would run away	…the dogs would run away
[2_2]	Whenever the bear was approaching // the people //	…would run away	…the dogs would run away
[2_3]	Whenever the bear was approaching // the people ///	…would run away	…the dogs would run away
[2_4]	Whenever the bear was approaching // the people ////	…would run away	…the dogs would run away
[3_1]	Whenever the bear was approaching /// the people /	…would run away	…the dogs would run away
[3_2]	Whenever the bear was approaching /// the people //	…would run away	…the dogs would run away
[3_3]	Whenever the bear was approaching /// the people ///	…would run away	…the dogs would run away

Condition	Sentence	Disambiguating region	
[3_4]	Whenever the bear was approaching /// the people ////	…would run away	…the dogs would run away
[4_1]	Whenever the bear was approaching //// the people /	…would run away	…the dogs would run away
[4_2]	Whenever the bear was approaching //// the people //	…would run away	…the dogs would run away
[4_3]	Whenever the bear was approaching //// the people ///	…would run away	…the dogs would run away
[4_4]	Whenever the bear was approaching //// the people ////	…would run away	…the dogs would run away

Notes: 1) In each condition, the first number corresponds to the size of the early boundary and the second number corresponds to the size of the late boundary; 2) "/" = manipulated "no-boundary"; 3) "//" = original /ip/ boundary; 4) "///" = manipulated mid-range boundary; 5) "////" = manipulated /IPh/-compatible boundary.

2.9. Results

The results of Experiment 2 were almost identical to the results found in Experiment 1, for both main effects and interactions. The acceptability judgment scores for each condition in EC and LC, and their corresponding percentages are summarized in Table 12. Differences in acceptability judgments between EC and LC (EC-LC) in each condition are presented in Figure 6.

2.9.1. Main effects

As in Experiment 1, we found significant main effects of Syntax [F(1,18) = 7.6, p<.02], Early Boundary [F(3,54) = 27.9, p<.0001] and Late Boundary [F(3,54) = 9.1, p<.01]. The main effect of Syntax reflected the general EC preference across all conditions, indicating that EC was, again, the overall preferred structure (see Table 10). Note again that reporting the acceptability proportion of a given condition without teasing-apart the syntactic structures does not accurately reflect the actual preference given to it when computed in each structure independently (the bottom row compared to the second row in Table 10). Importantly, in contrast to Experiment 1, we found that LC was *more acceptable* than EC in three conditions whose pattern of prosodic phrasing was most compatible with LC phrasing (i.e., [1_3], [1_4], and [2_4]), as illustrated by the negative bars in Figure 6. Interestingly, condition [2_4], which was preferred in its LC version in the current experiment, is similar to condition [1_3], which was preferred as EC in Experiment 1 (in both, the early boundary is the original ip and the late boundary is the digitally manipulated IPh). This finding might suggest that increasing the prosodic range by adding a boundary weaker than the original ip would shift the acceptability of some of the conditions towards LC. That is, the presence of the "no-boundary" condition seemed to have affected not only those conditions which incorporated it but the perception of the other conditions as well.

Table 12. Averages and corresponding percentage of conditions acceptability in EC and LC

Condition	*4_1*	*4_2*	*4_3*	*4_4*	*3_1*	*3_2*	*3_3*	*3_4*	*2_1*	*2_2*	*2_3*	*2_4*	*1_1*	*1_2*	*1_3*	*1_4*	*Mean*
EC	4.06	4.00	3.76	3.36	4.03	4.05	3.70	3.17	4.04	4.00	3.64	3.10	3.95	3.78	3.40	3.02	3.69
%EC (EC/6)	68	67	63	56	67	67	62	53	67	67	61	52	66	63	57	50	
LC	2.26	2.39	2.32	2.37	2.59	2.68	2.60	2.67	3.12	3.19	3.31	3.46	3.76	3.72	3.82	3.83	3.01
%LC (LC/6)	38	40	39	39	43	45	43	44	52	53	55	58	63	62	64	64	
EC+LC	6.32	6.39	6.08	5.73	6.62	6.73	6.31	5.84	7.16	7.19	6.95	6.57	7.72	7.50	7.22	6.85	
%EC (EC/EC+LC)	64	63	62	59	61	60	59	54	57	56	52	47	51	50	47	44	

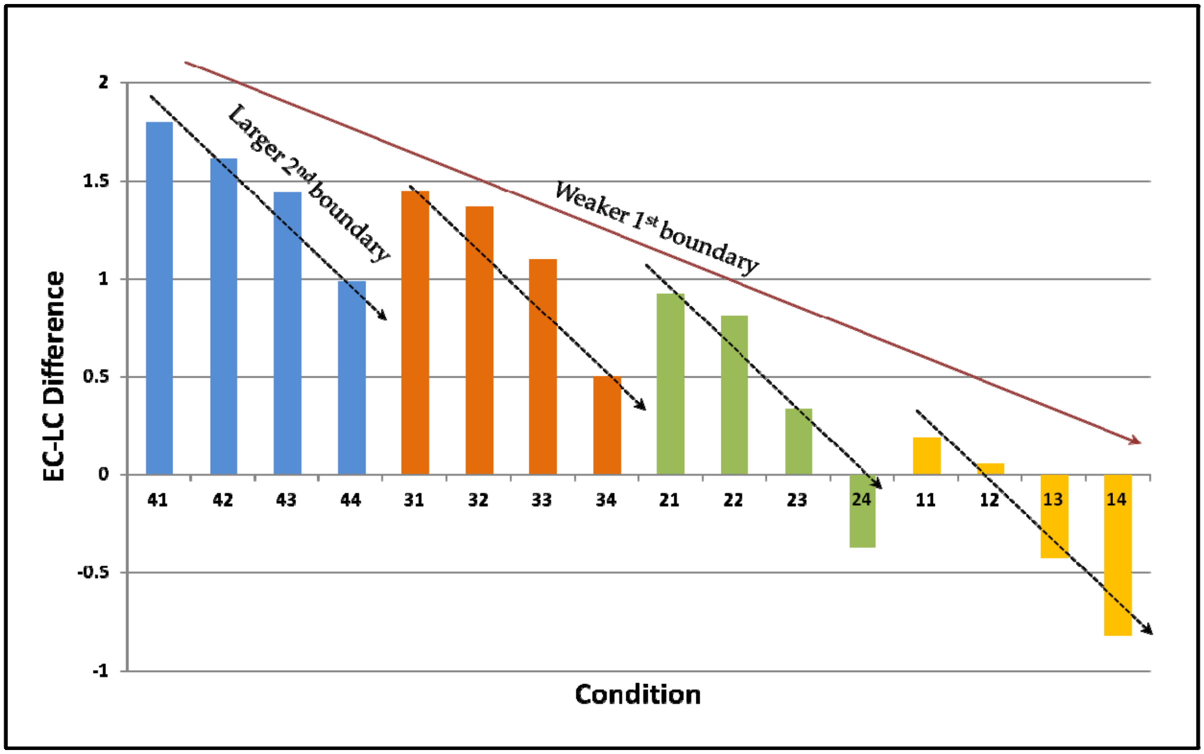


Figure 6. Difference in acceptability between EC and LC structures as a function of prosodic boundary size. X-axis: Displays the prosodic pattern for each condition (e.g., 4_1: Boundary #1 = IPh, Boundary #2 = "no-boundary"; 2_2: Boundary #1 = ip, Boundary #2 = ip; 1_4: Boundary #1 = "no-boundary", Boundary #2 = IPh). Y-axis: EC acceptability minus LC acceptability for each prosodic structure.

To illustrate the main effects of Early and Late Boundary in a manner comparable to Experiment 1, we placed them again in a matrix (Table 13), and found the same exact pattern of results. That is, as the boundary size increased, the overall acceptability decreased in a graded manner (1 > 2 > 3 > 4). This was true for both boundary 1 (downward pointing arrow) and boundary 2 (leftward pointing arrow). Follow-up analysis for the factor Early Boundary revealed that all boundary sizes were significantly different from one another ($p<.0001$; 3 vs. 4 ($p<.01$)). Follow-up analysis for the factor Late Boundary revealed that almost all pairwise comparisons were significantly different from one another ($p<.01$), with the exception of boundary sizes 1 vs. 3, which showed a trend ($p=.082$), and boundary sizes 1 vs. 2, which were non-significant. Nevertheless, based on the findings of Experiment 1 and given the similar pattern of results, we expected that here too, the effects of Early and Late Boundary locations would not be additive to the effect of Syntax, but would rather show distinct patterns of interaction with each structure.

Table 13. Total acceptability scores for each prosodic condition (EC+LC) as a function of boundary position.

		Boundary 2			
		1	2	3	4
Boundary 1	4	6.32	6.39	6.08	5.73
	3	6.62	6.73	6.31	5.84
	2	7.16	7.19	6. 95	6.57
	1	7.72	7.50	7.22	6.85

2.9.2. Interactions

As expected, we found highly significant interactions of Syntax × Early Boundary [$F(3,54) = 31.1$, $p<.0001$] and Syntax × Late Boundary [$F(3,54) = 21.3$, $p<.0001$], indicating that the main effects of Early and Late Boundary were not the same in EC and LC structures. Follow-up analyses for each structure revealed, again, a significant main effect of Early Boundary [$F(3,54) = 42.5$, $p<.0001$] but not of Late Boundary ($p>.4$) in LC, and a significant main effect of Late Boundary [$F(3,54) = 27.3$, $p<.0001$] but not of Early Boundary ($p>.1$) in EC. To determine whether the same graded pattern of results found in Experiment 1 was replicated, we conducted pairwise comparisons for each effect. In LC, the scores of all Early Boundary size levels differed highly significantly from one another (p<.0001), and showed a decrease in mean acceptability with each increasing level of boundary size (1 > 2 > 3 > 4). In EC, most Late Boundary size levels differed highly significantly from one another (p<.0001; p<.002 for 2 vs. 3), with the exception of boundary sizes 1 vs. 2, whose contrast did not reach significance (p>.4). Nevertheless, all levels exhibited the same direction of acceptability found in LC. Figure 7 illustrates the pattern of results for each structure separately.

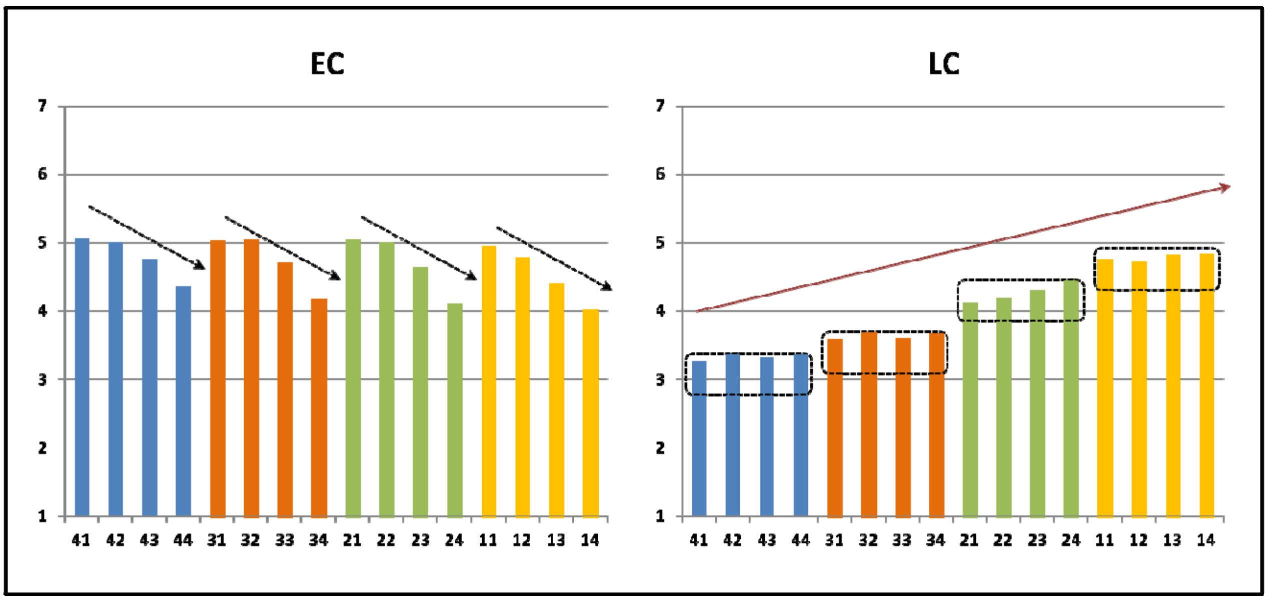


Figure 7. Acceptability scores per condition in EC and LC as a function of prosodic boundary size.

As in Experiment 1, we found no significant three-way interaction of Syntax × Early Boundary × Late Boundary [F<1]), suggesting that the main effect of Syntax and the two-way interactions were additive.

2.9.3. Same-size boundary conditions

A repeated measures ANOVA with within-subject factors *Syntax (2)* × *Prosody (4)*, performed for the 4 conditions that included 2 boundaries of equal size, revealed a main effect of Prosody [$F(3,54) = 24.3$, $p<.0001$]. Recall that in Experiment 1, three boundary levels allowed only 3 pairwise comparisons between conditions, one of which showed a trend but was non-significant. In the current experiment, the four boundary levels allowed us to perform 6 pairwise comparisons, which should aid validating or disconfirming the effects observed in Experiment 1, and determine whether the differences between conditions were reliable. Indeed, follow-up analyses revealed that all same-size boundary conditions were highly significantly different from one another ($p<.0001$; 1 vs. 2 and 3 vs. 4 ($p<.01$)). Moreover, the direction of the effect was similar to the general pattern of acceptability (illustrated in Table 11, by a diagonal starting at the top right corner), where conditions containing larger boundaries were less acceptable than conditions containing smaller boundaries. Interestingly, in addition to the expected significant main effect of Syntax [$F(1,18) = 9.3$, $p<.01$], we also found a significant interaction of Syntax × Prosody [$F(3,54) = 8.1$, $p<.0001$]. A follow-up analysis revealed that while conditions [2_2], [3_3] and [4_4] received significantly lower scores in EC than in LC ($p<.01$), condition [1_1], although going in the same direction, was not significantly different between structures ($p>.4$). This finding suggests that our "no-boundary" condition was sufficiently neutral to be perceived as equally acceptable in both structures.

2.10. Discussion

In Experiment 2, we replicated and validated the findings of Experiment 1, and extended them by confirming our hypothesis that adding a weaker boundary size would shift the structural preference towards LC in at least some of the conditions bearing a typical LC prosodic phrasing profile (weak early boundary and strong late boundary). Interestingly, while condition [1_3] received higher acceptability rates in LC, condition [2_3] received higher acceptability rates in EC. This finding is especially important, given that the "no-boundary" differed from the original ip only by a subtle reduction in duration. On the other hand, like the ip boundary, the "no-boundary" was followed by a pitch reset, which previous research has found to be indicative of the presence of a prosodic break (de Pijper & Sanderman, 1994; Swerts, 1997). Since the tonal contour characteristic of the ip category was unaffected by this manipulation, the categorical view would treat the "no-boundary" as a mere variation of ip (see Carlson et al., 2001). The fact that an LC preference was found in condition [1_3] but not in [2_3], therefore, stands in contradiction to this view.

Crucially, unlike in Experiment 1, here we found evidence that the early EC bias was reversed, provided the late boundary was sufficiently large. Specifically, the difference between the early "no-boundary" and the late boundaries in conditions [1_3] and [1_4] was large enough to trigger higher scores in LC, while a smaller (or no) difference between early and late boundaries in conditions [1_1] and [1_2] was not. Interestingly, the critical difference, or perceptual distance, necessary to permit this reanalysis consisted of two boundary levels, as condition [2_4] also received a higher score in LC. It should be noted that although we refer to this finding in terms of number of boundary size differences, we do not assume these boundaries are equidistantly distributed. By discussing the difference in so-called number of boundary sizes, we allude to the perceptual difference that was required to create this effect, which depends on multiple variables (e.g., speech rate, constituent length). Had we employed additional boundaries within the same range, or spaced them differently, the outcomes may have been different. But from the present results, we can cautiously conclude that these boundaries have different measurable effects on both syntactic structures. Based on the consistent pattern found in both experiments, we can also assume that additional boundary sizes would display the same pattern.

Although currently not addressing the relationship between strengths of early and late appearing boundaries on comprehension, the BDH, which was extended to account for the effect of boundary order, can be further extended to accommodate these findings. Results show that the early appearing boundary maintained its effect on parsing, unless the late boundary was substantially larger. Nevertheless, the size of the interfering early boundary was taken into account in the process. The penalty for this revision can be illustrated by the relatively lower score given to condition [2_4], whose earlier boundary size was larger (58%), compared to condition [1_4], whose early boundary size was smaller (64%). Although both late boundaries were large enough to reverse the EC preference, the mental deletion of a larger early boundary in [2_4] affected preference more gravely than the mental deletion of a smaller early boundary in [1_4]. Interestingly, it seems that once the late boundary surpassed a critical threshold to allow this reanalysis, its absolute size did not matter, as reflected by the same scores (64%) given to conditions [1_3] and [1_4].

The additional boundary also allowed us to examine whether the general EC preference reflected a structural bias, created by the progressive aspect of the optionally-transitive verb, or the size of the early prosodic boundary. If verb aspect affected preference, it should appear at all early boundary levels, as it should have had an early influence on parsing that was independent of prosodic information. If, on the other hand, the size of the early boundary was mainly responsible for this effect, no significant differences should be found between EC and LC when the early boundary is relatively weak. In Experiment 1, we found non-significant difference between structures in Early Boundary level 1 ($p>.07$); however, we could not draw firm conclusions based on this effect as it was rather small. In Experiment 2, on the other hand, we found strong evidence that the difference between the scores for EC and LC at the early boundary position was non-significant at levels 1 ($p>.3$) and 2 ($p>.1$), but was significantly different at levels 3 ($p<.002$) and 4 ($p<.0001$), suggesting that listeners' judgments were influenced by the strength of each boundary rather than by a pre-existing structural bias. This outcome is in line with previous evidence showing that prosodic boundaries (in this type of structure) are capable of overriding other early lexical and structural cues, such as late closure and transitivity bias, that may otherwise influence parsing (see Itzhak et al., 2010).

Importantly, Experiment 2 also served to validate two outcomes that stand in contradiction to the predictions of both theories: (i) same-size boundary conditions should receive similar scores (either because they cancel each other's effects, or because they are not informative relative to one another); (ii) boundaries are realized categorically. In contrast to the first assumption, we found significant differences between the scores of all 6 pairwise comparisons, which, in contrast to Experiment 1, were all highly significant. Second, we found highly significant differences between (almost) all boundary levels in EC and LC, suggesting that the quantitative differences in boundary strength had a gradient effect on processing (see Wagner & Watson, 2010, for review). The only exception was the non-significant difference between the scores of boundaries 1 and 2 in EC. However, similarly to Experiment 1, the extended version of the BDH can account for this finding. That is, the interference effect of the two smallest levels of late boundary on EC acceptability was negligible in the presence of the stronger early bias towards EC structure. This was not the case for LC, where each level of early boundary cumulatively affected the structure's acceptability. Thus, the parser was affected differently by the same boundary size, depending on its order of appearance in the sentence.

2.11. General Discussion

The present study investigated the influence of multiple prosodic boundaries of different sizes on the resolution of temporarily ambiguous English EC/LC garden-path sentences. Given this field of research is relatively new and under-investigated, our primary goal was to compare the outcomes to the predictions made by two prominent hypotheses – the Informative Boundary Hypothesis and the Anti-Attachment Hypothesis – regarding the expected prosody-syntax mapping in such instances. Our second goal was to examine whether listeners perceive prosodic boundaries in a categorical or a continuous fashion, as the majority of reports assume exclusively categorical perception (Clifton et al., 2002; Watson & Gibson, 2005), while recent findings suggest that listeners are able to identify and respond to more subtle differences among boundaries (Wagner, 2005; Wagner & Crivellaro, 2010). To achieve these goals we developed a novel paradigm, both in terms of stimulus creation and task, which we hypothesized would overcome some of the previous shortcomings we identified in relevant research.

Our results show that when presented with two boundaries at each closure position, EC structure is favored, even when the prosodic pattern should strongly support an LC structure. LC preference was only found when the late boundary was substantially larger than the early boundary, as seen in Experiment 2. But the most surprising finding, which was only possible to ascertain by examining each structure separately using an acceptability rating scale, was the unique effect of each boundary site on acceptability; that is, EC acceptability was exclusively affected by the size of the late boundary, while LC acceptability was exclusively affected by the size of the early boundary. Both patterns proved to be reliable, as they were replicated and extended in Experiment 2. While neither the IBH nor the AAH could account for the majority of these findings, we propose an extended version of the BDH (i.e., eBDH), which can accommodate them. We first present and discuss the principles of the eBDH in light of the outcomes, and then address the subject of categorical perception of boundaries.

2.11.1. The BDH extended

According to the original version of the BDH (Steinhauer & Friederici, 2001), retro-actively deleting a superfluous boundary is more costly than mentally generating a missing boundary. This prediction was validated in a previous study demonstrating that listeners exhibited a very large garden-path effect when presented with an ambiguous sentence (LC) containing a superfluous IPh boundary, compared to a very small garden-path effect when the ambiguous sentence (EC) contained no prosodic boundaries (Pauker et al., 2011; see also Bögels et al., 2013, for similar findings in Dutch). Directly contrasting the effect of superfluous boundaries of various sizes on both EC and LC acceptability enabled us to make the following observations:

1) The degree of difficulty associated with deleting a superfluous boundary depends on the size of that boundary: a larger superfluous boundary would be more difficult to override than a smaller superfluous boundary.
2) When the disambiguation of a sentence is heavily influenced by the presence of a prosodic boundary at either one of two potential closure/attachment sites, the order of boundary appearance gives an advantage to the earlier boundary. Consequently,

deleting an early superfluous boundary would be more difficult than deleting a late superfluous boundary, even if both are of the same size.

3) In order to overcome the bias created by the early boundary, the late boundary must be considerably larger. Once the critical threshold is surpassed, increasing the late boundary size does not seem to have an effect on acceptability ([1_3] = [1_4] in LC). However, the difficulty of overriding a larger early boundary (see point 1) would still be reflected by the score given to the reanalyzed sentence, such that the larger it is, the lower the acceptability (e.g., [1_4] > [2_4] in LC).

2.11.2. EC preference

Our finding that EC was the preferred structure in all conditions suggests that all early boundary sizes triggered closure. In fact, had we not incorporated the "no-boundary" condition in Experiment 2, we would not have observed a single instance in which LC is convincingly preferred over EC. This stands in contradiction to the assertion made by the AAH that only IPh boundaries can serve as a cue for syntactic closure. The IBH, on the other hand, allows more degrees of freedom with regard to the type of combinations that would make one structure more favored compared to the other. However, it too would not predict the effect the early boundary would have on parsing decisions. This point is crucial, because the IBH clearly states that no particular boundary position should be more dominant compared to the other in directing the parser. Rather, the relative strength between boundaries should determine the syntactic preference. The AAH also does not make any assumptions regarding an advantage of one particular boundary position over another. However, our results strongly indicate that the early boundary, even when relatively weak compared to the late boundary (in most cases), created a strong expectation for an early closure, which was very difficult to override. We have suggested that the eBDH can provide a valid explanation for these results.

The overall EC preference also stands in contradiction to an explicit expectation of the IBH (and an implicit assumption of the AAH, given their similar predictions) that LC should be the baseline preference (see Carlson et al. 2001; p. 59). To determine the baseline preference for the structures assessed, previous investigations testing these theories relied on published literature (with the exception of Carlson et al. [2001; Experiment 1], who used a written questionnaire to establish the baseline preference) (see Clifton et al., 2002; Watson & Gibson, 2005). Similarly, it is generally assumed in the literature that LC rather than EC should be the default structure in EC/LC ambiguities (namely, more easily processed, all things being equal). This assumption originates from the seminal reading study by Frazier and Rayner (1982) and later from the Kjelgaard and Speer (1999) auditory study, showing a processing advantage for LC compared to EC (in terms of RTs and error rates), not only in the baseline conditions (where the location of the boundaries was ambiguous) but also in the conflicting conditions (where a prosodic boundary was placed at a misleading position). We have previously criticized these conclusions by arguing that listeners could have more easily created an alternative grammatical revision for the LC garden-path sentences by mentally inserting a second boundary (e.g., *When the maid cleans # the rooms – they're immaculate*), while such simple revision was not possible for the EC garden-path sentences (e.g., *When the made cleans the rooms # are immaculate*), forcing a more elaborate revision process (see section 1.3). Several later studies also found that listeners show greater difficulty processing conflicting LC conditions than conflicting EC conditions (Schafer et al., 2000; Steinhauer et al., 1999; Steinhauer & Friederici, 2001). Although some of these data are based on reading paradigms, they nevertheless show an effect in the opposite direction than the one reported by Frazier and Rayner (1982) and Kjelgaard and Speer (1999). It could also be, as discussed earlier, that the progressive aspect of the first verb modulated syntactic parsing, shifting structural preference towards EC (see Frazier et al., 2006). However, since previous studies using both the simple and progressive tenses of the critical verbs resulted in similar ratings for the well-formed conditions (~90%; see Walker et al., 2001 compared to Pauker et al., 2011), and since we did not find significant difference between the EC and LC scores at all Early Boundary levels in the current study, which should have been present if verb aspect rendered EC the baseline structure, it is highly unlikely that verb aspect alone was the cause of the EC preference. Given that previous studies have demonstrated that prosodic boundary information in auditory EC/LC ambiguities overrides any competing lexical or syntactic biases (Itzhak et al., 2010), it is more likely that in the current study, the presence of the early boundary in all sentences was mainly responsible for the EC preference. This also explains why none of the earlier EC/LC studies, the majority of which used a single boundary, has ever found such a strong EC preference.

2.11.3. Differentiated pattern of results for EC and LC

The most surprising and unexpected result was the finding that the acceptability of each structure was influenced differentially by only one boundary: the EC acceptability was modulated by the late boundary, and LC acceptability was modulated by the early boundary. Moreover, the acceptability for each boundary size was significantly different from the next, in a graded manner (1 > 2 > 3 > 4). While none of the hypotheses can account for these findings, we have proposed that the eBDH can explain them in terms of the amount of interference created by the size of the superfluous boundary. Such an outcome could only have been found if listeners made an effort to understand the sentences. These findings are especially exciting given that traditional paradigms have been unable to determine whether reanalysis and correction actually take place. A low accuracy rate for a garden-path sentence, for example, cannot by itself reveal whether subjects simply rejected it (when given the choice to either reject or accept the sentence), whether they attempted to repair it and failed, or whether they managed to repair it and the score reflects the cost of this reanalysis. Similarly, a forced-choice between two structures does not necessarily reflect the amount of acceptability for that particular condition in the same manner that a separate rating for each structure does.

2.11.4. Gradient vs. categorical perception

The second goal of the present study was to examine whether the perception of boundary strength is categorical or rather continuous. One of the challenges facing the categorical view is that it does not provide a systematic measure by which to determine boundary size. The ToBI system, for example, does list the various contours for a given boundary, but admits that each category's characteristics are not absolute, but are rather determined relative to other tonal events in the sentence. In addition, we have seen that the same category may vary considerably acoustically, both in terms of duration and fundamental frequency (see Introduction section). This is of course quite logical, as spoken language is extremely variable. But for the same reason, it is difficult to maintain that the large differences between different instances of the same category, as permissible by ToBI, cannot also influence listeners' judgments variably. A number of studies propose that listeners are sensitive to subtle variations in boundary size (de Pijper & Sanderman, 1994; Sanderman & Collier, 1996, 1997) and can reliably make their parsing decisions in a gradient manner (Wagner & Crivellaro, 2010; Milotte et al., 2008).

When addressing the categorical approach, we must also address the materials used in our experiments. Recall that the original boundaries were designed to conform to the ToBI definition of the ip category (H-L contour, pitch reset, pre-final lengthening), in order to ensure that the manipulated conditions would also be ToBI-compatible. Specifically, we aimed at creating natural-sounding IPh-compatible boundaries with a H-L-L% contour. These accents and tones were adopted from the Kjelgaard and Speer (1999) study where similar sentences were used. According to Carlson et al. (2001, p.67), a "substantial variation in the physical signal is not sufficient to affect interpretation in the absence of variation in the phonological category of prosodic boundaries". Since all manipulated conditions used in the current study were derived from the original ip and varied only in terms of duration, while the tonal contour remained unchanged, it could be claimed that all boundaries should be considered instantiations of the ip category. That is, the resulting differences between them could be considered quantitative rather than qualitative, preventing listeners from discerning them, based on the categorical view (see Carlson et al., 2001; for review see Cutler et al., 1997 and Wagner & Watson, 2010). In this case, the categorical account would predict that no significant difference in acceptability should be observed for any of the conditions. Alternatively, if the manipulation successfully created an IPh boundary (level 3 in Experiment 1 and level 4 in Experiment 2), the mid-range boundary should belong to either the IPh or the ip category (e.g., either 3=4 or 3=2). Similarly, the ip and "no-boundary" should be considered variations of the same category. If this were the case, then the categorical view would predict that significant differences in acceptability between the IPh boundary (level 4) and boundaries 1 and 2. The acceptability for boundary 3 (mid-range) would either be significantly different than 1 and 2, but not from 4, if it is perceived as an IPh, or significantly different than 4, but not from 1 and 2, if it is perceived as an ip. Contrary to these predictions, our findings seem to be compatible with a continuous account. The results of both experiments show that the acceptability for all boundary levels, in both EC and LC, was significantly different, and reflected the parametric manipulation, such that each increasing boundary size lowered the acceptability in a gradient manner[8]. That is, whether our listeners heard IPh and mid-range boundaries or a variety of ip boundaries, they were able to reliably and consistently distinguish among them. Neither outcome supports a purely categorical view.

Our findings may also explain the reason why within-categorical differences in Carlson et al. (2001) were not found. Recall that in their study, Carlson et al. used adjunct attachment ambiguities, which, based on a written questionnaire, were found to be highly biased towards low-attachment interpretation. They recorded two types of ip boundaries, which differed only in terms of the lengthening of the pre-boundary word and added pause duration, while the tonal properties remained the same. The longer ip boundary was named long-ip. The IBH predicts that a larger early boundary (after *learned*) should promote low-attachment preference, while a larger late boundary (after *telephoned*) should promote high-attachment preference. Therefore, if the long-ip were perceived as larger than a "regular" ip, it should have promoted an increased high-attachment preference. Since they failed to find any differences between the ratings of sentences containing regular and long-ip boundaries, Carlson et al. concluded that the two boundaries were perceived similarly, as both were variants of the same category. However, note that the long-ip was placed only at the late boundary position, as shown below:

---

[8] The only exception to the rule was the insignificant difference between the scores of boundaries 1 and 2 in EC. However, as discussed earlier, this outcome most likely reflected the negligible amount of interference these weaker boundary levels had on EC acceptability.

(Susie learned) ip (that Bill telephoned) ip (after John visited)
(Susie learned) ip (that Bill telephoned) long-ip (after John visited)

According to the eBDH, once the early boundary was encountered, it was not only registered, but also strengthened the already strong low-attachment bias. In order to reverse this bias, a much larger boundary was necessary at the late boundary position, which was not present. It might be possible that a larger ip-derived boundary, as well as a more sensitive scoring system, might have allowed a shift in structure preference. Another possibility is that the sentences were already too strongly syntactically-biased towards low-attachment to be influenced by a larger boundary, in which case even an IPh boundary would not have been able to make a difference. Indeed, the difference between the [ip–long-ip] and the [ip-IPh] conditions was very small, as the first received 35% high-attachment rate and the second received 38% high-attachment rate. Importantly, the long-ip condition had an average pause duration of 286 ms, while our longest pause duration was 240 ms (IPh-compatible boundary in Experiment 1 = +320 ms overall; 80 ms pre-final lengthening) and was as short as 187 ms (IPh-compatible boundary in Experiment 2 = +250 ms overall; 63 ms pre-final lengthening). Still, we managed to find significant differences between the scores of our boundary sizes in both experiments.

2.12. Conclusion

The current study offers several novel insights into the effect of prosodic boundaries on the acceptability of early and late closure ambiguities in English. First, we found that the eBDH, rather than any of the more long-standing hypotheses, can best account for the findings of both experiments. In particular, it explains the dominant effect of the early boundary on parsing in terms of order of appearance, making it more costly to ignore compared to the weaker effect of the late boundary. Crucially, it also explains that the judgments for each structure were most affected by the size of the boundary interfering with the correct prosodic phrasing of that structure. We also found that qualitative boundary manipulations affected listeners' judgments in a continuous manner, with larger interfering boundaries leading to lower acceptability scores. This is, to our knowledge, the first time such effects have been demonstrated, as the task we used required scoring of each structure separately. By using an acceptability rating scale as an outcome measure, we were also able to detect subtle differences between the effects of various boundary sizes, which suggested that they were perceived differently from one another, contrary to the categorical view of boundary size perception. However, in order to determine the online effects of these types of boundaries on parsing, as well as the magnitude of the garden-path effects they create, an objective online measure (e.g., event-related potentials; EEG) would be required. Such a study is currently being analyzed in our lab.

PREFACE TO CHAPTER 3

Study 1 examined the interaction of two prosodic boundaries of various sizes on the resolution of Early and Late closure (EC/LC) sentences. Three important and novel findings were obtained. First, results demonstrate a primacy effect of prosodic boundaries on processing, reflected by an overall EC advantage, even in conditions more consistent with LC structure (containing a larger late boundary). The strong bias of even the smallest early boundary was only overturned by a maximally distant late boundary whose duration differed from the first substantially ([1_3], [1_4], [2_4]). Second, acceptability ratings showed a gradient quantitative effect of boundary size, supporting a gradient account of boundary processing. Third, only the interfering boundary in each structure drove acceptability. That is, the EC scores were decreased when the early boundary increased, and the LC scores decreased when the late boundary increased. The combined results cannot be accounted for by either the AAH or the IBH. However, the Boundary Deletion Hypothesis (BDH), predicting that interfering boundaries increase processing difficulties, best accounts for these findings. Since the BDH has been previously tested using two boundaries in LC, but not in EC, the present findings allowed us to validate its assumptions for both structures, and extend them such that they account for a variety of boundary sizes. However, given the behavioral nature of these findings, it is impossible to determine whether graded differences between boundary sizes can affect online processing in a similar manner.

Study 2 was directly derived from the results of Study 1, and was designed to test whether the outcomes found in the two behavioral studies can be observed in online processing. In particular, we aimed to examine whether the EC advantage and graded prosodic pattern would affect the magnitude of the CPS at both boundary positions, and the garden-path effects at the disambiguating region. Since this was the first study to investigate multiple prosodic boundaries of various sizes using ERPs, and given no previous studies have reliably shown a CPS to boundaries smaller than an IPh, our primary goal was to examine the effect of prosody on the magnitude of the CPS at each boundary position. The combined insights from both studies should provide a better understanding of the prosody-syntax interface, and open a range of possibilities for future investigations.

CHAPTER 3: Study 2

Electrophysiological and behavioral influence of gradient prosodic boundary sizes on ambiguous sentence resolution

Efrat Pauker

Karsten Steinhauer

ABSTRACT

The immediate effect of a single prosodic boundary on spoken language comprehension has been consistently demonstrated, both behaviorally and electrophysiologically. This event-related potentials (ERPs) study investigated how the brain processes temporarily ambiguous English garden-path sentences (EC/LC) containing two competing prosodic boundaries of various sizes. Data show gradient effects of boundary size on the acceptability ratings, as well as on the magnitude of EEG components. Specifically, we found that even the smallest boundary sizes modulated the amplitude and latency of the closure positive shift (CPS), an ERP component previously elicited only at major prosodic boundaries. The results also reflected the primacy effect of boundaries found behaviorally (rendering EC the overall favored structure), thereby exhibiting a larger garden-path effect (P600) in LC compared to EC (N400). These outcomes are inconsistent with the majority of existing hypotheses on prosodic phrasing, which assumes purely categorical boundary sizes, and in line with more recent studies suggesting gradient quantitative boundary sizes may be sufficient to explain parsing decisions.

3.1. Introduction

The way in which words in an utterance are grouped together is essential to language comprehension. The sentence '*most of the time travelers worry about their luggage'*, conveys two different meanings, depending on whether the words *time* and *travelers* are grouped into the same phrase ([*most of the* ***time travelers***] [*worry about their luggage*]) or into separate phrases ([*most of the* ***time***] [***travelers*** *worry about their luggage*]). In spoken language, one of the most effective and salient phrasing vehicles is prosodic boundaries. Prosodic boundaries are defined as perceptual breaks in the speech stream that are marked by a variety of acoustic cues, including pre-boundary syllable lengthening, pitch contour variations, pauses, and pitch resets – the composition of which depends on the degree of juncture between words (Cutler et al., 1997; Lehiste, 1973; Streeter, 1978; Swerts, 1997; Wightman et al., 1992). Suprasegmental in nature, prosodic boundaries can be detected even in the absence of clear segmental information, such as lexical and semantic content. Prosodic phrasing is in many ways speaker-dependent, as it may greatly vary based on such anatomical and idiosyncratic factors as the speaker's pitch range and speech rate (Mixdorff, 2002); in addition, there is a considerable amount of optionality with respect to the placement of boundaries in a sentence (Frazier et al., 2004; Schafer et al., 2000), since "the syntactic structure of a sentence affects but does not dictate its prosodic phrasing" (Frazier et al., 2004; p.4). Nevertheless, a variety of factors has been found to add to, interfere with, and limit the range of prosodic phrasing choices, including structural constraints (Frazier & Fodor, 1978b; Gorrell, 1995), constituent length (Clifton et al., 2006; Watson & Gibson, 2004b), semantic context, and co-occurrence frequency between words (see Wagner & Watson, 2010, for a review). For example, some linguistic theories argue that prosodic boundaries are aligned with the right edges (or end) of syntactic constituents (Selkirk, 1986, 2000; Truckenbrodt, 1999).

Consequently, the perceived grouping in speech is dependent on the relationship between these factors. According to accounts like the *Rational Speaker Hypothesis* (RSH), which links the production and perception of prosodic phrasing, listeners apply their knowledge as speakers in order to interpret utterances; they do so based on the particular patterns of prosodic phrasing, which are assumed to underlie speakers' consistent, intentional, and rational use of prosody, reflecting (at least in part) this variety of constraints (Clifton et al., 2002; Frazier, Carlson, et al., 2006). Indeed, literature demonstrates that speakers produce specific patterns of prosodic phrasing to communicate different syntactic structures[9], and that listeners can use such prosodic phrasing to disambiguate utterances (Beach, 1991; Carroll & Slowiaczek, 1987; Clifton et al., 2006; Kjelgaard & Speer, 1999; Kraljic & Brennan, 2005; Lehiste, 1973; Price et al., 1991; Pynte & Prieur, 1996; Schafer et al., 2000, 2005; Snedeker & Trueswell, 2003; Warren, 1985; Warren et al., 1995; see Cutler et al. 1997 for review). Although morphosyntactic principles do not necessarily constrain prosody, as multiple phrasing options may exist for a given structure (and in turn, syntactic structures cannot always be predicted from the prosodic structure), in some contexts, particular prosodic phrasing is obligatory and thus expected by listeners (see Shattuck-Hufnagel & Turk, 1996, for review). For example, boundaries (henceforth indicated by a hash mark, '#') are required between parentheticals ('*The surgeon # said the orderly # would perform the operation*'), non-restrictive appositives ('*my wife # the brain surgeon…*'), and at the right edges of initial subordinate clauses ('*While he was sleeping #…*') (see Frazier, Carlson, et al., 2006). One such well-studied structure is the early (EC) and late (LC) closure ambiguity (Frazier & Rayner, 1982b), as in (1):

---

[9] Note that speakers' ability to reliably disambiguate sentences prosodically was found to be higher when the task explicitly called for disambiguation (Cooper et al., 1978), than when they were unaware of the ambiguity (Allbritton et al., 1996; Snedeker & Trueswell, 2003).

(1) *After the chef cooked (#1)* <u>*the cake*</u> *(#2)…*

(1a) *…coffee was served* (LC)

(1b) *…was served* (EC)

In this structure, the end of the initial subordinate clause requires a boundary. In order to comprehend the sentence without difficulties, it is necessary to locate the closure position of this constituent. The ambiguity hinges on whether the ambiguous NP (underlined) serves as the direct object of the verb in the subordinate clause (LC interpretation – [1a]) or as the subject of the matrix clause (EC interpretation – [1b]), making the clause either longer or shorter, respectively. Since the optionally transitive verb (*cooked*) initially allows both interpretations, listeners expect either an early (#1) or a late (#2) boundary to cue the closure location of the subordinate clause, in order to resolve this ambiguity. In reading studies, the preferred (or default) interpretation is LC, which consequently leads to processing difficulties (garden-path effects) when the disambiguating word in EC (*was*) is encountered (Frazier & Rayner, 1982b). According to one of the most prominent theories accounting for this type of structural ambiguity, namely the garden-path model (Frazier, 1987; Frazier & Fodor, 1978b), to reduce the load on working memory, the parser initially constructs a single output, based on various syntactic heuristics. In case this 'guessing' strategy is realized as being incorrect - when conflicting or disambiguating information is met downstream - the parser is said to be led down a proverbial garden-path, and is then required to engage in a revision process. One of the structural principles suggested by the model, which is of particular relevance to the presence study, is the Late Closure (LC) principle, which assumes that "when possible, attach incoming lexical items into the clause or phrase currently being processed" (Frazier & Rayner, 1982b; p.180). In reading studies, violations of the Late Closure principle result in increased reading times or regressive eye movements (see Frazier, 1987; Frazier & Rayner, 1982b). However, auditory studies have demonstrated that the presence of a prosodic boundary at the early closure site (#1) reliably disambiguates these structures before the disambiguating word unfolds, thus preventing the ambiguity altogether (Kjelgaard & Speer, 1999; Marslen-Wilson et al., 1992; Pauker et al., 2011; Speer et al., 1996; Steinhauer et al., 1999; Warren et al., 1995).

A seminal study by Kjelgaard and Speer (1999) found that early prosodic boundaries not only cancel the initial effect of the Late Closure preference, but can even reverse it in favor of an Early Closure preference. Using EC/LC sentence pairs (e.g., *whenever the guard checks #1 the door #2 it's/is locked*), Kjelgaard and Speer created three conditions varying only in terms of prosodic phrasing: (1) cooperating, (2) baseline, and (3) conflicting. In the cooperating condition, the syntactic and prosodic boundaries of each respective structure coincided, creating well-formed, naturally produced sentences. The baseline conditions contained weaker boundaries at both early and late closure positions ("neutralized" prosody), and was designed to be compatible with both EC/LC interpretations, based on a norming study. In the conflicting condition, the prosodic boundaries were aligned with the syntactic boundaries of the opposite structures (e.g., early boundary in LC), resulting in a prosody-syntax mismatch. The level of processing difficulty of each condition was measured using error rates, as well as response and naming times, in four experiments. Results revealed that cooperating prosodic boundaries equally facilitated the processing of both EC and LC structures across tasks, cancelling any a priori LC preference, while conflicting boundaries led to increased processing difficulties (garden-path effects) in both structures. Importantly, in both the conflicting and the baseline conditions, LC sentences showed a significant processing advantage over EC sentences (as reflected by shorter

RTs and naming times), which the authors interpreted as indicating that the LC structure serves as the syntactic default. Although this was taken to suggest that the predictions of the LC principle hold for the auditory modality as well (see Kjelgaard, Titone, & Wingfield, 1999), a later study using similar material (Walker et al., 2001) did not observe an LC structural preference. In addition, the fact that the LC sentences were more easily processed in the conflicting condition, while no differences between structures were found in the cooperating condition, may not support an *overall* LC structural preference explanation, but rather indicate that the prosody-syntax mismatch in LC sentences was easier to recover from.[10] Recall that both cooperating and conflicting conditions contained a single, naturally produced boundary, compatible with one specific interpretation until the disambiguating word was encountered. Since both structures were easily disambiguated by the same type of prosodic information in the well-formed sentences, they should have equally confused listeners in the mismatch sentences. Moreover, it is not obvious why the *syntactic* operations involved in the revision from an initial EC parse towards an LC structure should be easier than those from LC to EC, since both require similar changes in syntactic dominance relations among constituents (see Gorrell, 1995). A recent study (Pauker et al., 2011) suggested an alternative account for the apparent LC advantage in the mismatch condition that focuses on prosodic, rather than syntactic, factors. They argued that because all LC sentences contained a pronoun in the second verb phrase (*it's*), which always co-referred to the preceding NP (*… the door it's locked*), listeners could have applied a minimal reanalysis procedure by mentally assuming another boundary at the late closure position

[10] Similarly, for the advantage of LC over EC in Kjelgaard and Speer's 'baseline' condition, one could ask if the prosodic pattern in the baseline condition was well chosen. Since the 'baseline' prosody was *defined* as being 'equally compatible with EC and LC', the mere fact that processing differences were then found anyway seems to be highly problematic.

([*Whenever the guard checks*] # [*the door #2 – it's locked*]); in which case the second clause would contain a topicalized NP, resulting in a completely well-formed structure.[11] This is an option already briefly considered – but then rejected – by Kjelgaard and Speer (1999). Note that such an alternative revision process was not an option in EC mismatch sentences ([*Whenever the guard checks the door*] # [*is locked*]): here the incompatible boundary needs to be mentally removed, and the compatible boundary before *"the door"* needs to be established, in order for the sentence to make sense[12]. Similarly, no alternative topicalization strategy was available in the LC materials of Walker et al. (2001), who used lexical NPs instead of pronouns in NP2 position (**[While the man parked] # [cars #2 – bikes were waiting]*). As one would expect based on Pauker et al.'s data interpretation, this study did not replicate Kjelgaard and Speer's finding of an LC preference. The main reason why Pauker et al. (2011) favored the topicalization revision in LC sentences over the works of a general (syntactic) LC principle as an account for the processing advantage in LC mismatch sentences has to do with the *prosodic* differences involved. According to the authors, deleting a prosodic boundary is expected to be much more difficult than inserting a missing one. The topicalization revision in LC would not involve such a boundary deletion (just a boundary addition), whereas the EC condition would, thus explaining the larger processing difficulties in LC mismatches. We will come back to this important point below. In addition, the authors raised the possibility that a lexical bias towards transitive

---

[11] Interestingly, this alternative revision would also (a) minimize the syntactic revisions (in particular, the already established dominance relations do not have to be changed) and (b) result in a perfectly coherent conceptual representation of the propositions. That is, even though not expressed syntactically, 'the door' is still the most likely referent of what was checked by the guard. In fact, work by Fernanda Ferreira has demonstrated, that is *exactly* how participants interpret these sentences (using a 'good enough' strategy that is not completely covered by the actual syntactic structure (Ferreira & Patson, 2007).

[12] It is possible that listeners instead 'corrected' the lexical input from *"is"* to *"it's"* (rather than its prosodic and syntactic structure), resulting in the LC version rather than the EC version targeted by the authors. As this is also an option that fully depends on the consistent use of pronouns in Kjelgaard and Speer (1999), methodological concerns cast strong doubt on the validity of the findings reported in this influential paper.

interpretation of the optionally transitive verb in Kjelgaard and Speer (1999), found in their corresponding norming study, might also have contributed to the LC preference in the baseline condition. That is, the absence of a clear early boundary (obligatory in this context for the EC structure) in combination with a lexical LC bias might have led participants to assume sufficiently strong evidence for an LC structure.

Pauker et al.'s reinterpretation of Kjelgaard and Speer's data was motivated by a number of other findings. A study in German had demonstrated the compelling role of covert phonological phrasing in silent reading and, similar to Walker et al. (2001), also failed to find indications for an LC advantage in mismatch conditions. In this study, Steinhauer and Friederici (2001) used commas as subvocal generators of prosodic boundaries in EC/LC ambiguities. Contrary to Kjelgaard and Speer's findings, the "reverse" LC garden-path sentences, created by introducing a superfluous comma at the early closure position, were significantly harder to process compared to the "classical" EC garden-path sentences, created by omitting the comma (required according to German punctuation rules) at the early boundary position. In order to 'repair' these sentences, readers had to ignore the comma in LC and assume a comma in EC. Since the former proved more challenging than the latter, the authors proposed the Boundary Deletion Hypothesis (henceforth referred to as BDH) according to which "the mental deletion of a previously assumed pause/comma/boundary may be more costly than the postponed insertion of an initially omitted pause/comma/boundary" (Steinhauer & Friederici, 2001; p.290). These findings have since been confirmed in the auditory modality as well, both in English (Pauker et al., 2011) and in Dutch (Bögels et al., 2013; see below for details). In addition, a follow-up study in English by Itzhak et al. (2010) demonstrated that in this context, prosodic phrasing is indeed more influential than both the structural preferences (Late Closure principle; Frazier & Fodor, 1978b) and lexical biases (transitivity of the initial verb; Garnsey et al., 1997; Pickering et al., 2000) that have been found to guide initial sentence interpretation in reading (Frazier & Rayner, 1982b). They showed that in the absence of audible prosodic boundaries, the parser relies on these lexical and structural heuristics, similarly to the strategy observed in reading studies (Frazier & Rayner, 1982b), whereas the presence of overt prosodic boundaries completely overrides both factors.

3.1.1. Types of prosodic boundaries and their status in prosody research

While earlier research on prosodic phrasing focused primarily on the impact of one single prosodic boundary on processing by manipulating its presence versus absence and location, recent studies have begun to explore the interaction between two competing boundaries on ambiguity resolution, both in terms of size and location (Carlson et al., 2001; Clifton et al., 2002; Frazier, Carlson, et al., 2006; Schafer, 1997; Watson & Gibson, 2004a, 2005). Whereas the manipulation of the number of prosodic boundaries in a sentence is relatively easy to do, characterizing their size – and the impact of size manipulations on sentence processing – is subject of an ongoing theoretical debate. Since these issues constitute a central challenge for this line of research in general, and for the present study in particular, the following paragraphs will provide a brief overview of the theoretical framework and functional implementation of the principles used to define prosodic boundary size.

Researchers have long acknowledged that acoustic parameters associated with prosodic boundaries do not provide a consistent and accurate measure for phrasing on their own, due to their high variability both within and across speakers (Shattuck-Hufnagel & Turk, 1996). In the past few decades, an approach to prosodic phrasing has become a convention, which assumes prosody can be parsed, much like syntax, into constituents forming a hierarchical order, and that those constituents are marked by distinctive acoustic cues. These prosodic categories are thought to represent both the grouping and rhythmic prominence (stressed and unstressed syllables) of words in the utterance (Beckman & Pierrehumbert, 1986; Hayes, 1989; Nespor & Vogel, 1986; Selkirk, 1986; for review see Shattuck-Hufnagel & Turk, 1996). Thus, spoken sentences are assumed to be exhaustively parsed into smaller units (of the same type, at each level) – the largest of which spans the entire utterance and the smallest of which marks the syllabic level (some accounts postulate levels smaller than the syllable, depending on the language), as shown in Table 1 below (Shattuck-Hufnagel & Turk, 1996; p.206):

Table 1. Leading models for prosodic structure hierarchy. Adapted from Shattuck-Hufnagel & Turk (1996; Fig. 2).

Nespor and Vogel (1986), Hayes (1989)	Selkirk (1978, 1986)	Beckman, Pierrehumbert (1986, 1988)
Utterance	(Utterance)	
Intonational Phrase	Intonational Phrase	Full Intonational Phrase
Phonological Phrase	Major Phrase	Intermediate Intonational Phrase
---	Minor Phrase	Accentual Phrase
Clitic Group	---	
Prosodic Word	Prosodic Word	
Foot	Foot	
Syllable	Syllable	
	Mora	

The theoretical accounts illustrated in this figure all assume two prosodic categories at the phrase level, namely the smaller intermediate (or phonological) phrase (ip), and the larger intonational phrase (IPh). The right edges of these constituents are similarly referred to as ip and IPh boundaries, which conform to the same categorical distinctions, such that IPh boundaries are expected to be perceived as larger than ip boundaries (Rossi, 1997).

Importantly, the number and nature of these levels is not universally agreed upon; for example, some theories assume more than one level of phonological phrase (Selkirk, 1986; for review see Shattuck-Hufnagel & Turk, 1996), and hence, another boundary level. This point is crucial, as the notion of prosodic categories at the phrase-level is still debated, yet much research on prosodic phrasing in English has subscribed to this ip/IPh dichotomy. As put by Ostendorf (2000; p.265):

"Just as phones or other sub-word units are used as an intermediate level between acoustic features and words in large vocabulary continuous speech recognition, there is a practical need for intermediate prosodic units to allow for generalization over the many possible combinations of prosodic patterns that can occur. Unfortunately, there is no clear theoretical consensus on the appropriate units, *in particular the number of levels of phrase constituents* and the grammar of these constituents are issues of debate."

In parallel to this debate, a need has arisen for the development of a practical empirical tool incorporating the core assumptions of these theories. The most widely used system created to implement these theoretical principles into a workable framework is the Tones and Break Indices (ToBI) transcription system (Beckman & Elam, 1997; Beckman & Hirschberg, 1994; Beckman & Pierrehumbert, 1986; Silverman et al., 1992), which is used to describe the prosodic grouping and tonal events in an utterance in an integrated manner. Although other labeling systems exist beside ToBI (e.g., Rhythm and Pitch – RaP; Dilley & Brown, 2005), it is currently the most widespread framework, and considered a standard for American English (note that versions of ToBI exist for several other languages, like Greek and Japanese). In particular, ToBI clearly distinguishes between the ip and IPh categories, describing each with a characteristic set of intonation contours (tones) and pause length (break indices), and, therefore, different boundary sizes. Within each ip, perceptually prominent (stressed) words are marked with a *pitch* *accent* (*), reflecting rising (H*), falling (L*), or more complex (e.g., L*+H) pitch variations. The pitch accent is immediately followed by a high (H-) or a low (L-) *phrase accent*, which delineates the intonation contour between this point and the right edge of the ip. Every IPh is comprised of at least one ip, and ends with a high (H%) or low (L%) *boundary tone*, appearing on the final syllable or word (depending on its length) of the phrase. The system of break indices associates a different degree of disjuncture with every prosodic constituent, on a scale from 0-4. IPh boundaries are the most prominent and are assigned the strongest break size (4), whereas ip boundaries are assigned the next smaller break size (3) (Beckman et al., 2005; Pitrelli et al., 1994). Both tones and break labels are considered relative rather than absolute markers, in order to account for a variety of prosodies. Note that based on this system, IPh boundaries always coincide with their final ip boundaries, and share the same tonal patterns – with the exception of the boundary tone (e.g., L*H-L%; H*L-L%). This observation is important, because it illustrates that the tonal difference between these boundaries is rather subtle, and can be made ambiguous if both right edges go in the same direction (L-L%). In this case, the boundary would be perceived as containing a relatively longer falling pitch compared to a boundary with varied contour (H-L%). An important perceptual difference between the ip and IPh boundaries is often made based on their durational properties, according to which IPh boundaries are characterized with longer pre-final lengthening and a silent interval, i.e., a pause (see Shattuck-Hufnagel & Turk, 1996 for review).

Most researchers exploring the manner in which two boundaries affect the syntactic parsing of American English sentences assume that listeners process boundary information in a categorical and discrete manner, as defined by the ToBI system. An important implication of the categorical distinction between ip and IPh boundaries is that no differences in processing are expected between variants of the same category. That is, while acknowledging that boundaries of a specific prosodic category can be produced in many different ways, proponents of the categorical view argue that the "syntactic structure can influence how an utterance is divided into ip's or IP's, but all boundaries of a given kind are equivalent and thus *continuous variation within a category plays no role at the syntax-prosody interface*" (Snedeker & Casserly, 2010; p.1239). This assumption is currently supported by a single study showing that increasing only the durational properties (pre-final lengthening and pause) of an ip boundary (while maintaining its tonal structure) did not impact listeners' syntactic preference compared to a "regular" ip boundary (Carlson et al., 2001; Experiment 2). Since the ToBI-defined ip-compatible intonational contour remained the same (H*L*L-) in both productions of the ip boundaries, the authors argued that listeners attended to the phonological identity, rather than acoustic prominence, of the boundaries. However, this explanation may not actually capture the reason for this finding. First, the long and short ip's were placed at the late boundary position in a sentence with an adjunct attachment ambiguity (e.g., *Susie learned #1 that Bill telephoned #2 after John visited*), where a larger late boundary was expected to trigger high (matrix) attachment preference (i.e., VP1 – *learned*). However, since in this structure prosodic boundaries are optional and the overwhelming syntactic preference is for a low attachment interpretation, boundaries showed a rather limited effect. In fact, using an IPh boundary, in the same position in the same sentence, only showed a slightly higher (38%) high-attachment preference compared to the one found with ip boundaries (35%), even though the pitch contours were different. This suggests that the choice of sentence material did not allow one to observe clear behavioral effects even for large (between-category) prosodic differences, let alone for more fine-grained within-category variations. Second, recall that ToBI describes the degree of disjuncture of the category as well as its tonal contour. However, it only serves as a relative/subjective measure. The difference between ip (break index 3) and IPh (break index 4) is not an accurate and objective measure, as it depends on the transcriber's judgment. In fact, it has been demonstrated that agreement between transcribers on the identity (as opposed to the presence vs. absence) of edge tone and pitch accent was low (< 50%) or inconsistent for most types, suggesting that even highly trained ToBI labelers did not reliably identify these key elements (see Hirst, 2005; Syrdal & McGory, 2000; Wightman, 2002). With respect to break indices, Syrdal and McGory (2000) reported a high agreement for break level 4 (~80%), but a low agreement for break level 3 (~35%). In Carlson et al.'s (2001) study, the "regular" ip contained a break spanning an average of 60 ms (SD 54 ms), whereas the "long" ip contained a break spanning an average of 286 ms (SD 83 ms), both presumably labeled 3, as this is the only type of disjuncture allowed for ip's. By comparison, in another experiment of the same study (Carlson et al., 2001; Experiment 4) the IPh boundary in one condition contained a 382 ms (SD 58 ms) break, which was almost the same as the 362 ms (SD 55 ms) break of an ip boundary in another condition, both located at the same position in the sentence. Both IPh and ip boundaries in that experiment also had comparable pre-final lengthening duration (539 and 549, respectively). The main difference between them was a more complex, and thus possibly more prominent, tonal variation in the IPh (H*L-H%), compared to a more monotonous tonal variation in the ip (H*L*L-). This begs the question: where does one draw the line between boundary categories, and is there, in fact, any?

This is indeed one of the major challenges and a source of criticism regarding ToBI's labeling guidelines, which require transcribers to interpret the data based on both visual and auditory inspection, rather than rely on their perceptual impressions alone (see Dilley, Breen, Bolivar, Kraemer, & Gibson, 2006). Wightman (2002; p.3) states that "the restriction that certain break indices can be used only in combination with specific intonation labels is one of the most controversial aspects of the ToBI system. The linkage between the tiers further *de-emphasizes the perceptual experience of the listener*: The ToBI guidelines even suggest that, once *either* the tonal *or* phrasal labels have been produced by the listener, the redundant labels be inserted automatically to save time and increase inter-transcriber agreement". Also, since ToBI is theory-based, it is subjected to modifications. Over the years the number of levels and their identity have been changed and adjusted, with earlier accounts suggesting seven levels (Price et al., 1991),which were eventually scaled down to five (Wightman, 2002). Nevertheless, there is no universal agreement regarding the theoretical validity of these particular categorical distinctions. As put by Selkirk (1984; p.29):

> "… language may exhibit more than one level of phonological phrase, in which case finer terminological distinctions can be made: PhP^{1}, PhP^{2}, . . . , PhP^{n}. With this terminology then, an intonational phrase is a special case of a phonological phrase, one that is associated with a characteristic tonal contour and that has an important function in representing the "information structure" of the sentence. The unit utterance, if it existed, would also be a phonological phrase in this sense." (Note: PhP refers to the phonological phrase).

While resolving this debate is beyond the scope of this paper, the reader should be aware that such strict and perhaps abstract categorical differences, as well as rigid labeling guidelines, likely ignore finer differences between or within categories, which may be meaningful to the syntax-prosody mapping.

Indeed, some studies suggest that listeners are aware of non-categorical prosodic differences, which in turn influence the way they perceive and process sentences. By varying pitch and durational cues to boundary strength, they have reported that boundary sizes can be discerned in a gradient quantitative, rather than a qualitative, manner, showing that listeners can detect and process a wider range of boundary strengths (de Pijper & Sanderman, 1994; Sanderman & Collier, 1996, 1997; Wagner, 2005; Wagner & Crivellaro, 2010; for review see Wagner & Watson, 2010). An important study by de Pijper and Sanderman (1994), found that both naïve and trained listeners can reliably and similarly perceive boundaries in a gradient manner, on a scale from 1 to 10, both in natural and delexicalized productions. Importantly, larger boundaries were most strongly correlated with a pitch reset, and longer duration of pre-final lengthening and (to a lesser degree) silence interval. Sanderman and Collier (1996, 1997) tested five levels of digitally-synthesized boundary strengths using an acceptability task, and showed listeners were able to significantly distinguish between them on a 10 point scale. Importantly, unlike de Pijper and Sanderman (1994), here the pre-final lengthening played a smaller part compared to other cues (e.g., pitch variations), suggesting no particular set of cues is responsible for this outcome; rather, participants seem to gravitate towards a finer scale of boundary strength since it likely reflects a more natural approximation of natural speech production (as this type of synthesized material was preferred to almost the same degree as natural material). Finally, a more recent study by Wagner and Crivellaro (2010) provided strong evidence for gradient boundary strength processing of sentences with a prosodic bracketing ambiguity (e.g., *the tires may wear down the road*), using six levels of digitally manipulated boundary levels. The boundaries were manipulated by increasing the duration of the pre-final lengthening and pause in five equal steps (0 being the original; 400 ms increase overall) with a 40-60% ratio. Listeners then chose one of the two possible phrasing options and rated how confident they were with their selection. Results showed participants' parsing decisions were significantly influenced by the relative boundary strength in a cumulative manner, a finding which challenges a purely categorical view. It is important to mention that the particular syntactic ambiguity (verb particles/preposition) was selected based on findings demonstrating that it is reliably disambiguated by prosodic bracketing (Price et al., 1991).

Against this background, Pauker, Wagner, and Steinhauer (In preparation) recently conducted two experiments testing both the influence of two boundaries on sentence comprehension, as well as whether continuous differences in boundary size affect listeners' parsing decisions in a cumulative manner, or whether they display exclusively categorical effects. Listeners were presented with highly controlled digitally-manipulated EC/LC sentences, each containing two prosodic boundaries at the early (#1) and late (#2) closure positions (e.g., *whenever the bear was approaching $_{\#1}$ the people $_{\#2}$ (EC): …would run away*; *(LC): …the dogs would run away*), which differed only in terms on their relative sizes (3-4 levels per boundary). The task required an acceptability judgment on a (5 or 7 point) scale. To allow the evaluation of both categorical and non-categorical accounts, the originally recorded boundaries were ToBI-defined ip's, to ensure the stimuli were compatible with those used in previous studies. In addition, to maintain a high acceptability and an ambiguity of the sentences prior to the disambiguating words, the pre-selected pitch contour was similar to the one used by (Kjelgaard & Speer, 1999) in their baseline condition (H*L-). Each ip boundary also contained a pitch reset, followed by a return to the pre-boundary $F_ø$ level (as opposed to a downstep), to clearly signal prosodic juncture (Swerts, 1997)[13]. This enabled the creation of natural-sounding ToBI-compatible IPh boundaries containing a (H*L-L%) pitch contour, by adding a total of 320 ms (in the first experiment) and 250 ms (in the second experiment), with a 25-75% ratio of pre-final lengthening to pause duration, using a Praat (Boersma & Weenink, 1996) script. Two additional

[13] It should be noted that unlike (Kjelgaard & Speer, 1999), both boundaries were produced such that the duration properties were highly matched, to avoid any preliminary bias towards one interpretation.

boundary sizes were also created: a mid-range boundary size, designed to be perceived as larger than the original ip and smaller than the manipulated IPh, representing a perceptually "neutral" position between the two categories (a total of 80/100 ms in experiments 1 and 2, respectively), and a "no-boundary" condition (used only in the second experiment), designed to simulate the absence of a phrase boundary (by subtracting 70 ms from the pre-final lengthening of the original ip boundary). The results of both experiments showed that (i) even small early boundaries biased listeners towards EC, with much larger second boundaries required to override this bias; (ii) surprisingly, only the strength of the *incompatible* boundary (1^{st} in LC, 2^{nd} in EC) drove acceptability ratings – a finding predicted by the Boundary Deletion Hypothesis (BDH), according to which deleting a superfluous boundary should be more demanding than mentally creating a needed boundary; (iii) most importantly, they found a strictly gradient pattern of the acceptability rates within each structure. The findings were taken to demonstrate that listeners are sensitive to subtle, non-categorical differences between prosodic boundaries, which in turn cumulatively affect the degree of processing difficulty. Such outcomes cannot be explained by a purely categorical account. Based on these findings, the authors extended the BDH by adding the following predictions:

(1) The difficulty associated with mentally deleting an incompatible boundary increases with increasing boundary size.
(2) In structures that require a prosodic boundary for disambiguation (at one of two closure/attachment positions), the mental deletion of an early superfluous boundary is more difficult than that of a late superfluous boundary, if both are of the same size.
(3) In order to override the structural preference induced by an early boundary, the competing late boundary must be substantially larger.

3.1.2. ERPs in prosody research

Despite providing strong preliminary evidence for our assumptions, the majority of the reviewed studies employed behavioral measures, which were likely influenced by offline decision processes, and do not permit the inspection of these processes in real time. Since the late 1990s, event-related brain potentials (ERPs) have been successfully used to investigate the prosody-syntax mapping in an objective online manner. ERPs are sequences of negative- and positive-going waveforms (‘ERP components’) that reflect the brain’s neural activity following the presentation of a target stimulus, such as a tone, a syllable, or a word in a sentence. Each ERP component is assumed to reflect a relatively specific mental process. As subjects’ brain activity is continuously recorded with a high temporal resolution while they engage in language processing tasks, researchers are able to examine multiple ERP components for any part of a sentence and link them to the corresponding linguistic events of interest. In psycholinguistic ERP research, a number of ERP components have been identified that typically occur within 1000 ms after word presentation and seem to be associated with distinct language-related neurocognitive processes. Their respective functional significance ranges from early pattern recognition (P200) and phoneme discrimination (mismatch negativities, MMN) during the first 250 ms to syntactic reanalysis (P600) and mental rehearsal (sustained negativities) after 500 ms. Two well-established ERP components relevant to the present paper are the N400, a centro-parietal negativity around 400 ms post word onset that reflects lexical-semantic processing difficulties (for review see Kutas & Federmeier, 2011; Kutas & Hillyard, 1980, 1984), and the P600, a parietal positivity after 600 ms that has been linked to morpho-syntactic processing difficulties, including reanalyses in garden-path sentences (Osterhout & Holcomb, 1992; for review see Steinhauer & Connolly, 2008).

A seminal study by Steinhauer et al. (1999) reported the first ERP evidence demonstrating the immediate influence of prosodic boundaries on parsing, using spoken German structures with temporary attachment ambiguities. In addition, Steinhauer et al. discovered that a bilaterally distributed positive waveform, largest over the midline electrodes and spanning about 500 ms, was elicited relative to the prosodic boundaries in these sentences. The authors interpreted this component as signaling the closure of intonational phrases (i.e., IPh boundaries), and termed it the *Closure Positive Shift* (CPS). The authors also examined whether the CPS reflected the acoustic rather than the phonological level of processing, particularly whether the CPS relied on a specific acoustic cue, such as the presence of a silence interval interrupting the speech signal, or on the array of prosodic boundary markers (e.g., pre-final lengthening, pitch variation), shown to be used by speakers. To this end, they created new conditions by removing the entire pause from the corresponding original sentences, while maintaining the other intonational and durational properties of the boundaries, and were able to replicate this component, thereby demonstrating that the CPS indeed taps the phonological level (for an overview seeBögels et al., 2011; Steinhauer, 2003). The CPS has been successfully replicated in German (Isel et al., 2005; Pannekamp et al., 2005; Steinhauer & Friederici, 2001), as well as cross-linguistically, including in Dutch (Bögels et al., 2010; Kerkhofs et al., 2007, 2008), Japanese (Mueller et al., 2005), Chinese (Li & Yang, 2009), English (Pauker et al., 2011; Steinhauer et al., 2010), and Korean (Hwang & Steinhauer, 2011). The CPS was found to reliably reflect the perception of prosodic phrasing in other contexts as well, including in 'jabberwocky' sentences devoid of semantic information (Pannekamp et al., 2005), sentences stripped of all segmental content, including lexical and syntactic information (Pannekamp et al., 2005; Steinhauer & Friederici, 2001), and in reading, triggered by commas (Steinhauer &

Friederici, 2001). Importantly, the CPS was generated even in the complete absence of acoustic or orthographic boundary markers, due to an expectation of a boundary after long subject NPs in silent reading (Hwang & Steinhauer, 2011), and at the late closure position in spoken EC sentences, where the initial verb was transitively biased (Itzhak et al., 2010). This evidence demonstrates that the CPS is a universal marker of implicit and explicit prosodic phrasing.

Specifically relevant to the present investigation is a recent study by Pauker et al. (2011), who presented listeners with EC/LC structures, based on the sentence material used in Kjelgaard and Speer (1999) and Walker et al. (2001), but using different sub-conditions. They recorded two well-formed LC (condition A) and EC (condition B) pairs with cooperating IPh boundaries, which were then cross-spliced to create prosody mismatch conditions C (with no boundaries) and D (with two boundaries). Thus, conditions A/D and B/C were matched lexically but not prosodically, as in the following example:

(2) [A] When a bear is approaching the people # the dogs come running
[B] *When a bear is approaching # the people come running*
* [C] When a bear is approaching *the people come running*
* [D] *When a bear is approaching # the people # the dogs come running*

Compared to condition A, in condition D, the superfluous (conflicting) boundary at the early closure position meant NP2 was flanked by two large prosodic boundaries, thus preventing the assignment of a theta role to this constituent (it was neither integrated as the direct object of the verb in the subordinate clause, i.e., Patient, nor as the subject of the matrix clause, i.e., Agent). This anomaly, perceived relative to the onset of the late boundary (when it became clear NP2 was stranded), elicited a strong garden-path effect, reflected by a biphasic N400-P600 pattern.

On the other hand, Condition C, without any boundaries (resembling classical garden-path sentences in reading), elicited a weak garden path effect, as reflected by a small P600 at the disambiguating word (*come*), and higher acceptability rates (~53% vs. 28%) in the judgment task. These results confirmed, for the first time in an auditory study, the prediction of the BDH that structures containing a superfluous incongruous boundary should present greater processing difficulties compared to structures with a missing boundary, in contrast to Kjelgaard and Speer's (1999) claim that LC structures should generally be processed more easily. Moreover, as expected, a CPS component was reliably elicited relative to each boundary in conditions A, B, and D, presenting a highly consistent pattern throughout. These findings demonstrate that prosodic boundaries are more informative to the processing of EC/LC ambiguities than other syntactic (LC principle) and lexical (transitivity) biases, and that the degree to which these boundaries interfere with parsing is reflected by the magnitude of the garden-path effects.

While providing crucial evidence regarding the online elicitation and interaction of prosodic boundaries with parsing, this study had several limitations. First, the prosodic manipulations were not symmetrical, and therefore, did not allow a direct comparison of their influence on both EC and LC sentences (e.g., EC sentences with two boundaries). Second, because the garden-path and CPS components in condition D were elicited relative to the same events (the onset of the late boundary), the CPS was superimposed by the P600, which made it difficult to inspect these components separately. Third, like in almost[14] all other reviewed auditory EEG studies, the CPS was evoked using exclusively IPh boundaries, and therefore the findings cannot contribute to the categorical vs. gradient boundary processing debate. To address

---

[14] The exception being Li and Yang (2009), who found no differences between CPS evoked at ip and IPh boundaries, but whose data was also rather noisy.

some of these limitations, a behavioural study conducted by Pauker et al. (In preparation) employed a fully symmetrical prosodic design, and provided compelling behavioral evidence for the gradient perception of two competing boundaries, each of which was realized with three (Exp. 1) or four (Exp. 2) different sizes, thus parametrically manipulating the boundary strength at both positions (see Methods section in Chapter 2 of this thesis).

However, to explore whether and to what extent these findings are also reflected by online measures, it is necessary to replicate the behavioural study using ERPs. ERP measures are expected to reveal gradient or categorical boundary perception both at the boundary positions (reflected by the CPS) and at disambiguating positions (reflected by garden path components, such as N400s and P600s). Moreover, the current design also overcomes a limitation of the Pauker et al. (2011) ERP study, which tested two boundaries only in the LC condition, resulting in a predictable structure at onset of boundary #2 and a surprisingly early P600 effect at boundary onset (see Bögels et al., 2013, for dicussion). Since the new stimulus materials of Pauker et al. (In preparation) contained two boundaries in both EC and LC structures, listeners would be able to disambiguate the sentences only *after* the late boundary unfolds, thereby likely allowing a separation of the CPS and subsequent garden-path effects.

In the present study, we thus adopted the improved paradigm of Pauker et al. (In preparation) to investigate the influence of two prosodic boundaries of varying strengths on the processing of EC/LC ambiguities. The main objective was to address the following research questions:

(i) Can the behavioural results of Pauker et al. (in preparation) be replicated during an ERP experiment?

(ii) If so, do ERP online measures in listeners provide additional information regarding the 'categorical' versus 'gradient' processing views of differences in boundary size?

(iii) Can the CPS be elicited by boundaries *smaller* than an IPh?

(iv) If so, does the CPS amplitude *reflect the size differences* among boundaries, or is the CPS rather an 'all or none' brain response reflecting the mere presence of perceived boundaries?

(v) Irrespective of the CPS findings, do ERP garden path effects (N400s, P600s) mirror the gradient pattern of processing difficulties observed in the judgment data of Pauker et al. (in preparation)?

(vi) Lastly, with respect to the extended BDH (henceforth: eBDH), one of the most interesting questions was whether we would find ERP evidence supporting the notion that garden path effects depended primarily on the strength of the incompatible (superfluous) boundaries (i.e., the first one in LC, but the second one in EC).

3.1.3. Specific Hypotheses

Based on the eBDH as well as on the combined findings of Pauker et al. (2011), and Pauker et al. (In preparation), given that virtually all of our sentences contain a superfluous boundary that requires deleting, all conditions were expected to exhibit a certain degree of processing difficulty. According to the eBDH, an advantage for the early boundaries should be observed in these structures. Therefore, the behavioural data should replicate the overall EC preference. Moreover, processing difficulty was predicted to be greater in the LC versions, which contained an early incongruous boundary, compared to the EC versions, which contained a late incongruous boundary, as reflected by stronger garden-path effects in LC. The CPS was expected at least at the largest (IPh) boundaries, irrespective of position (early versus late) and structure (EC vs LC). If smaller boundary sizes prove sufficient to elicit a CPS, its amplitude might either (a) be relatively constant across boundary sizes or (b) increase with boundary size.

The former case (a) would point to an 'all-or-none' type of brain response[15], whereas the latter case (b) would suggest a gradient electrophysiological reflection of either (b1) acoustic or (b2) phonological boundary processing. Since a 'gradient' view of boundary processing (e.g., Wagner & Crivellaro, 2010) is, in principle, compatible with many distinct degrees of boundary strength *both* at the acoustic and at the phonological level of processing, it would assume cumulative quantitative differences between the CPS components at each boundary size (i.e., a gradient pattern of up to four different levels of CPS amplitudes) for *both* (b1) and (b2), i.e., irrespective of the level of processing reflected by the CPS (acoustic or phonological). In contrast, the classical 'categorical' view of boundary processing would permit many degrees of *acoustic* boundary strength, but can allow only two (or maximally three) levels of *phonological* processing: no boundary/ip boundary/IPh boundary. Therefore, if the actual CPS amplitude in our data is found to distinguish between exactly two levels of boundary strength (e.g., levels 1+2 versus levels 3+4), this result would support the categorical view and, moreover, suggest that the CPS reflects *phonological* rather than acoustic processing levels. In contrast, a gradient CPS amplitude would in principle still be compatible with both the gradient and the categorical view. In such a scenario, the ERP garden-path effects would need to be consulted, because they can be based *only* on phonological (but not acoustic) processing (or else the distinction between acoustic and phonological representations would be meaningless). Here, the categorical view cannot permit more than two levels of processing difficulty (e.g., reflected by two levels of P600 amplitudes as a function of boundary size), whereas more amplitude levels would support the

[15] Note that this pattern would, in principle, also be compatible with a phonological interpretation of the CPS from a categorical perspective, if one assumes that our prosodic manipulation only resulted in variants of the ip category (as only durational manipulations were administered, as in Carlson et al., 2001). However, given the gradient behavioral data in Pauker et al. (In preparation), this theoretical option is extremely unlikely (and would cause more problems for the categorical view than it would resolve).

gradient view. If the gradient pattern found for the acceptability ratings in Pauker et al. (In preparation) reflects the immediate judgments at the lexical disambiguation point (and not second guessing at a later processing stage), then one would expect a gradient pattern of ERP garden path effects at this position as well. Moreover, the strength of these garden path effects should primarily depend on the late boundary in EC, but on the early boundary in LC.

3.2. Methods

3.2.1. Participants

Forty undergraduate students from McGill University (20 women, age range: 18–25 years) were recruited by advertisement and paid for their participation. All were right-handed (Edinburgh Handedness Inventory; Oldfield, 1971) native speakers of English with no known history of hearing impairment or brain injury. Prior to their participation, each subject signed a written informed consent. Nine subjects (4 women) were later excluded from further analysis due to excessive eye blinks and other movement artifacts exceeding 40% of the trials in one or more of the experimental conditions.

3.2.2. Materials

The 40 EC/LC sentence pairs and the thirty-two experimental conditions (16 versions per EC and LC structure) used in the current study were the same as those in Pauker et al. (In preparation, Experiment 2; see Table 2). For each of the 80 sentences, sixteen prosodic versions had been created by varying the boundary strength at both positions from a very weak ip boundary (level 1) to a strong IPh boundary (level 4). These boundary manipulations were carried out by applying a Praat script to the original recordings (which contained both boundaries at level 2) that either weakened the original boundary by subtracting 70 ms boundary duration (resulting in boundary level 1) or increased the boundary strength by adding 100 ms or 250 ms (resulting in levels 3 and 4, respectively)[16]. These manipulations affected only the duration of (a) the pre-boundary syllable (pre-final lengthening, 25%) and (b) the subsequent pause (75%), while leaving the original intonation contour (H*-L) intact. Combining the 4 size levels of the first boundary with the 4 levels of the second boundary resulted in the 16 prosodic conditions of each sentence, thus totaling 640 EC and 640 LC sentences. (For further details see Pauker et al., In preparation).

---

[16] We would like to thank Dr. Michael Wagner for developing the Praat script used for these manipulations.

Table 2. Sample stimuli for the sixteen experimental conditions

Condition	Sentence	Disambiguating region	
		EC	**LC**
	Conj NP1 VP1 B1 NP2 B2	*Would VP2*	*NP3 Would VP2*
[1_1]	Whenever the bear was approaching / the people /	…would run away	…the dogs would run away
[1_2]	Whenever the bear was approaching / the people //	…would run away	…the dogs would run away
[1_3]	Whenever the bear was approaching / the people ///	…would run away	…the dogs would run away
[1_4]	Whenever the bear was approaching / the people ////	…would run away	…the dogs would run away
[2_1]	Whenever the bear was approaching // the people /	…would run away	…the dogs would run away
[2_2]	Whenever the bear was approaching // the people //	…would run away	…the dogs would run away
[2_3]	Whenever the bear was approaching // the people ///	…would run away	…the dogs would run away
[2_4]	Whenever the bear was approaching // the people ////	…would run away	…the dogs would run away
[3_1]	Whenever the bear was approaching /// the people /	…would run away	…the dogs would run away
[3_2]	Whenever the bear was approaching /// the people //	…would run away	…the dogs would run away
[3_3]	Whenever the bear was approaching /// the people ///	…would run away	…the dogs would run away
[3_4]	Whenever the bear was approaching /// the people ////	…would run away	…the dogs would run away
[4_1]	Whenever the bear was approaching //// the people /	…would run away	…the dogs would run away
[4_2]	Whenever the bear was approaching //// the people //	…would run away	…the dogs would run away
[4_3]	Whenever the bear was approaching //// the people ///	…would run away	…the dogs would run away
[4_4]	Whenever the bear was approaching //// the people ////	…would run away	…the dogs would run away

Notes: 1) In each condition, the first number corresponds to the size of the early boundary and the second number corresponds to the size of the late boundary; 2) "/" = manipulated "no-boundary"; 3) "//" = original /ip/ boundary; 4) "///" = manipulated mid-range boundary; 5) "////" = manipulated /IPh/-compatible boundary.

To allow a precise time-locking of the ERP waveforms to the various relevant events in the speech signal (i.e., onsets and offsets of relevant words, syllables, and pauses), twelve cue points were inserted into each of the speech files of the *original recordings* of both EC and LC sentences (corresponding to boundary conditions 2_2), marking these positions with 1-millisecond accuracy relative to sentence onset. Software scripts were used to extract the exact timing values for all cue points from each individual speech file. The positions of these twelve markers are illustrated in Table 3.

For the other fifteen prosodic conditions that had undergone digital manipulations of their boundary sizes, the timing information of all cue points *following* a given boundary manipulation changed according to the type of manipulation by -70 ms (level 1), +100 ms (level 3), or +250 ms (level 4). Since these changes were constant and applied equally to all sentences as a simple function of boundary strength, it was easy to adjust the cue points' timing information for all conditions relative to sentence onset. One exception was the pause onset in conditions 3 and 4, because it was affected only by the pre-final lengthening, and thus shifted only by 25% of the total manipulation time (i.e., 25 ms instead of 100 ms for boundary level 3). In each condition, the timing information of the cue points (relative to sentence onset) was used to calculate the duration of each segment (= part of the speech signal between two cue points; typically corresponding to words, syntactic phrases, and pauses). Each segment was assigned a letter from *a* (= sentence-initial determiner) to *l* (= sentence offset). Most importantly, the cue points were later superimposed onto the recorded EEG signal in order to identify the onset of each segment and to extract the ERPs time-locked to it (for details see Pauker et al., 2011).

Table 3. Sample stimuli for each structure with cue points

Syntax	***Sentence***												
	Onset	*Conj*	*NP1*	*VP1*	*PFL1*	*P1*	*NP2*	*PFL2*	*P2*	*NP3*	*would*	*V2*	*Offset*
		a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*
LC		*Whenever*	*the girl*	*was baking*			*the cake*			*the house*	*would*	*smell good*	
EC		*Whenever*	*the girl*	*was baking*			*the cake*				*would*	*smell good*	

Notes: 1. The vertical lines mark the placement of each cue point. All cue points mark the onset of the described constituents/events. Note that EC has one less cue point, because it does not contain NP3.
2. Conj = Conjunction (Whenever).
3. PFL1 = Pre-Final Lengthening on the pre-1st boundary word.
4. P1 = First pause.
5. PFL2 = Pre-Final Lengthening on the pre-2nd boundary word.
6. P2 = Second pause.

The 1280 sentences were equally distributed across four different lists, i.e., each participant was presented with 320 experimental sentences in four blocks of 80 items. To minimize repetitions of similar sentences, each list contained exactly 4 prosodic versions of each EC and each LC sentence, one version in each block. Moreover, all versions contained exactly 10 items in all 16 prosodic conditions, both for EC and LC structures. Although this is theoretically the optimal distribution of sub-conditions and items, this initial design had one important disadvantage: it contained not enough trials that could be expected to elicit a garden path effect for in EC sentences. ERP analyses require a sufficient signal-to-noise ratio (SNR), and thus a sufficient number of 30 or more trials per condition. The original design contained 10 EC trials in each of the three conditions that showed a clear preference for LC in Pauker et al. (In preparation) and could, therefore, be expected to elicit a garden path effect for EC disambiguation (i.e., boundary combinations 1_3, 2_4, and 1_4). Even collapsing across these three conditions would potentially not provide the required SNR, especially as a number of trials must be excluded from analysis due to eye-blinks and other artifacts. For these reasons, we increased the number of trials in these three conditions from 10 to 15, and reduced the number of trials in other condition that were less problematic (especially those that showed a strong EC preference). The resulting distribution of items across prosodic conditions, which universally applies to all lists (and both EC and LC) is shown in Table 4 below (see Appendix 2 for the number of trials per condition in each experimental version). This goal was achieved by identifying individual trials with an EC-biased prosody (e.g., EC sentence # 17 with 3_1 prosody) in each list and replacing them with an LC-biased speech file of the same sentence (e.g., EC sentence # 17 with 1_4 prosody). Much care was used to ensure that (a) within subjects, these exchanges never affected any of the 40 sentences more than once, and that (b) across subjects, neither the replaced nor the replacing file was ever used twice. Note also that these adjustments were done equally for EC and LC structures, such that none of the16 prosodic patterns was associated more with one structure than with the other (which could have provided an undesirable cue towards its ultimate disambiguation). Importantly, the overall distribution of trials was still sufficiently balanced to permit ERP analyses contrasting the four levels of each boundary (see Results section for details). The behavioral data would indicate if this modification had any impact on the participants' processing strategies or prosodic discrimination skills.

Table 4. Number of experimental trials per condition

Stronger LC preference				Stronger EC preference			
Boundary1 = #1		Boundary1 = #2		Boundary1 = #3		Boundary1 = #4	
Condition	N	Condition	N	Condition	N	Condition	N
[1_1]	10	[2_1]	9	[3_1]	9	[4_1]	8
[1_2]	10	[2_2]	9	[3_2]	9	[4_2]	8
[1_3]	15	[2_3]	9	[3_3]	9	[4_3]	8
[1_4]	15	[2_4]	15	[3_4]	9	[4_4]	8
Total	50		42		36		32

The four lists were then pseudo-randomized based on five criteria: (i) degree of predicted prosodic anomaly, (ii) syntactic structure, (iii) semantic content (e.g., animals, sports, professions), (iv) prosodic structure, and (v) sentence length. The randomization rules used to create the experimental lists are displayed in Table 5 below. To control for order (sequence) effects, 4 additional lists (mirror images) were created by reversing both the block and the sentence order of each original list.

Table 5. Pseudo-randomization criteria and rules used to create the 4 original experimental lists

Criteria (starting with highest priority)	Variables	Randomizing rule
1. Degree of prosodic anomaly	(i) high; (ii) intermediate; (iii) low	No more than 3 repetitions of the same level in a row
2. Syntactic structure	(i) EC; (ii) LC	No more than 3 repetitions of the same structure in a row
3. Semantic field	8 semantic fields (e.g., sports)	No more than 2 sentences of the same semantic field on a row
4. Prosodic structure	16 prosodic conditions (4 × 4)	No more than 2 identical prosodic conditions in a row
5. Sentence length	(i) short; (ii) long	No more than 5 short or 5 long sentences in a row

3.2.3. Procedure

Participants were seated in a comfortable chair in an electro-magnetically shielded and sound-attenuated booth, approximately 80 cm in front of a computer monitor and listened to spoken sentences presented binaurally via insert-phones (Etymotic Research Inc., Elk Grove Village, IL). Subjects were instructed to press one of 5 marked keyboard keys, each representing a degree of acceptability on a continuum between “completely acceptable” and “not acceptable”, to indicate the degree of acceptability of each presented sentence (acceptability judgment task). Stimuli were presented using Presentation software (Neurobehavioral Systems), which has a timing accuracy of 0.1 ms. Each trial began when a fixation cross (i.e., “+”) appeared on the screen 1500 ms before sentence onset, and remained visible until the end of the sentence. At both sentence onset and offset, the stimulus presentation computer sent a trigger code (specifying the sentence condition) to the EEG system, marking these two events in the continuous EEG signal. Following sentence termination, a response prompt appeared on the screen (i.e., “Please rate!”) until a keyboard key was pressed or 5 seconds had elapsed (whichever came first). Next, a second prompt ("!!!") appeared on the screen for 1500 ms, indicating the time period during which subjects were encouraged to blink their eyes before the next trial began (a procedure found to significantly reduce eye blinks during sentence presentation). At the beginning of each session, participants were given a short practice block (i.e., 5 EC and 5 LC sentences with varying prosodic structure, derived from items not used in the experiment), after which further questions were clarified, if necessary. All subjects were given the same written instructions, which were presented on the screen prior to session beginning. The instructions did not contain examples of acceptable or unacceptable sentences or any other criteria for sentence scoring.

3.2.4. EEG recording

EEG was continuously recorded (500 Hz/32 bit sampling rate; Neuroscan Synamp2 amplifier) from 20 cap-mounted Ag/AgCl electrodes (Electro-Cap International, Inc., Eaton, OH) placed according to the standard International 10-20 System in the following sites: FP1, FP2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, Oz, O2. Vertical and horizontal eye-movements were monitored with bipolar EOG electrode arrays placed above and below the left eye and on the outer canthus of each eye, respectively. All EEG electrodes were referenced against the right mastoid, and an electrode placed half-way between Fpz and Fz served as the ground. Electrode impedance was kept below 5 kΩ.

3.2.5. Behavioral Data analysis

Acceptability ratings were computed by transforming the responses into a numerical scale (0 – 4) and then averaging them separately in each condition. The data were subjected to a 3-way repeated-measures ANOVA with the factors *Syntax* (2: EC vs LC), *Early Boundary* strength (4), and *Late Boundary* strength (4).

3.2.6. ERP Data analysis

The EEG data were analyzed using EEProbe software (ANT, The Netherlands). Single subject averages were computed separately for each experimental condition following filtering (0.16-30 Hz bandpass) and artifact rejection. In order to identify the EEG signals triggered by relevant segments of the speech files, a Perl script identified the individual sentence onset triggers and inserted new triggers based on the timing information of the cue points. All EEG epochs contaminated with EOG and movement artifacts exceeding a 30 μV threshold were excluded from the averaging procedure. Only the data of those 31 subjects with a minimum of 25 trials in each condition of each comparison entered the statistical analysis. ERP analyses were carried out for the CPS and the pre-CPS negativity at both boundary positions, as well as for garden-path components at the disambiguating constituents following the second boundary. ERP components were quantified in terms of amplitude averages in representative time windows. These time-windows were selected based on the related ERP literature and on visual inspection of the waveforms. Due to the complexity of the stimulus materials, ERPs had to be time-locked to various positions, and only a subset of analyses could be based on standard pre-stimulus baselines (-200 to 0 ms). Further details will be explained in the context of each analysis.

Global analyses of variance (ANOVAs) for repeated-measures were carried out for the data of each time window, separately for the four midline electrodes (Fz, Cz, Pz and Oz) and two different arrays of lateral[17] electrode sites. Analyses at two distinct lateral arrays were necessary because some effects were lateralized and required the inclusion of the most lateral electrodes (e.g., F7/8, C3/4, T5/6) for which no occipital level existed. By contrast, other ERP effects were

[17] (i)Lat12 – 12 lateral electrodes (2 levels of factor Laterality, 3 levels of factor AntPost): F7, F3, T3, C3, T5, P3, F8, F4, T4, C4, T6, P4 ; (ii) Lat8 – 8 electrodes (4 levels of factor AntPost): F3, C3, P3, O1, F4, C4, P4, O2.

maximal at occipital electrodes (O1, Oz, O2), for which no lateral electrode level existed. Statistical analyses for the ERP data were carried out either across syntactic conditions, i.e., collapsing across EC and LC conditions (especially for the CPS) or separately for EC and LC structures, such that (unlike in the behavioral analysis) the factor 'Syntax' was not a relevant factor. The ANOVAs for ERP data from the lateral electrodes included the two condition factors: Early Boundary (early boundary size), and Late Boundary (late boundary size), as well as the three topographical factors: Anterior-Posterior (AntPost), Hemisphere (Hemi), and – for the array with 12 electrodes – Laterality (Lat). The ANOVAs for the ERP data from the midline electrodes included the same factors except for the topographical factors of Hemisphere and Laterality. Additional ANOVAs followed up on main effects with more than 1 degree of freedom and significant interactions. In order to avoid violations of sphericity (Type 1 error) the Greenhouse-Geisser correction was applied to all repeated measures with more than one degree of freedom in the numerator, in which case we report the original degrees of freedom and the corrected p-values.

3.3. Results

3.3.1. Behavioral Data

The behavioral data largely replicated the results of Pauker et al. (In preparation), both in terms of main effects and interactions. Differences in acceptability judgments across conditions were tested using an ANOVA with within-subject factors *Syntax (2)* × *Early Boundary (4)* × *Late Boundary (4)*. The scores for each condition in EC and LC and their corresponding percentages are summarized in Table 6. Differences in acceptability judgments between EC and LC (EC minus LC) in each prosodic condition are presented in Figure 1, showing the same gradient pattern found in our behavioral experiments. Most importantly, in line with the eBDH, we again found that the acceptability scores were primarily driven by the strength of the *incompatible* boundary, in both EC and LC.

3.3.2. Main effects

As in our previous study, we found a significant main effect of Syntax [$F(1,30) = 29.2$, $p<.0001$], which reflected the general EC preference across conditions (see Table 6). We also replicated the previous finding of Exp 2 showing that LC was *more acceptable* than EC in condition [1_4], whose pattern of prosodic phrasing was most compatible with LC phrasing, as illustrated by the negative bar in Figure 1. Conditions [1_3] and [2_4], which were preferred as LC in Exp2 (but not Exp1) of Pauker et al. (In preparation), received nearly identical scores in both structures.

Table 6. Averages and corresponding percentage of conditions acceptability in EC and LC

Condition	*4_1*	*4_2*	*4_3*	*4_4*	*3_1*	*3_2*	*3_3*	*3_4*	*2_1*	*2_2*	*2_3*	*2_4*	*1_1*	*1_2*	*1_3*	*1_4*	*Mean*
EC	2.97	3.05	2.70	2.29	3.13	2.86	2.62	2.23	3.16	2.96	2.63	2.19	2.94	2.85	2.53	2.09	2.70
%EC (EC/4)	74	76	68	57	78	72	66	56	79	74	66	55	73	71	63	52	68
LC	1.30	1.35	1.46	1.35	1.53	1.71	1.71	1.66	1.91	2.23	2.15	2.16	2.38	2.50	2.47	2.58	1.90
%LC (LC/4)	32	34	36	34	38	43	43	41	48	56	54	54	59	63	62	65	48
EC+LC	4.26	4.40	4.16	3.64	4.66	4.57	4.33	3.88	5.07	5.19	4.79	4.35	5.32	5.35	5.00	4.67	
%EC (EC/EC+LC)	70	69	65	63	67	63	61	57	62	57	55	50	55	53	51	45	59

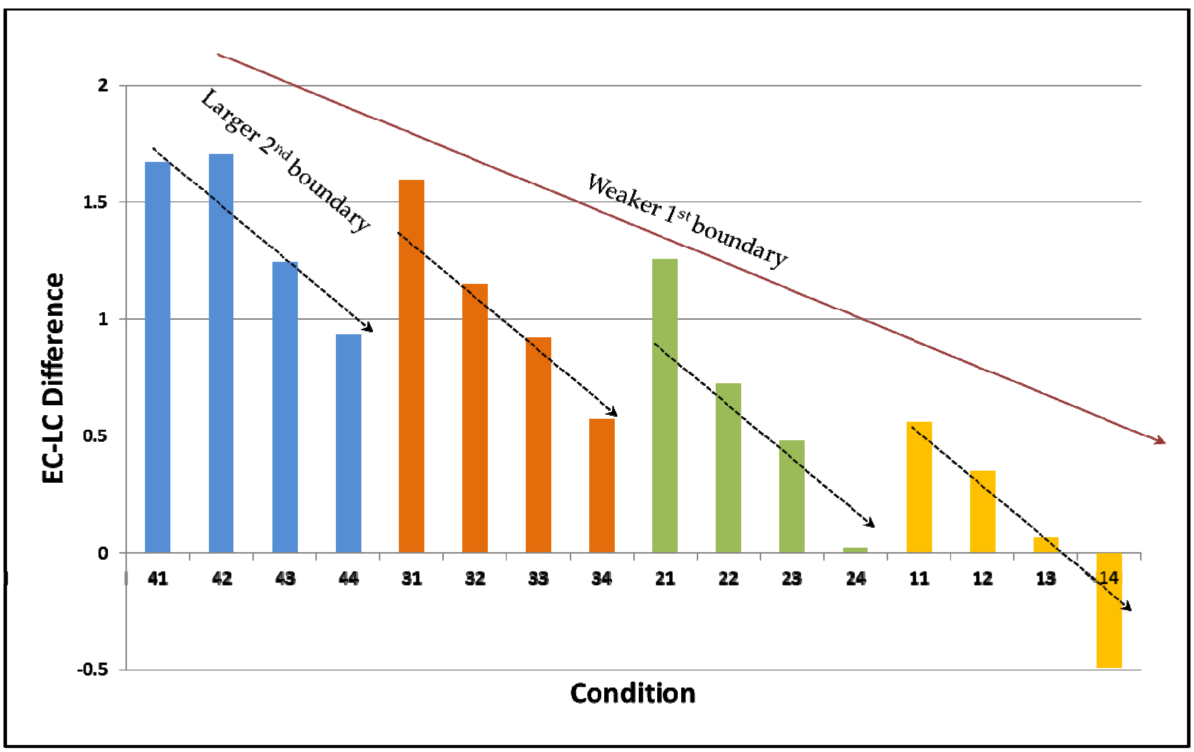


Figure 1. Difference in acceptability between EC and LC structures as a function of prosodic boundary size. X-axis: Displays the prosodic pattern for each condition (e.g., **4_1**: Boundary #1 = IPh, Boundary #2 = "no-boundary"; **2_2**: Boundary #1 = ip, Boundary #2 = ip; **1_4**: Boundary #1 = "no-boundary", Boundary #2 = IPh). Y-axis: EC acceptability minus LC acceptability for each prosodic structure.

We also replicated the main effects of Early Boundary [F(3,90) = 30.1, p<.0001] and Late Boundary [F(3,90) = 23.5, p<.0001]. To illustrate these effects, we placed the total values of each condition in a scores matrix (see Table 7), and found again that as the boundary size increased, the overall acceptability decreased in a graded manner (1 > 2 > 3 > 4). This was the case for Boundary 1 (downward pointing arrow) and partially so for Boundary 2 (leftward pointing arrow) where, however, the scores for level 1 tended to be slightly lower than those for level 2. This deviation from the overall robust pattern stems from the scores of Late Boundary size 1 in LC (see also section 2.1.2 below). Follow-up t-tests showed that the scores of all boundary sizes differed highly significantly from one another (Early Boundary: [p's < .0001; 1 vs. 2 and 3 vs. 4 (p<.01)]; Late Boundary:[p's < .0001; 1 vs. 3 (p<.01)]), with the exception of boundary sizes 1 vs. 2 of factor Late Boundary, whose ratings did not differ from each other (p>.1). Since each prosodic boundary was compatible with one syntactic structure and incompatible with the other one, we expected these Boundary main effects would be qualified by an interaction with factor Syntax, which was indeed the case (see section 3.3.3 below).

Table 7. Total acceptability scores for each prosodic condition (EC+LC) as a function of boundary position.

Boundary 1		Boundary 2: 1	2	3	4
	4	4.26	4.40	4.16	3.64
	3	4.66	4.57	4.33	3.88
	2	5.07	5.19	4.79	4.35
	1	5.32	5.35	5.00	4.67

3.3.3. Interactions

As expected, we found highly significant interactions of Syntax × Early Boundary [F(3,90) = 56.3, p<.0001] and Syntax × Late Boundary [F(3,90) = 35.0, p<.0001], indicating that the main effects of Early and Late Boundary did not apply equally to the two syntactic structures. The follow-up analysis in LC revealed a very strong main effect of Early Boundary [F(3,90) = 60.7, p<.0001] and, surprisingly, a small but significant main effect of Late Boundary [F(3,90) = 4.0, p<.02), which was not found in our previous experiments. In EC, on the other hand, in accordance with Pauker et al. (In preparation), we found a main effect of Late Boundary [F(3,90) = 47.5, p<.0001] but not of Early Boundary (p>.1). The pairwise comparisons confirmed that the scores in each structure replicated the graded pattern found in Pauker et al. (In preparation), as illustrated in Figure 2. In LC (Fig. 2, right panel), all levels of Early Boundary size were significantly different from one another (all p's < .0001), and exhibited incrementally larger scores for the smaller boundary sizes (1 > 2 > 3 > 4). In the Late Boundary follow-up analysis, we found that the scores for levels 2, 3, and 4 were indistinguishable (p>.8), whereas level 1 received significantly lower scores than all other levels ([1 vs. 2]: p<.01; [1 vs. 3]: p<.03; [1 vs. 4]: p<.04). This pattern is also reflected by a significant interaction of Syntax × Early Boundary × Late Boundary [F(9,270) = 2.3, p<.04], which was absent (p>.4) when we analyzed the data without Late Boundary level 1. In EC, the scores of all Late Boundary size levels differed significantly from each other, including levels 1 and 2 that did not differ in our previous study (p < .02 for 1 vs. 2; all other p's < .0001).

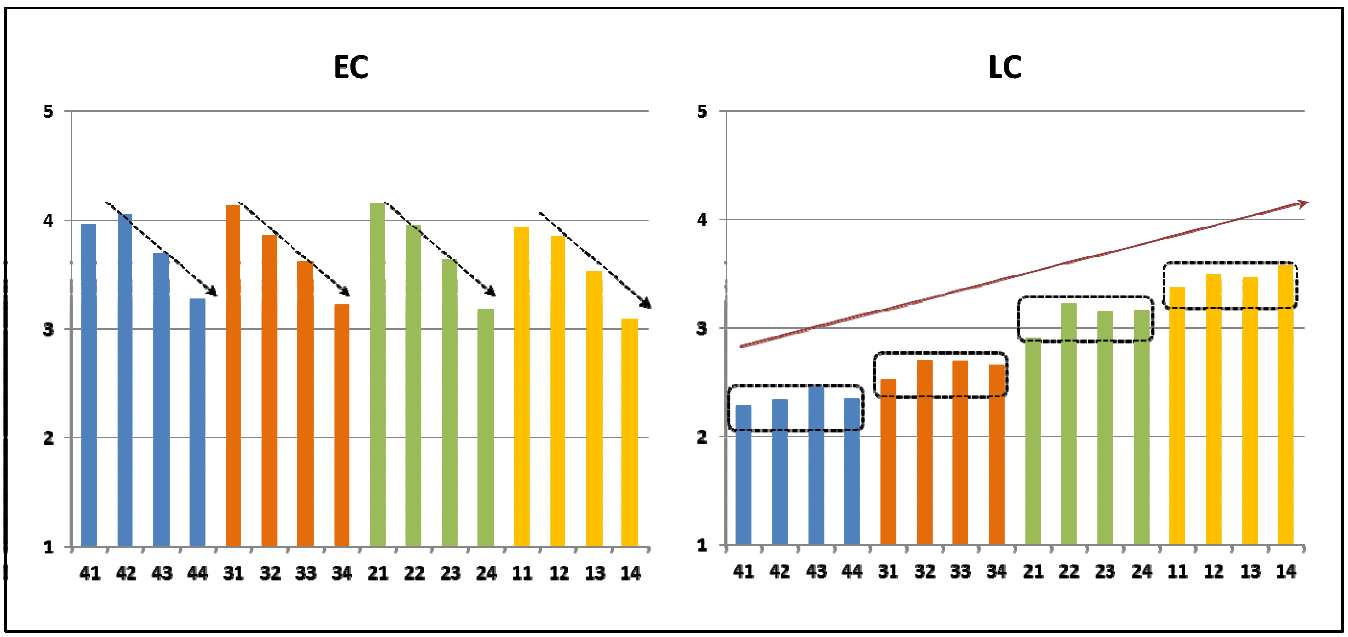


Figure 2. Acceptability scores per condition in EC and LC as a function of prosodic boundary size

3.4. ERP Data

As argued in Pauker et al. (2011; see also Kerkhofs et al., 2007), due to the variability in both (i) the duration of the critical (here: prosodically manipulated) regions in each condition and (ii) the constituent and word length across sentences, it is usually not possible to accurately examine auditory ERPs by time-locking them to the sentence onset. Instead, ERPs need to be time-locked to the relevant events in the speech signal. Because we predicted CPS components at each boundary location and garden-path effects at the disambiguating region of each structure, we used time-locking positions relative to these specific events. Since the majority of ERP components were larger over medial and posterior sites, we typically report the results for the midline electrodes and an array of medial lateral electrodes (excluding electrodes F7, F8, T3, T4, T5, T6; see Methods section). In one case where an effect was prominent at more lateral electrodes, we present the analysis conducted for that region (using an alternative array of lateral electrodes). In general, before we present statistical analyses of the respective ERP effects, we will briefly describe both specific challenges for the respective ERP analyses as well as our choice of time-locking points and baselines. Moreover, any methodological issues that are directly related to the quantification of, and distinction between, ERP components will also be discussed within the Results section, allowing for immediate reference to the corresponding ERP plots. In contrast, the Discussion section will primarily focus on the implications of our findings for psycholinguistic theories of boundary processing.

3.4.1. ERP effects at the early boundary position

At the first boundary position, EC and LC conditions share the same speech signal and differ only in terms of boundary size. Therefore, in order to quantify the effect of the prosodic manipulation at that region, we collapsed data across both syntactic structures (and irrespective of the strength of the second boundary) to create four new conditions: [B1_1], [B1_2], [B1_3], and [B1_4] (where B1 stands for Boundary1 followed by the boundary size level). For example, condition [B1_1] was created by collapsing the following prosodic conditions (across EC and LC): [**1**_1], [**1**_2], [**1**_3], [**1**_4] – all sharing the same early boundary size. Figure 3 and Figure 4 illustrate the ERPs for conditions [B1_1] through [B1_4], time-locked to the onset of NP2 ("***the*** people") following the first boundary, and Figure 5 illustrates the ERPs for conditions [B1_2] through [B1_4] time-locked to the onset of the pre-final lengthening on the pre-boundary word ("approach***ing***").

< Figure 3 here >

The first analysis, time-locked to the onset of post-boundary NP2, was important to reliably distinguish the CPS from other positive-going waveforms such as the onset P200 of the NP2. Onset P200 components reflect primarily the physical characteristics of a stimulus, including its relative visual or acoustic contrast against the respective 'background'. Thus auditory P200s are known to increase in amplitude after a period of silence (including after prosodic boundaries) and can, therefore, be mistaken for CPS components (see Kerkhofs et al., 2007; Männel & Friederici, 2009; Pauker et al., 2011; Steinhauer, 2003). Crucially however, whereas the CPS has been shown to be elicited early *during* the pause (likely triggered by pre-final lengthening of the pre-boundary word), onset-P200s of post-boundary words are not elicited until *after* the pause, i.e., approximately 200 ms after onset of the corresponding post-boundary word.

ERP plots of this analysis are displayed in Figures 3 and 4 between -600 and 1000 ms relative to NP2 onset. Since the four conditions differ immediately prior to this point due to our prosodic boundary manipulations, a standard pre-target baseline interval would have affected the ERPs and was not an option. Instead, by investigating the ERPs of all conditions from sentence onset, we determined the optimal time interval for a 'distant baseline' that preceded the prosodic manipulations without distorting the signals. This interval was identified between -520 and -320 ms relative to the onset of NP2 in all four conditions[18]. Recall that the underlying speech signals of the four conditions differed only with respect to the prosodic manipulations (i.e., amount of pre-final lengthening and pause duration) administered to the preceding word (VP1). Therefore, any ERP differences appearing *before* the vertical line at 0 ms should reflect the influence of the strength of each boundary size. By contrast, any differences related to the onset P200 component of the post-boundary word (NP2) should appear after the vertical line at 0 ms. Figure 3 shows a broadly distributed CPS in all four conditions, slightly right lateralized, with a posterior amplitude maximum, starting around -300 ms and spanning about 800 ms before returning to baseline. Importantly, its amplitude increases with increasing boundary size. With some variance in its distribution characteristics, the present CPS bears the same characteristics as CPS components elicited by naturally produced IPh boundaries, both in terms of latency and duration (Steinhauer et al., 1999; Toepel, Pannekamp, & Alter, 2007; but see Kerkhofs et al., 2007; Pannekamp et al., 2005; Pauker et al., 2011 for reports of CPS components with a somewhat more frontal distribution). This finding was a first indication that listeners processed our

[18] This particular interval was initially determined based on the administered prosodic manipulations: the longest manipulation in [B1_4] (IPh) consisted of 250 ms and the shortest manipulation in [B1_1] ("no-boundary") consisted of -70 ms (i.e., 250 + 70 = 320). In a second step, we plotted ERPs relative to an earlier trigger point to confirm that there were no systematic ERP differences across conditions.

digitally-generated prosodic boundaries similarly to more naturally-generated prosodic boundaries. To date, only one other study we are aware of reported a CPS to a boundary smaller than an IPh (Li & Yang, 2009), demonstrating a CPS for ip boundaries in Chinese, which *did not differ statistically* from a CPS elicited by an IPh boundary. In contrast, here we observe a *graded pattern* of results where both the amplitude and duration of the CPS increase as the boundary size increases, most prominent over posterior sites (see Table 8). As anticipated, a more frontally distributed P200 component appearing around 200 ms post NP2 onset superimposes the ongoing CPS wave, also showing an increasingly larger amplitude in each condition, which reflects the increase in physical contrast (silence versus word onset) with increasing boundary (and pause) size. As a result, the CPS appears to have two peaks after the vertical 0 line. In order to better illustrate the difference between early CPS and post-boundary onset P200s, we created difference waves by subtracting the ERPs of condition [B1_1] (which contained no pause or pre-final lengthening) from those of conditions [B1_2], [B1_3], and [B1_4]. These difference waves are displayed in Figure 4 at occipital electrodes and clearly show that the positive shift of the CPS precedes both the onset of NP2 as well as that of the P200 and peaks around 0 ms, whereas the additional P200 differences have a latency of 200 ms post-onset. This distinction was also useful in selecting an optimal time window to quantify the CPS (from -300 to 100 ms) that is not influenced by the P200.

< Figure 4 here >

3.4.1.1. *Statistical analysis of the CPS1 at NP2*

Table 8 summarizes the analyses for the -300 to 100 ms time window, where a significant main effect of factor BSize and an interaction of BSize × AntPost at both the lateral and midline electrodes, as well as a three-way interaction of BSize × AntPost × Hemi at lateral electrodes reflect the differences between the four conditions in the global ANOVAs. Follow-up comparisons revealed that the effect of BSize was largest over posterior sites, most prominent over the occipital region. Table 9 illustrates the mean amplitude in each condition at level 4 of factor AntPost, i.e., at electrodes O1, O2, and Oz (corresponding to Figure 4). Pairwise comparisons revealed that of the six possible pairwise comparisons, five were significantly different from one another (conditions [B1_1] and [B1_2] being the exception, were statistically indistinguishable from one another), either overall, or at more posterior regions. Finally, the three-way interaction seems to point towards slight differences in the AntPost distribution of the right and left hemispheres. However, follow-up analyses at each level of AntPost did not reveal any consistent differences.

Table 8. CPS1 effect across all four conditions (-300..100 msec relative to NP2 onset)

		CPS1 *(-300..100 ms)*	
		Global ANOVA + Follow-ups	
Source	*df*	*F*	*p*
Lateral electrodes			
BSize	***3, 90***	***5.72***	***<.01***
BSize × AntPost	***9, 270***	***7.79***	***<.0001***
AntPost3 (P3, P4)	*3, 90*	*9.8*	*<.0001*
AntPost4 (O1, O2)	*3, 90*	*16.65*	*<.0001*
BSize × AntPost × Hemi	***9, 270***	***2.73***	***<.02***
AntPost2 × Hemi1	*3, 90*	*3.03*	*<.04*
AntPost3 × Hemi1	*3, 90*	*8.34*	*<.0001*
AntPost3 × Hemi2	*3, 90*	*9.53*	*<.0001*
AntPost4 × Hemi1	*3, 90*	*14.96*	*<.0001*
AntPost4 × Hemi2	*3, 90*	*16.7*	*<.0001*
Midline electrodes			
BSize	***3, 90***	***6.08***	***<.001***
BSize × AntPost	***9, 270***	***5.89***	***<.001***
AntPost2 (Cz)	*3, 90*	*3.16*	*<.03*
AntPost3 (Pz)	*3, 90*	*10.35*	*<.0001*
AntPost4 (Oz)	*3, 90*	*14.94*	*<.0001*

		Pairwise comparisons									
		1vs3		***1vs4***		***2vs3***		***2vs4***		***3vs4***	
		F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Lateral electrodes											
BSize	*1,30*	*3.78*	*.061*	*7.17*	*<.02*	*5.9*	*<.03*	*10.95*	*<.01*	*---*	*---*
BSize × AntPost	*3,90*	*6.52*	*<.01*	*7.98*	*<.01*	*9.94*	*<.001*	*15.19*	*<.0001*	*---*	*---*
AntPost3 (P3, P4)	*1,30*	*6.07*	*<.02*	*12.09*	*<.01*	*9.82*	*<.01*	*19.4*	*<.001*	*4.16*	*.05*
AntPost4 (O1, O2)	*1,30*	*20.32*	*<.0001*	*20.87*	*<.0001*	*19.82*	*<.001*	*24.45*	*<.0001*	*3.49*	*.072*
BSize × AntPost × Hemi	*3,90*	*---*	*---*	*---*	*---*	*5.24*	*<.01*	*4.13*	*<.02*	*---*	*---*
AntPost2 × Hemi1	*1,30*	*---*	*---*	*---*	*---*	*---*	*---*	*6.42*	*<.02*	*---*	*---*
AntPost3 × Hemi1	*1,30*	*---*	*---*	*---*	*---*	*9.56*	*<.01*	*14.59*	*<.001*	*---*	*---*
AntPost3 × Hemi2	*1,30*	*---*	*---*	*---*	*---*	*8.99*	*<.01*	*20.1*	*<.0001*	*4*	*.055*
AntPost4 × Hemi1	*1,30*	*---*	*---*	*---*	*---*	*15.71*	*<.001*	*20.23*	*<.0001*	*3.24*	*.082*
AntPost4 × Hemi2	*1,30*	*---*	*---*	*---*	*---*	*23.35*	*<.0001*	*26.71*	*<.0001*	*3.23*	*.083*
Midline electrodes											
BSize	*1,30*	*4.99*	*<.04*	*7.58*	*<.01*	*6.32*	*<.02*	*11.57*	*<.01*	*---*	*---*
BSize × AntPost	*3,90*	*4.04*	*<.04*	*7.7*	*<.01*	*6.03*	*<.01*	*11.25*	*<.001*	*3.01*	*.051*
AntPost2 (Cz)	*1,30*	*---*	*---*	*3.81*	*.06*	*3.05*	*.091*	*7.23*	*<.02*	*---*	*---*
AntPost3 (Pz)	*1,30*	*7.78*	*<.01*	*14.1*	*<.001*	*9.03*	*<.01*	*19.14*	*<.001*	*4.59*	*<.05*
AntPost4 (Oz)	*1,30*	*17.88*	*<.001*	*22.4*	*<.0001*	*17.01*	*<.001*	*22.01*	*<.0001*	*---*	*---*

Table 9. CPS1 - Amplitude means (μV) at Occipital electrodes (-300..100 msec relative to NP2 onset). Standard deviations are in brackets.

		Electrode	
Condition	***O1***	***Oz***	***O2***
[B1_1]	*-0.38 (0.76)*	*-0.20 (0.63)*	*-0.42 (0.82)*
[B1_2]	*-0.44 (1.17)*	*-0.22 (0.89)*	*-0.50 (1.09)*
[B1_3]	*0.41 (0.90)*	*0.64 (0.95)*	*0.40 (1.03)*
[B1_4]	*0.76 (1.13)*	*1.00 (1.12)*	*0.80 (1.16)*

Note. Standard deviations are in indicated in parentheses.

3.4.1.2. Pre-CPS negativity

As can already be seen in Figure 3, the CPS was preceded by a negative deflection that also seemed to vary across conditions. A number of previous studies reporting CPS components have mentioned this biphasic ERP pattern and have referred to the negativity as a 'pre-CPS' negativity (Bögels et al., 2010; Pauker et al., 2011). The majority of previous experiments analyzed these components (or just the CPS) either relative to sentence onset (Pannekamp et al., 2005; Steinhauer et al., 1999) or further downstream (typically at or after the pause; see e.g., Kerkhofs et al., 2007; Pauker et al., 2011) and were not able to determine the onset of the negativity with much precision. However, in a recent study we hypothesized it might be triggered by the pre-final lengthening of the last syllable preceding the boundary (Pauker et al., 2011) . In our present study we were able to use a time-locking position directly at the pre-final lengthening on the last syllable of VP1 (henceforth referred to as ***PFL1***), which was meticulously marked to create the prosodic manipulations in each sentence. Although this position is not ideal to examine the CPS itself, partly because it is followed by much variability in duration among conditions (resulting in a different latency of the CPS in each condition), it does allow us to better inspect and characterize the pre-CPS negativity. Unlike the CPS analyses at NP2, the time-locking to PFL1 was compatible with a standard pre-target baseline (-300 to 0 ms).

Figure 5 displays the ERPs for boundary sizes 2 through 4. We can see that in line with previous accounts (Bögels et al., 2010; Pauker et al., 2011), the positive deflection of the CPS (now showing the expected time-shifts due to pre-final lengthening) is preceded by a broadly distributed negativity most prominent at frontal and central electrodes of the right hemisphere, emerging about 150 ms relative to PFL1 onset (at 0 ms) and peaking around 300 ms (in condition [B1_4]) before shifting into the positive range of the CPS. Importantly, as with the CPS, here too we found a graded effect of boundary size, exhibiting increasingly larger – negative – amplitudes for larger boundary sizes. The only exception was condition [B1_1] (not shown), which appeared more negative in comparison to conditions [B1_2] and [B1_3] at certain electrodes. However, recall that [B1_1] is the only condition containing no silence interval (condition [B1_2] contains a minimal, naturally produced [~]50 ms pause typical for other ip boundaries described in literature; see Carlson et al., 2001), and no pre-final lengthening (as it was substantially shortened). For this reason, in this condition the cue points marking the onset of PFL1, pause1 and NP2, almost coincide. Since we have already established that each condition elicited a CPS preceding NP2, we can determine that the effect we observe in B1_1 time-locked to PFL1 is in fact a (small) CPS, which, due to the differences in duration between conditions, is elicited the earliest and overlaps with the negativities seen in the other three conditions. Since it contains the smallest boundary, the CPS in condition B1_1 also has the smallest amplitude and, therefore, appears more negative than some of the conditions. In order to avoid any confounds between these distinct effects, we excluded condition B1_1 from the pre-CPS1 analysis.

< Figure 5 here >

Based on the plots in Figure 5, it seems that the magnitude of the pre-CPS negativity reflects the amount of the pre-final lengthening, given that with increasing boundary size each condition (B1_2 - B1_4) exhibits an increasingly larger negativity. However, a potential concern is that the observed effect is influenced by the delay in CPS onset latency between conditions. In contrast to previous studies, we have employed a systematic duration manipulation exactly at the time-locking point where we chose to quantify the pre-CPS negativity. Therefore, depending on how the components are quantified, there may exist a confound between the CPS-related positivity in some conditions and pre-CPS-related negativity in others. For example, the largest negativity is found in condition B1_4, which also exhibits the largest positivity. If the CPS is viewed as the beginning of its positive slope (as typically done in CPS studies), then the negativity and the CPS would need to be quantified together. Thus, the combined effect may simply reflect the overall size of the CPS rather than two distinct processes, rendering the two components virtually impossible to tease apart. Moreover, it may seem as if the CPS components in Figure 5 do not show a clear graded pattern anymore; however, this impression is primarily due to latency differences, i.e, the CPS in B1_4 is now aligned with the P200 component of B1_3 (rather than with its CPS), and so forth. These observations will be relevant for the quantification of ERP boundary effects in future studies. However, one key observation seems to suggest that the negativity may be a distinct component. Specifically, roughly 120 ms after the vertical 0 line, all conditions exhibit a shared negativity, following which they begin to diverge. We can see CPS1 in condition B1_2 directly following this early negativity. Condition B1_4, on the other hand, shows a distinct negative deflection, which is roughly the same size as (or even larger than) the preceding shared negativity (see also Figure 6 for an illustration of these effects at right frontal electrodes where the negativity was found to be largest). Both its early onset and its more anterior distribution compared to the CPS strongly suggest that the pre-CPS negativity and the CPS are distinct, although tightly linked ERP components.

3.4.1.3. Statistical analyses of the pre-CPS negativity at PFL1

Given that the region of anterior-lateral electrodes where this component tends to be most prominent are not included in our standard array of electrodes, we report analyses for the wider array of electrodes, which includes factor Laterality. Table 10 summarizes the results for the 250 to 450 ms time window (relative to PFL1), where we found a main effect of BSize and an interaction of BSize × AntPost, in both lateral and midline electrodes. Follow-up comparisons revealed all levels of BSize were significantly different from one another, confirming a graded effect. However, the difference between B1_4 and the other conditions was very large and broad, as reflected by a main effect of BSize and a lack of interactions, while the difference between conditions B1_2 and B1_3 was smaller and only significant over the frontal (electrodes F4 and F8; [$F(1,30) = 5.61$, $p<.03$]) and less so over the central (electrodes C4 and T4; [$F(1,30) = 3.95$, $p=.056$]) regions of the right hemisphere, as reflected by an additional interaction of BSize with factors AntPost × Hemi in the global ANOVA. The distribution profile found here resembles the one reported in previous studies, where the pre-CPS negativity was also right-lateralized, albeit more prominent over medial sites (Bögels et al., 2010; Pauker et al., 2011).

Table 10. Pre-CPS1 negativity effect in boundary sizes 2, 3, and 4 (250..450 msec relative to PFL1 onset)

		Pre-CPS1 negativity *(250..450 ms)*	
		Global ANOVA + Follow-ups	
Source	*df*	*F*	*p*
Lateral electrodes			
BSize	*2,60*	*11.86*	*<.0001*
BSize × AntPost	*4,120*	*3.48*	*<.04*
AntPost1 (F3, F4, F7, F8)	*2,60*	*7.45*	*<.01*
AntPost2 (C3,C4,T3, T4)	*2,60*	*14.77*	*<.0001*
AntPost3 (P3, P4,T5, T6)	*2,60*	*10.12*	*<.001*
Midline electrodes			
BSize	*2,60*	*7.49*	*<.01*

		Pairwise comparisons					
		2vs3		***2vs4***		***3vs4***	
		F	*p*	*F*	*p*	*F*	*p*
Lateral electrodes							
BSize	*1,30*	---	---	*21.15*	*<.0001*	*14.65*	*<.001*
BSize × AntPost	*2,60*	*8.95*	*<.01*	---	---	---	---
AntPost1 (F3, F4, F7, F8)	*1,30*	*3.07*	*.089*	---	---	---	---
AntPost2 (C3,C4,T3, T4)	*1,30*	---	---	---	---	---	---
AntPost3 (P3, P4,T5, T6)	*1,30*	---	---	---	---	---	---
Midline electrodes							
BSize	*1,30*	----	----	*12.25*	*<.01*	*9.55*	*<.01*
BSize × AntPost	*2,60*	*7.26*	*<.01*	---	---	---	---
AntPost2 (Fz)	*1,30*	---	---	---	---	---	---
AntPost2 (Cz)	*1,30*	---	---	---	---	---	---
AntPost3 (Pz)	*1,30*	---	---	---	---	---	---

3.4.2. ERP effects at the late boundary position

3.4.2.1. CPS and pre-CPS negativity at the late boundary position

The initial analyses for the ERP effects at the late boundary followed the same logic as those at the early boundary. Again, we collapsed across all EC and LC conditions whose *late* boundary sizes were the same, thereby creating four new conditions: [B2_1], [B2_2], [B2_3], and [B2_4] (where B2 stands for Boundary2 followed by the boundary size; e.g., condition [B2_4] was comprised of conditions [1_**4**], [2_**4**], [3_**4**], [4_**4**]). This allowed an inspection of the events related to the late boundary while counterbalancing the effects at the early boundary across conditions. As with CPS1, we expected CPS2 to be triggered by events time-locked to the pre-final lengthening of NP2 ("the peo***ple***"; henceforth referred to as PFL2). As CPS2 is triggered by the exact same speech signal in both EC and LC, it should exhibit the same effect in both structures. Moreover, collapsing across EC and LC increased the statistical power of the analyses.

The ideal time-locking position for CPS2 and the garden-path components was the onset of the post-boundary disambiguating word (*Would* in EC and *NP3* in LC). Similar to the CPS1 analysis, this position allowed us to tease apart the CPS effect of interest and the subsequent onset P200s of the post boundary words.

To examine all structures while avoiding the duration shift between late boundary conditions, it was again necessary to use a baseline interval preceding the prosodic manipulations at PFL2. In contrast to the analyses for CPS1, no shared time interval relative to the time locking point (i.e., the post-boundary word) could be identified that met the criteria of showing minimal variability across conditions. This was likely a result of enhanced variability due to the early boundary effects. An alternative way of establishing a distant baseline is to time-lock the baseline interval to a *shared segment* preceding the prosodic manipulation at PFL2 ('reference displacement technique', e.g., Friederici, Pfeifer, & Hahne, 1993). We identified a time window from -500 to -100 ms *relative to PFL2* as appropriate for our purposes. Note that this baseline interval varies across conditions in its temporal distance to the post-boundary word (i.e., it is furthest away from it in condition B1_4). One major advantage of this approach is that durational differences between conditions (for the distance between the baseline interval and the event of interest to which the ERPs are time-locked) do not affect the ERPs. As systematic durational differences (due to the prosodic manipulations) are exactly the challenge in our materials, this approach is highly appropriate. On the other hand, a serious potential risk of this technique is that any intervening slow waves occurring between the baseline interval and the target event are not compensated for and will affect the waveforms at the target event. Fortunately, the only slow waves to be expected after PFL2 onset are the pre-CPS negativities, which (a) are part of the ERP effects of interest, and (b) can this way be visualized by increasing the average window into the negative time range (i.e., towards the baseline interval).[19]

< Figure 7 here >

Figure 7 depicts the ERPs of all four conditions between -600 and 1000 ms, time-locked to the onset of the post-boundary word. We can see that the fronto-central pre-CPS2 negativity

[19] To examine whether this technique was acceptable, we first used a second distant baseline further up-stream, i.e., relative to sentence onset, where all sentence conditions were supposed to be identical by definition. In an average from sentence onset, we tested the homogeneity of waveforms at our intended distant baseline window (i.e., directly preceding PFL2). After finding that this was largely the case, we set the most appropriate distant baseline relative to this event. Note that we could not place the distant baseline too far upstream from the time-locking point as this would have introduced additional noise, decreasing the size of the effects and even distorting them.

varies in latency between conditions (as it is triggered by, but not time-locked to, PFL2) and appears between -500 and -150 ms. It is followed by CPS2, starting at about -150 ms in condition B4 and at the vertical zero line in condition B1. It is superimposed by the onset P200, and then both return to baseline around 350 ms. As was observed at the early boundary position, the latency and duration of these components depend on the size of the boundary. Similar to CPS1, CPS2 was elicited prior to word onset, whereas the P200 is evoked following word onset (see Figure 8).

< Figure 8 here >

3.4.2.2. Statistical analyses of CPS2 and the pre-CPS negativity

Quantifying CPS2 in a manner similar to that used for CPS1, namely analyzing a time-window before the onset of the P200 (in this case, 0 to 100 ms) relative to the onset of the post-boundary word, revealed rather limited outcomes. This was likely due to both (a) a higher loss of trials in this particular analysis and (b) the fact that the distant baseline, although carefully selected, may still have added some noise to the data (see EC N400 garden-path effects below for some additional evidence). On the other hand, a visual inspection of the pre-CPS2 negativity and the CPS2 indicated that the shapes and graded patterns of the effects were similar to the ones observed at the early boundary. To overcome this challenge, we decided to quantify CPS2 and its preceding negativity independent of the baseline. Specifically, we tested whether the pre-CPSs negativity and the positive shift of the CPSs *combined* would show the same strong graded pattern found at CPS1. Since these two components (a) belong to a biphasic pattern that clearly begins prior to the onset of the post-boundary word and (b) have opposite polarities and were hypothesized to exhibit a similar graded order in terms of their amplitudes (i.e., larger amplitudes with increasing boundary size), contrasting the two components across boundary sizes was expected to reveal if these boundary effects were systematically influenced by the prosodic manipulation. Importantly, such an analysis has the advantage of avoiding any confounds with previous shifts in the data and would also eliminate any problems due to the distant baseline selection. To execute this analysis, we first identified two consecutive time-windows capturing the pre-CPS2 negativity (Time Window 1, or TW1) and the beginning of the CPS2 before the P200 component (Time Window 2, or TW2). TW1 was set between -400 and -150 ms and TW2 was set between -150 and 100 ms – each comprising 250 ms. We then subtracted the mean amplitude of TW1 from TW2 in each condition, and found that – numerically – the amplitude difference was indeed graded in the expected way (see Table 11).

Table 11. CPS2 - Amplitude means (μV) at Occipital electrodes using a time window analysis (-400..-150 msec vs. -150..100 msec relative to the onset of the disambiguating words – would/NP3)

		Electrode	
Condition	***O1***	***Oz***	***O2***
[B2_1]	*-0.84 (1.01)*	*-0.91 (1.03)*	*-0.84 (1.05)*
[B2_2]	*-0.76 (0.98)*	*-0.85 (0.97)*	*-0.81 (0.95)*
[B2_3]	*0.26 (0.96)*	*0.43 (0.88)*	*0.33 (1.01)*
[B2_4]	*0.92 (1.11)*	*0.98 (1.10)*	*1.05 (1.22)*

Note. Standard deviations are in indicated in parentheses.

To demonstrate that this pattern was also statistically meaningful, two outcomes were essential: First, we expected to find a main effect of TW, as the pre-CPS negativity in TW1 should be overall more negative than the CPS in TW2. Second, it was essential to find a significant interaction of TW × BSize, reflecting the differences between TW1 and TW2 across conditions, which is comparable to a BSize main effect for either of the two components. The overall amplitude difference between the time-windows should systematically increase from [B2_1] to [B2_4]. Finally, a BSize main effect (if found), reflecting an overall difference between conditions – regardless of time-window – was predicted to be smaller than the two-way interaction. The global ANOVAs at both lateral and midline electrodes showed the same pattern of results: a main effect of TW (lateral: [$F(1,30) = 11.66$, $p<.01$]; midline: [$F(1,30) = 9.54$, $p<.01$]) and a highly significant interaction of TW × BSize (lateral: [$F(3,90) = 26.93$, $p<.0001$]; midline: [$F(3,90) = 32.04$, $p<.0001$]), which were both larger than a main effect of BSize (lateral: [$F(3,90) = 5.02$, $p<.01$]; midline: [$F(3,90) = 2.93$, $p<.04$]). Similar to CPS1, the effect was most prominent at midline electrodes; therefore, here we report the results for the midline array only (see Table 12). As for CPS1, we found significant differences between all pairwise comparisons, save for conditions [B2_1] and [B2_2], which were statistically indistinguishable. A 3-way interaction of TW × BSize × AntPost revealed that, similar to CPS1, CPS2 also showed an increase in amplitude over the parieto-occipital region. Follow-up analyses confirmed the effect was significant in each of the pairwise comparisons except [B2_1] vs [B2_2].

Table 12. CPS2 effect across all four conditions using a time-window analysis (-400..-150 msec vs. -150..100 msec relative to the onset of the disambiguating words – would/NP3)

		CPS2 Time-window analysis *TW1: (-400..-150 ms) vs TW2: (-150..100 ms)*									
		Midline Electrodes									
Source	*df*	*F*						*p*			
		Global ANOVA									
TW	*1,30*	*9.54*						*.004*			
TW × BSize	*3, 90*	*32.04*						*<.0001*			
TW × BSize × AntPost	*9,270*	*8.09*						*<.0001*			
AntPost1 (Fz)	*3, 90*	*10.83*						*<.0001*			
AntPost2 (Cz)	*3, 90*	*28.57*						*<.0001*			
AntPost3 (Pz)	*3, 90*	*36.9*						*<.0001*			
AntPost4 (Oz)	*3, 90*	*34.55*						*<.0001*			
		Pairwise Comparisons									
		1vs3		***1vs4***		***2vs3***		***2vs4***		***3vs4***	
		F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
TW	*1,30*	*21.88*	*<.0001*	*---*	*---*	*19.83*	*.0001*	*---*	*---*	*3.26*	*.08*
TW × BSize	*1,30*	*9.78*	*.004*	*51.27*	*<.0001*	*20.15*	*<.0001*	*67.85*	*<.0001*	*20.26*	*<.0001*
TW × BSize × AntPost	*3,90*	*14.34*	*<.0001*	*7.84*	*.003*	*7.17*	*.005*	*7.54*	*.003*	*11.42*	*.0002*
AntPost1 (Fz)	*1,30*	*---*	*---*	*12.65*	*.001*	*---*	*---*	*26.25*	*<.0001*	*15.56*	*.0004*
AntPost2 (Cz)	*1,30*	*3.74*	*.06*	*43.05*	*<.0001*	*12.66*	*.001*	*64.98*	*<.0001*	*26.62*	*<.0001*
AntPost3 (Pz)	*1,30*	*17.21*	*.0003*	*69.2*	*<.0001*	*20.64*	*<.0001*	*65.92*	*<.0001*	*23.04*	*<.0001*
AntPost4 (Oz)	*1,30*	*33.93*	*<.0001*	*56.4*	*<.0001*	*30.36*	*<.0001*	*59.49*	*<.0001*	*5.4*	*.03*

3.4.2.3. Garden-path effects

At the lexically disambiguating constituents (NP3 in LC and the modal verb *'would'* of VP2 in EC), incompatibilities between prosodically driven parsing decisions and the actual structure were predicted to elicit ERP garden path effects reflecting processing difficulties. Moreover, if these online processing difficulties underlie the acceptability judgments of our behavioral data, then we would expect that the strength of the ERP effects should follow a graded pattern as well. Specifically, the eBDH would predict that the garden path effects should primarily be a function of the size of the respective *incompatible* boundary (the early boundary in LC, and the late boundary in EC).

3.4.2.4. N400 in EC

Figure 9 shows the ERPs of the four EC conditions manipulating boundary 2, again time-locked to the onset of the disambiguating post-boundary word of VP2 (*'would'*). We can see a posterior N400-like negativity between 350 and 550 ms, peaking around 450 ms, largest at electrode Oz (see also voltage map of the difference between [B2_4] minus [B2_1] in Figure 9). This effect exhibits a graded pattern in the opposite ordering observed for CPS2, i.e., condition [B2_4] first shows the largest CPS amplitude and then the largest negativity, gradually followed by conditions [B2_3], [B2_2], and [B2_1]. As we assume these deflections reflect the processing difficulties created by the prosody-syntax mismatch, we will henceforth refer to the negativity in EC as an N400 garden-path component.

< Figure 9 here >

Although the N400 observed here appears more posterior compared to the parietally maximal canonical N400 reported in literature (for review see Kutas & Federmeier, 2011), we suspect a temporal overlap between the large positivity of CPS2/P200 components and the smaller negativity of the N400 may be the reason for this outcome. In EC (unlike in LC, see below) it appears that the CPS2 shifts abruptly and relatively early towards the negative range around 300 ms. That is, the early part of this component is likely cancelled out by the ongoing positive components. Moreover, since the amplitude of the overlapping CPS2 and P200 components is larger over the centro-parietal region, where the N400 usually peaks, and smaller over the occipital region, it is likely that the overlapping positive effects diminished the amplitude of the N400 more strongly at centro-parietal electrodes. As seen above in the context of the CPS analyses, another potential concern were moderate (and difficult-to control) slow shifts due to the distant baseline. Figure 10a shows a comparison between the two (prosodically matched) EC and LC conditions [B2_4], where the EC condition elicits the largest N400 and LC is the prosodic condition with the highest acceptability level for this structure. As can be seen, in this comparison the N400 already displays a broader distribution and is clearly present at both parietal and central electrodes. Importantly, this contrast also used the distant pre-PFL2 baseline. Since both conditions are prosodically (and lexically) identical up to the disambiguation point, their ERPs should – in principle – be indistinguishable before 0 ms. However, the EC condition is actually more positive between -500 and 0 ms, especially at PZ, most likely due to slow shifts, and thus ultimately due to the distant baseline. If this difference is removed by establishing a new baseline from -500 to 0 ms (which is entirely legitimate for this particular contrast), the two conditions are virtually indistinguishable until 250 ms (see Figure 10b), and the profile of the negativity in EC – although still maximal at Oz – now looks much more like that of a typical N400 component.

< Figures 10a and 10b here >

Since the late boundary was the incompatible one for EC structures, an N400 effect that increases with increasing boundary size seems to be in agreement with the predictions of the eBDH. For manipulations of the *early* (compatible) boundary, however, no impact on the N400 was expected. To test this hypothesis, we also computed averages manipulating the *early boundary*, while using the same (distant) baseline interval and time-locking position. The corresponding ERPs are shown in Figure 11. As expected, no differences among the conditions are visible either for the N400 or any other time windows. To quantify the N400 in EC conditions we used a time window from 350 to 500 ms (for effects of both the late and the early boundary).

< Figure 11 here >

3.4.2.5. Statistical analysis of the N400 in EC

Analyses in the 350-500 ms time window for the *late* boundary revealed a significant two-way interaction of BSize × AntPost at both lateral [$F(9,270) = 2.72$, $p<.05$] and midline electrodes [$F(9,270) = 2.87$, $p<.04$], confirming that the N400 differences between conditions were largest over the posterior region, Follow-up analysis at the midline showed the effect was significant only at Oz [$F(3,90) = 4.75$, $p<.01$], whereas no significant effects were found at the lateral array. Pairwise comparisons between the four conditions at Oz showed significant differences between conditions [B2_1] vs. [B2_3] [$F(1,30) = 9.81$, $p<.01$], [B2_1] vs. [B2_4] [$F(1,30) = 10.11$, $p<.01$] and [B2_2] vs. [B2_4] [$F(1,30) = 6.39$, $p<.02$. In addition, similarly to the ordering pattern observed in Figure 9, the mean amplitude for each condition at Oz was found to be fully graded: [B2_1] (–.032μV) < [B2_2] (–0.76 μV) < [B2_3] (–1.23 μV) < [B2_4] (–1.63 μV). Analyses for prosodic manipulations at the *early* boundary did not reveal any significant effect or interaction involving factor boundary size (all F's < 1).

3.4.2.6. P600 in LC

Figure 12 displays the ERPs of the four LC conditions manipulating the *late* boundary, time-locked to the onset of the disambiguating NP3, again using the distant baseline. Between 600 and 1400 ms, condition [B2_1] elicits a positive-going waveform compared to all other conditions, which seems largest over centro-parietal electrode sites. That is, the condition that shows the smallest CPS and the smallest P200 appears to show a late positivity in the P600 time range. Since the late boundary is the compatible (cooperative) boundary for LC structures, the BDH would not have predicted a strong impact of the second boundary. However, [B2_1] should clearly be the most difficult one of the four conditions, and a similar pattern (just for boundary size 1) was also found in the behavioral data. Numerically, the 'easiest' (i.e., most compatible) condition [B2_4] shows a local negativity at OZ in the same time range. Note that, since the voltage maps are based on comparisons between the 'hardest' and the 'easiest' garden path conditions, this relative negativity shifts the maximum of the P600 difference wave toward occipital electrodes. Similar to the N400 analyses, we also computed contrasts to reveal the impact of the *early* boundary on the ERPs of the LC conditions. These are depicted in Figure 13. Since these conditions only differ at the first boundary position, but are matched for the second boundary (by averaging across all boundary size levels), we could use an (unusual) post-target onset baseline from 0 to 100 ms, thus not displaying the ERPs before 0 ms[20]. The LC plots for the early boundary manipulation show a P600-like positivity that resembles the one in Figure 12 for the late boundary. However, this positivity has an earlier onset (around 400 ms). Moreover, now the two conditions with the *largest* boundary size (B1_3 and B1_4) elicit such a component,

---

[20] Note that employing such a baseline had no effect on the absence of any N400 effects for the corresponding EC contrasts shown above in Figure 11.

whereas the other two conditions do not, at least not initially. As the early boundary was prosodically incompatible with an LC analysis, garden path effects for larger boundary sizes were predicted by the eBDH; however the effects do not display a graded pattern. Condition B1_1 seems to display a later and smaller temporary positivity. To capture this potential effect, and to be consistent across comparisons, the P600 in LC was generally analyzed in a time window between 600 and 1400 ms.

< Figure 12 here >

3.4.2.7. Statistical analyses for the P600 in LC

Analyses in the 600-1400 ms time window for the *late* boundary revealed a significant main effect of factor BSize at both lateral [$F(3,90) = 4.55$, $p<.01$] and midline [$F(3,90) = 5.1$, $p<.01$] electrodes, as well as an interaction of BSize × Hemi at lateral electrodes [$F(3,90) = 3.19$, $p<.05$], reflecting the differences among the four conditions in the global ANOVAs. Follow-up analyses showed the P600 was broadly distributed, most prominent over the centro-parietal region. In line with the visual inspection of Figure 12, the pairwise comparisons confirmed that the ERPs in condition [B2_1] were significantly more positive than conditions [B2_2] (Lateral: [$F(3,90) = 6.93$, $p<.02$]; Midline: [$F(3,90) = 6.59$, $p<.03$]), [B2_3] (Lateral: [$F(3,90) = 9.85$, $p<.01$]; Midline: [$F(3,90) = 8.75$, $p<.01$]), and [B2_4] (Lateral: [$F(3,90) = 6.83$, $p<.02$]; Midline: [$F(3,90) = 10.59$, $p<.01$]), which were found to be statistically indistinguishable from one another. Follow-up analyses also revealed the BSize × Hemi interaction pointed to the effect being more prominent at the left hemisphere between boundary sizes 1 and 3 only (Global: [$F(3,90) = 4.79$, $p<.05$]; Hemi1: [$F(3,90) = 12.97$, $p<.01$]; Hemi2: [$F(3,90) = 5.83$, $p<.03$]).

The analyses for the *early* boundary in LC sentences revealed only a significant main effect of boundary size [F(3,90) = 4.76; p < .005] without any further interactions, pointing to a broadly distributed P600 profile. Pairwise follow-up analyses confirmed that boundary sizes 3 and 4 (showing a P600) each differed from boundary sizes 1 and 2 (not showing a P600) [all *p*-values < 0.5]. In contrast, boundary size 3 did not differ from boundary size 4 [F < 1], nor did size 1 differ from size 2 [F <1].

< Figure 13 >

3.5. Discussion

This ERP study examined the effect of two competing prosodic boundaries of varying sizes on the processing of spoken English garden-path sentences (Early and Late Closure). Behaviorally, we reliably replicated the findings of Pauker et al. (in preparation), demonstrating the effect of boundaries in this task was consistent across studies. The ERP findings were also in line with the previous literature, showing CPS components at prosodic boundaries, preceded by small frontal negativities (i.e., pre-CPS negativities), across both structures. In EC, the size of the incompatible (late) boundary modulated a garden-path N400 on the disambiguating modal verb *would* in a graded manner, whereas size manipulations of the compatible (early) boundary did not affect the N400. In LC, a P600 on the disambiguating NP3 (e.g., *the dogs*) was found in the most difficult conditions, again as a function of the incompatible (early) boundary and, surprisingly, somewhat influenced by the compatible boundary as well. These results are in line with the predictions of the eBDH. The pattern as a whole can be better accounted for by a gradient than by a categorical view of boundary processing, but in absence of unambiguous evidence for consistently graded ERP garden path effects, the categorical view can also account for a number of data points. We will first discuss the behavioral results and then focus on the ERP effects, starting with CPS effects at the boundary positions and then turning to garden path effects.

3.5.1. Behavioral results

Here we replicated the results of Pauker et al. (In preparation), thus adding to the validity of the paradigm and supporting the extended BDH. First, we found that even small early boundaries biased listeners towards EC; much larger second boundaries were required to override this bias. Unlike Pauker et al. (In preparation), who found LC preference in three conditions, in the present study only condition [1_4] showed a clear LC preference. However, the other two conditions [2_4] and [1_3] were rated virtually equally in both structures and did not display the strong EC advantage observed in the other 13 conditions (see Figure 1). Second, only the strength of the *incompatible* boundary (Early Boundary in LC; and Late Boundary in EC) drove acceptability ratings in each structure. Third, these ratings were significantly different between the boundary levels, in a gradient manner – with the largest (incompatible) boundary receiving the lowest scores; compatible boundaries did not seem to have any impact (no gradient). The only exception was the finding that while three Late Boundary levels (i.e., 2, 3, 4) in LC were statistically indistinguishable from one another, the smallest boundary size (1; “no boundary”) received significantly lower scores by comparison. It appears that listeners perceived this boundary size as considerably weaker at the late boundary position, which further increased the expectation of an EC structure, resulting in a stronger garden-path effect. Since all conditions carried at least some indication of an early prosodic boundary, which was perceptually sufficient to trigger a strong EC expectation across the board, we observe this deviance (reflected by a 3-

way interaction) from the overall consistent pattern only in LC. This outcome supports the claim of the extended BDH regarding an early boundary advantage, at least for the structures and boundary sizes used here. Interestingly, this asymmetry between structures is also reflected in the ERP data, when comparing the garden-path effects between EC and LC.

Given the overall preference for EC as well as the observation that only much larger late boundaries could overturn this bias, it is conceivable that the weak late boundary at size level 1 was immediately identified as 'irrelevant' for any change of the initial EC preference towards an LC preference. However, 'irrelevant' in this context should not be confused with 'non-informative'. To the contrary, under the assumption that this weak late boundary may have confirmed the initial EC preference beyond any reasonable doubt, it was likely the reason why only in this condition a clear LC garden path effect was observed – including in ERPs. An important question regarding our understanding of the integration of multiple boundaries is how this very boundary would be processed in complete absence of any early prosodic boundary. In our opinion, it may well be that under these circumstances a small boundary could make the difference and potentially override lexical and other non-prosodic biases favoring an EC analysis. At least the data by Itzhak et al. (2010) suggest that prosodic information should be able to overturn such biases, but they were based on strong IPh boundaries. If confirmed, such a finding would be strong evidence for the 'global' processing view of the 'relative boundary strength hypothesis', the Rational Speaker Hypothesis, and the Informative Boundary Hypothesis put forward by Frazier et al. (2006), according to which the strength of previous (competing) boundaries should be decisive in determining the relevance and impact of a later boundary.

A last point concerns the relationship between the present study and our previous studies. The fact that the judgment data of this ERP study almost perfectly replicated those of our behavioral studies is important for two reasons. First, it suggests that the slight modifications of the design (e.g., the asymmetric distribution of trials across conditions) and of the experimental setting (recording EEG data) did not have any effect on the way in which participants processed the sentences. Secondly, and related to the first point, the replication of results during the EEG experiment also means that the ERP patterns found in the present study are most likely representative of the real-time processing in those previous studies as well.

3.5.2. ERP results

In the following section we discuss the ERP results, first at the early and then at the late boundary position in the sentences. As both syntactic structures were acoustically identical at certain time-locking points and were thus analyzed together (by collapsing across matching conditions), we distinguish between the shared and non-shared ERP effects. Specifically, both CPS analyses are considered a shared effect, whereas the garden-path effects are structure-specific. As the shared effects take place first, we review and interpret the ERP data in that order.

3.5.2.1. Shared effects at boundary 1

Although the CPS component has been reliably demonstrated using similar material in previous studies (Itzhak et al., 2010; Pauker et al., 2011; Steinhauer et al., 1999), it was unclear whether the prosodic manipulation employed in this study would be sufficient to trigger this component. The first reason for this uncertainty was that the boundaries (apart from the original ip) were created digitally and were derived from a weak boundary, rather than reduced from a large boundary (to ensure the sentences sound as natural as possible). To our knowledge, no other study has shown such material can elicit a CPS. Second, the CPS components reported in previous studies in English were elicited by IPh boundaries ranging from 400 to 500 ms on average, while the duration of the largest boundary used in the present study was only 320 ms (~70 ms of the original ip + 250 ms duration manipulation). In fact, our largest boundary was smaller than some ToBI-defined ip boundaries used in previous behavioral studies, both in terms of break and pre-final lengthening duration (see Carlson et al., 2001). Moreover, the mid-utterance IPh boundaries in previous studies are usually characterized with a salient falling and rising (L-H%) tone, while in the present study we used an ambiguous pitch contour, usually characteristic to ip boundaries (H*L-) to create the IPh boundary (H*L-L%). This specific contour was also used for the same purpose by Kjelgaard and Speer (1999). Taken together, these pitch and duration differences made our boundaries considerably weaker compared to those used in those used in the published literature. Third, to date, only one other ERP study we are aware of reported a CPS at a boundary smaller than an IPh (Li & Yang, 2009), demonstrating a CPS for ip boundaries in Chinese, which *did not differ statistically* from a CPS elicited by an IPh boundary. Finally, as mentioned earlier, since all boundaries shared the same pitch contour and differed only in terms of relatively subtle duration changes, resulting in boundaries that the categorical view treats as variants of the ip category (Carlson et al., 2001), no differences whatsoever should have been found between conditions.

For these reasons our results are quite striking: we showed not only that the CPS component reflects the detection of boundaries of a size below the IPh level, but also that it is sensitive enough to mirror the subtle differences between them. Specifically, we found that increasing the boundary size across all levels elicited *CPS components (and pre-CPS negativities) whose amplitudes also increased in a (largely) gradient manner*. Results show that all boundary levels, with the exception of levels 1 and 2, were significantly different from one another.

Although these results appear to strongly support accounts assuming gradient (rather than categorical) boundary processing, we should consider a very important question, namely, which level does the CPS tap? Does it underlie an acoustic/perceptual or a phonological level of processing? The answer to this question is crucial for our interpretation of the results. Both the gradient and categorical views agree that boundaries can be produced in a variety of ways. The gradient approach does not distinguish between an acoustic and a phonological level. That is, it is implicitly assumed there is no separate phonological level. The categorical approach, on the other hand, asserts that while variants of the same category may be *perceived as acoustically different*, listeners ultimately associate them with *the same phonological category*, which bears distinctive phonological characteristics (Carlson et al., 2001). Therefore, if the CPS reflects processing at the acoustic/perceptual level, then the pattern of results may be explained simply in terms of the acoustic differences between the boundaries. Such an account would be acceptable by both approaches. In that case, it would be necessary to determine whether the garden-path effects show gradient or categorical differences between conditions, as this would indicate the parser integrated the boundaries differently. If, however, the CPS taps a higher-order phonological level, then the current results advocate a gradient account, as either a clear division (in case of two different categories) or no differences at all (if all boundaries are variants of the ip category) should have emerged between conditions – neither of which is the case here.

Some evidence suggests it may indeed reflect the phonological rather than perceptual level. Although the CPS has been shown to be elicited by delexicalized/low-pass filtered sentences (Steinhauer & Friederici, 2001) and hummed speech (Pannekamp et al., 2005), which may suggest it is a lower-order perceptual marker of boundaries, it could also be argued that it actually reflects a higher-order suprasegmental level of processing. Recall that ToBI defines prosodic boundaries in terms of pitch variations and breaks, and that the theory of prosodic phonology (which underlies the notion of prosodic units) claims there is “a many-to-one and one-to-many mapping between prosodic structure and these acoustic dimensions” (Shattuck-Hufnagel & Turk, 1996; p. 238-239). We suggest that the phonological information crucial for the marking of boundaries ‘survived’ the delexicalization process while segmental information was eliminated. In addition, past research has demonstrated that the CPS can be evoked in the absence of overt acoustic markings, by processes triggering “subvocal phonological sentence phrasing” (Steinhauer & Friederici, 2001; p. 286), including long subject NPs (Hwang & Steinhauer, 2011) and commas (Steinhauer & Friederici, 2001) in silent reading. However, perhaps the strongest evidence for the CPS representing a higher-level process is a study by Itzhak et al. (2010) in which this component was elicited in the total absence of a prosodic boundary, merely due to a strong *expectation* of a boundary, created by lexical and syntactic factors (e.g., transitivity bias and structural preference).

The magnitude of the CPS has also been shown to be modulated by various factors, including acoustic and linguistic information. For example, the smaller CPS found in the comma study compared to the larger CPS found in the auditory study, was explained in terms of a lesser activation of phonological representations. That is, it seems that the degree to which a CPS is activated is dependent on the number and saliency of multiple higher and lower order cues, including acoustic, phonological and syntactic information. Future studies would be needed to explore this issue further. Finally, the behavioral results, which show a graded scoring pattern that is directly driven by the influence of each boundary size on comprehension, are consistent with the magnitude of the CPS. These outcomes would not have been expected by a categorical account.

As seen in previous studies (Bögels et al.,2010; Pauker et al., 2011), the CPS was preceded by a central negativity (so-called Pre-CPS negativity), which was somewhat right-lateralized, especially over frontal and central electrodes. As hypothesized in Pauker et al. (2011), this negativity was triggered relative to the pre-final lengthening on NP2 (PFL1), a time-locking position which had not been available in previous studies. Moreover, it displayed a similar gradient pattern as the CPS, where its magnitude increased with that of PFL1. Boundary level 1, containing no pre-final lengthening, was therefore an exception, showing no negativity but rather an early onset of a small CPS. Depending on the exact quantification of the CPS, the presence of this negativity can add to the overall amplitude of the positive shift. In fact, as we have shown, even after the introduction of multiple trigger points to identify the ERPs of smaller segments in the speech signal, teasing apart the negativity from the CPS proper is a non-trivial task. Future work will have to determine whether this negativity is a mandatory sub-component of a biphasic CPS-complex.

3.5.2.2. Shared effects at boundary 2

The CPS elicited at the late boundary ("CPS2") was very similar in essence to the CPS found at the early boundary ("CPS1"), in terms of scalp distribution, latency and amplitude. More importantly, the gradient pattern observed with CPS1 was replicated; all boundary levels, with the exception of levels 1 and 2, were significantly different from one another in a graded manner (although all levels exhibited a numerically gradient difference). We also replicated the pre-CPS negativity, which also showed the same gradience, albeit in an inverse order (boundary level 4 induced the largest negativity). However, a few differences between these components at the late vs. the early positions should be noted. First, CPS2 seems to shift into the positive range more sharply and later in time compared to CPS1; second, the pre-CPS negativity at the late boundary is larger, spans over a longer period of time, and is more broadly distributed compared to the one observed at the early boundary. These differences are especially noticeable when comparing the components side by side (see Figure 13 below). Given these differences, we hypothesize that the pre-CPS2 negativity may have been superimposed by a second negativity. Such a combined negativity may have cancelled out some of CPS2's early positivity, and may also have resulted in an apparent delay of the CPS component (compared to CPS1). We briefly discuss two possibilities of what this additional negativity preceding CPS2 may have been: an N400 and an expectancy negativity (similar to that in Pauker et al., 2011).

< Figure 13 here >

To explain the N400 hypothesis, it is important to remind the reader of the manner by which our data was analyzed. To obtain enough trials in each of the four late boundary conditions, we kept the early boundary size constant by collapsing across all conditions containing the same late boundary size (i.e., [B2_X] = [1_X] + [2_X] + [3_X] + [4_X]), thereby creating ERPs of a medium-sized early boundary on average (with a mean size of 2.5). Based on our behavioral data, we know that even a relatively small early boundary size (smaller than 2.5) was perceived as a strong indicator of an EC structure. When PFL2 is encountered, it provides sufficient information of an upcoming second boundary. Importantly, if this second boundary is large, NP2 (which should have served as the subject of the subsequent clause) becomes stranded between the two boundaries and would not receive a theta role, a processing problem known to elicit N400 effects (e.g., Friederici and Frisch, 2000). The stronger the indication of a late boundary is, the stronger the violation and the N400 should be, and this is compatible with our findings. This particular violation greatly resembles the one found in the prosody-syntax mismatch condition D in Pauker et al. (2011), which contained two IPh boundaries flanking NP2 (e.g., *When the bear was approaching # **the people** # the dogs come running*). This violation was interpreted as an interruption of an early process of preliminary theta role assignment that is largely guided by the syntactic structure as well as by the prominence of the NP.[21] The fact that we do not observe a biphasic N400-P600 pattern, as did Friederici & Frisch (2000) or Pauker et al. (2011), suggests a weaker type of violation, compatible with its transient nature in our present study. In the other two studies, the lack of a theta role signaled an outright violation. As Bögels et al. (2013) correctly pointed out, this was actually a weakness of the D condition in Pauker et al., because the second boundary was a reliable predictor of the ultimate (lexical) disambiguation towards an LC structure (such that the P600 was elicited 'too early'). In our present design both EC and LC sentences were presented in all prosodic conditions, such that garden path effects were postponed until later (see below).

A second possibility is that the larger pre-CPS negativity at the second boundary was influenced by a co-occurrent expectancy negativity, similar to that observed in Pauker et al. (2011) in conditions A and C. In that paper, we argued that the lack of a first boundary may have triggered a strong expectancy regarding the presence vs. absence of a boundary at the second position. At first, this seems inconsistent with our present finding of a larger negativity after *larger* boundary sizes at the first position. However, given the much more complex design of our

[21] For example, the assumed grammatical function of subject for the first appearing (animate) NP in a new clause strongly predicts (at least in English) it is also the "active participant, from a thematic perspective; i.e., the Actor (Van Valin & LaPolla, 1997) or Proto-Agent (Dowty, 1991; Primus, 1999)" (Schlesewsky & Bornkessel, 2004p. 1214; see also Bornkessel & Schlesewsky, 2006 for a theoretical framework). Due to the two boundaries, the 'stranded' NP in condition D would neither receive a theta role from the preceding verb nor a proto-role based on its grammatical function and prominence. This interpretation is consistent with previous findings showing an N400 is elicited by thematic information processing difficulties (Friederici & Frisch, 2000; Frisch, Hahne, & Friederici, 2004; Frisch & Schlesewsky, 2001; for review see Friederici & Weissenborn, 2007).

current experiment, it could be argued that here a strong initial boundary might have resulted in different types of expectations. Moreover, the rather frontal distribution of the present effect also seems more in line with the profile of an expectancy-related negativity than with an N400. However, the observation that the current negativity was triggered at the first acoustic marker of the boundary (PFL1), not at – or right after – the first boundary position as in Pauker et al. (2011), suggests that an expectancy-based account is rather unlikely.

3.5.2.3. Non-shared garden-path effects at lexical disambiguation points

P600 in LC

Somewhat similar to the LC garden-path condition D in Pauker et al. (2011), LC conditions with incompatible prosodic structures were found to elicit a parietal positivity at the lexically disambiguating third NP (NP3; "*the dogs*"). This effect was most evident in the two conditions that contained the largest *incompatible* (early) boundaries (as predicted by the BDH), but was also found for the condition with the weakest *compatible* (late) boundary, mirroring our behavioral findings. As in Pauker et al. (2011), we interpret this deflection as a P600 effect, reflecting the processing difficulties associated with garden-path sentences (Osterhout & Holcomb, 1992, 1993; Osterhout et al., 1994), which is potentially superimposed by a domain-general response-related P300 effect, previously shown to be evoked by working memory updating and reorganization (Donchin, 1981; Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001). Despite their lexical similarity, however, in condition D, the P300/P600 complex co-occurred with the CPS evoked at the second IPh boundary, resulting in a very large, yet more focal, positivity (~5μV) with a relatively early latency. Here, on the other hand, the positivity displayed a much smaller amplitude (~2μV) as well as longer duration and later latency. In the next paragraphs we discuss several factors that may have contributed to this difference: (a) prosodic structure, (b) the type of prosody-syntax mismatch, and (c) the procedure (task and response type).

First, recall that the ERP garden-path effects in condition D of Pauker et al. (2011) were elicited before the disambiguating word was even encountered. As discussed by Bögels et al. (2013), since this was the only condition containing two IPh boundaries, participants could rely on prosodic, rather than lexical, information to detect the violation. In order to avoid similar strategic learning in the present study and to tease apart the CPS and P600, we created a fully symmetrical design, where EC and LC were identical until the offset of the late boundary. Thus, listeners had to rely on lexical information for disambiguation, rather on the earlier prosodic cues. Consequently, the onset of the garden-path effect was delayed compared to the one observed in condition D, and did not co-occur with the CPS. One might argue that the latency difference between the studies might also have contributed to amplitude differences, since the positive components no longer co-occurred in the present study. However, the larger P600 in Pauker et al. (2011) was quantified in a contrast between condition D and its matched control condition A, which also contained a boundary (and a CPS), thereby separating the P600 from the CPS amplitude.

Second, as the P600 was elicited on NP3, the garden path effect is qualitatively different from that in Pauker et al.'s (2011) condition D. In their condition D, the biphasic N400-P600 pattern was likely evoked because NP2 was initially stranded without any theta role due the presence of the superfluous early IPh boundary separating NP2 from the preceding subordinate clause, and a late IPh boundary preventing its attachment to the subsequent clause ("*when a bear is approaching* ***# the people #*** *…*"). In absence of any filler sentences in that study that could have licensed a second boundary, the stranded NP was sufficient to signal the ungrammaticality of the sentence (see also point three below). By contrast, in the present study, all conditions eliciting the P600 were characterized by the presence of a weak late boundary following a larger early boundary, which created an expectation for an EC parse. When NP2 was encountered, it was thus assigned the grammatical function of subject NP, creating an expectation for a verb to follow. The rather weak second boundary did not seem to have changed this EC preference. Since the incoming input was an NP instead of the expected verb, it functioned as a trigger for a necessary structural (and prosodic) revision, very similar to many previous garden paths in the ERP literature (e.g., Osterhout and Holcomb, 1992, 1993). In fact, the P600 in the present study is quite comparable to effects found in those studies.

Third, the response paradigm and administered task used in the current study likely contributed to a longer latency and more variable garden-path effect (especially the smaller amplitude) compared to the one observed in Pauker et al. (2011). That study employed a *binary response paradigm* (forced choice acceptability judgment), which allowed participants to make very quick yes/no judgments, resulting in shorter RTs. Conversely, in the current study, the 5-point *scale response paradigm* was more complex as it demanded additional evaluation time, leading to slower RTs. A big advantage of this task was that participants could not simply rely on superficial cues to make a decision, and as reflected by the graded acceptability rates, they clearly evaluated the 32 different conditions with impressive attention to even small nuances in the prosodic pattern. Similar to RTs, this response paradigm is very likely to partly account for the longer latency of the P300/P600 (as the positive shift begins around 500 ms), which has been shown to accurately reflect “the time required to evaluate and categorize an event” (Johnson, 1986; p.379). The use of a scale for our judgment task certainly also explains the difference in P600 amplitudes across studies. In Pauker et al (2011), the second boundary was sufficient to determine the ‘no’ answer in the binary acceptability judgment, which is a standard paradigm to elicit large P300 components that reflect (i) stimulus categorization, (ii) task relevance, and (iii) response preparation (Donchin, 1981; Johnson, 1986). In other words, in the 2011 study, the P300/P600 effect was likely strongly dominated by the P300 component, and it is not even clear if structural reanalyses were consistently initiated. In the present study, by contrast, participants needed to carefully evaluate each stimulus to give the most appropriate response out of 5 choices. This scenario is much more likely to actually involve attempts to understand and – if necessary – reanalyze the sentence. In other words, the positivity in our data is more likely to actually reflect psycholinguistically relevant processes than the P600 in condition D of the 2011 study.

Interestingly, when comparing the ERP outcomes for the *late* (compatible) boundary manipulation to the corresponding behavioral data, it appears that condition [B2_1] was again statistically different than the other 3 conditions: whereas all of these elicited statistically indistinguishable ERPs, only [B2_1] showed a clear P600. This outcome mirrors the behavioral finding of an unexpected statistical difference between Late Boundary size 1 (“no boundary”) and the other boundaries, which was reflected by a small 3-way interaction. Importantly, if boundary sizes 1 and 2 were in fact variants of the same category (as may be indicated by the lack of difference between these components in the CPS analyses), no differences should have been observed between them here. Rather, both conditions [B2_1] and [B2_2] should have displayed similar effects compared to condition [B2_4]. Condition [B2_3] should have clustered with either the smaller or larger boundary types, depending on whether it should belong to the ip or IPh category. Alternatively, if all boundaries belonged to the ip category, no differences whatsoever should have been found between them. Neither of these potential outcomes was the case here. Although no clear effect of the compatible boundary was expected by the BDH, the condition showing the P600 is certainly the most difficult LC condition (with the smallest compatible boundary). The fact that the ERPs provided an online correlate of the effect observed in the behavioral data, suggests that the latter were indeed a consequence of the neurocognitive processes taking place right at the disambiguating NP3.

Perhaps the most interesting finding was the influence of the *early* boundary on the P600. Similar to our behavioral data, the magnitude of the P600 garden path was largest for those 2 conditions that contained the largest (incompatible) early boundaries, strongly suggesting that the processing difficulties at the disambiguation point were largely related to the strength of this boundary. This finding is exactly what the eBDH predicted. However, the amplitudes did not follow a strictly gradient pattern; instead they only distinguished between boundary sizes 1 and 2 (no garden path) versus 3 and 4 (garden path). This pattern is in line with the predictions of the categorical processing account. On the other hand, it is somewhat surprising that the behavioral data, which seem to be surprisingly well reflected by the ERP pattern in general, did show the gradient pattern. Given that all of these analyses used a distant baseline, it is possible that the lack of a gradient pattern was influenced by our baseline choice. Additional analyses employing other baseline options may reveal if this was the case. In sum, the P600 data confirm assumptions of the BDH but they seem to favor a categorical over a gradient view of boundary processing.

N400 in EC

When investigating the effects of the incompatible *late* boundary on EC sentences, condition [B2_4] with the largest boundary size elicited a clear occipitally-maximal negativity on the lexically disambiguating modal verb ("*Would*"), which we interpret as an N400 component.

The other three conditions showed similar N400s whose amplitudes decreased with decreasing boundary size.

We assume two reasons this effect shows a posterior, rather than the classic parietal, scalp distribution (for review see Kutas & Federmeier, 2011). First, the large positivity of the preceding, similarly-distributed, CPS2 partially overlapped with the smaller negativity of the N400 and most likely cancelled out (or considerably diminished) this effect at the parietal region. As shown in section 0, although CPS2 should not differ between structures, as it is elicited by the same lexico-prosodic information, in EC it returns to baseline about 200 ms earlier than in LC, and is then immediately followed by the N400. Second, due to the absence of a control condition for this structure (EC sentence containing a single early boundary), comparing the negative deflection between the EC conditions may not fully reflect the magnitude of the effect, because all of them elicit some negativity. A better comparison for this purpose was condition [B2_4] in LC, because it was identical before the disambiguating point while not eliciting a garden-path effect in this time window. Indeed, this comparison showed the N400 was in fact more broadly distributed. This comparison also revealed that the distant baseline had partly shifted the data.

Even more so than the garden-path effect in LC, the N400 amplitudes of the four conditions were numerically graded but only partially statistically different. Importantly, the significant differences were found between those conditions whose overall prosodic structure either provides a stronger ([B2_1]; [B2_2]) or weaker ([B2_3]; [B2_4]) indication of an EC phrasing. To understand the difference between structures, we should consider that unlike condition [B2_1] in LC, which resembles the classic prosodically-induced garden-path violations used in previous behavioral studies (Kjelgaard & Speer, 1999; Schafer et al., 2000; Speer et al., 1996; Walker et al., 2001; Warren et al., 1995), with an incongruent boundary at the EC attachment site, condition [B2_4] in EC shows greater resemblance to condition D in Pauker et al. (2011), as both contain two strong boundaries preventing the attachment of NP2 to the verb in the subordinate clause as well as its assignment as the subject of the subsequent clause. However, compared to these two LC garden-path conditions, the violation in EC is considerably weaker, both behaviorally and electrophysiologically. That is, the garden-path in LC resulted from a misleading expectation for an EC structure, whereas the violation in EC was induced by interference of the late boundary to the EC construction. To understand the nature of the violation in EC, let us consider the influence of the two boundaries on the sentence's interpretation. We have already established that the strong early boundary (size 2.5 on average, using the early boundary collapsing technique) creates a relatively strong expectation for an EC structure. Therefore, when NP2 unfolds, it is initially assigned the role of subject NP (and, based on the word order in English, the proto-role of "doer") of the new clause. However, the late boundary indicates NP2 will not receive a proper theta-role, which is reflected by an N400 effect. The following disambiguating word (*would*) is a modal verb, which cannot assign a theta role or determine the number of required arguments (in the same fashion a verb does), but it does indicate that a subject NP is required, with a proto-role of a "doer". This creates a "convenient" situation where an NP lacking an external theta role (assigned to the subject NP; Carnie, 2007) is followed by a modal verb with a role of proto-Agent that needs to be allocated to a missing NP. The N400 in this case is thus elicited not by the absence of a theta role, which the modal verb cannot assign, but due to the lack of an argument. Given the nature of this violation, the absence of a clear indication of a P600, which is expected in such garden-path violations (Bögels et al., 2010; Pauker et al., 2011; Steinhauer et al., 1999), was surprising. A small positivity at the same time-window as the P600 in LC was found insignificant even between the extreme conditions.

We assume two main reasons for the impedance of this effect: (i) the weak garden-path in EC and (ii) a potential overlap between the N400 and the P600 components.

At the point of disambiguation, the saliency of NP2 to the parser was higher because of the recent preceding violation; this, combined with the proximity of the interfering boundary and the already strong EC preference (which was difficult to override, even in the presence of a large second boundary), facilitated the mental deletion of the late boundary in order to repair the sentence. By comparison, mentally deleting the earlier, fully integrated boundary, as in the case of the garden-path in LC, proved more demanding to the parser. Even behaviorally, participants recovered from the strongest violation in EC (condition [1_4]) more easily than from the strongest violation in LC (condition [4_1]), as reflected by higher acceptability proportions in EC (52%) compared to LC (32%) for these conditions. In fact, even the difference between the ratings of the least and most compatible conditions was 50% larger in LC (33%) relative to EC (22%). Also, if we compare the garden-path effect here to the one found in condition C in Pauker et al. (2011), in which EC sentences were biased towards LC due to the absence of an early prosodic boundary (e.g., "*When the bear was approaching the people* ***come running***"), we learn that even though a P600 component was in fact elicited, it was very small in terms of amplitude and scalp distribution. Interestingly, both condition C and condition [1_4] in EC received the same acceptability scores and were both perceived as weak violations. This in in line with the extended BDH, which predicts an advantage of the earlier compared to a later appearing boundary in driving the syntactic preference in these structures, requiring a much larger late boundary in order to override this bias.

Another possibility is the existence of a temporal overlap between the N400 and the P600, leading to the cancellation (or reduction) of these components in the averaged ERP signal.

This is common in auditory studies, since "spoken syntactic violations are stretched out over time and vary in their duration. As a result it becomes more difficult to obtain clearly non-overlapping phases in the averaged waveforms" (Hagoort & Brown, 2000; p.1548). Since the violation in EC was weak and evoked a smaller garden-path effect as a result, such overlap may completely negate an existing effect, rather than simply reduce it. In order to determine whether the small positivity in EC is in fact a small P600, we created two new conditions comprised of three sub-conditions each, which corresponded to the goodness of fit between prosodic and lexical information. These conditions were time-locked to both the onset of *Would* (cue point j) as well as the second VP (*VP2*; cue point k) – a position used to quantify the P600 in condition C. Accordingly, conditions gpECj and gpECk represented the garden-path effect at Would and VP2, respectively, and were created by averaging conditions [1_4], [1_3] and [2_4] – all of which showed stronger LC preference in the behavioral results. Conditions nogpECj and nogpECk time locked to Would and VP2, respectively, were created averaging together the conditions showing the best fit to the EC structure: [4_1], [3_1] and [4_2]. The comparison at both positions (see Figures 14 and 15) showed a P600 component was evoked around 500 ms relative to Would in condition gpECj and 300 ms relative to VP2 in condition gpECk. Recall that the N400 was elicited between 350 and 500 ms relative to Would, which confirms our hypothesis that the two components partially overlap. Although not verified statistically, visual inspection and the embedded running t-test indicate that this effect is more reliable relative to VP2, likely due to the small size of the effect. This is in line with our assumption that the violation corresponds to the absence of an argument. Although this comparison is not ideal to demonstrate the individual effect of the late boundary sizes on processing, it does provide an indication such methodological consideration should be taken into account.

< Figures 14 and 15 here >

Finally, to examine whether the ERP data corresponds to the behavioral data, where the early boundary in EC showed no effect on processing, we created four conditions time locked to *Would*, which varied in the size of the *early* (compatible) boundary while the size of the late boundaries was kept constant. As expected, we did not observe any effect, i.e., all conditions showed the exact same pattern and were statistically indistinguishable. This finding mirrors the results of the behavioral study, strongly confirming the prediction of the BDH that the processing of this structure was exclusively affected by the size of the incompatible late boundary. As a whole, the garden path effects in the EC condition were in perfect agreement with the BDH and also reflected the graded pattern of amplitudes predicted by the gradient view of boundary processing. These data, along with the corresponding behavioral data, are difficult to account for within a framework assuming categorical boundary processing, because the N400 garden path effect can only be linked to the *phonological*, not the acoustic/perceptual processing.

3.6. Conclusion

In the present study we replicated the behavioral outcomes of Pauker et al. (in preparation) and confirmed that (i) even small early boundaries biased listeners towards EC, with much larger second boundaries required to override this bias, and that (ii) the strength of the incompatible boundary (early in LC, late in EC) drove the acceptability ratings. These findings can only be explained by the Boundary Deletion Hypothesis (BDH). Importantly, we also extended the findings of our previous study, showing not only a gradient pattern of acceptability in the behavioral, but also in the ERP data. Specifically, both CPS components showed larger amplitude in each increasing boundary size, with nearly all boundary levels statistically different from one another. While the garden-path effects showed a numerically graded pattern, only some conditions were found to be statistically different. However, several factors have contributed to these less clear-cut outcomes. First, using the averaging technique to compensate for the small number of trials in each condition resulted in an early boundary that was neither weak nor strong, which accounts for the relatively small size of garden-path the effects in both structures. Also, the size of the largest boundary used here was considerably smaller than that of boundaries that were previously reported to elicit such effect in similar material. Using a smaller number of conditions with more trials in each would undoubtedly increase the power of the observed results, and potentially also the differences between conditions. Given the relatively subtle violations and boundary markers used in his study, our findings are rather encouraging. Despite these shortcomings, the finding that the ERP data mirrored the behavioral results, and the fact that differences between smaller boundaries (such as 1 and 2 in LC) were observed, would not be expected if boundaries were perceived categorically. Even the fact we found numerical differences between the averages of the effects is indicative of a continuous grid, and suggest that the purely categorical approach to boundary size perception at the phrase levels should be reconsidered, at least in the context of EC/LC garden-path sentences, where the phrasing is a strong indicator of the intended parse. It should be noted, however, that the actual markers of boundaries, especially the tonal events, as described by ToBI, proved instrumental for the detection of boundaries. The findings demonstrate that subtle differences between prosodic boundaries are detected by the brain and affect the degree of processing difficulty. These outcomes cannot be explained by a purely categorical account. Moreover, they cast serious doubts on most current models of prosodic online processing and strongly support the BDH.

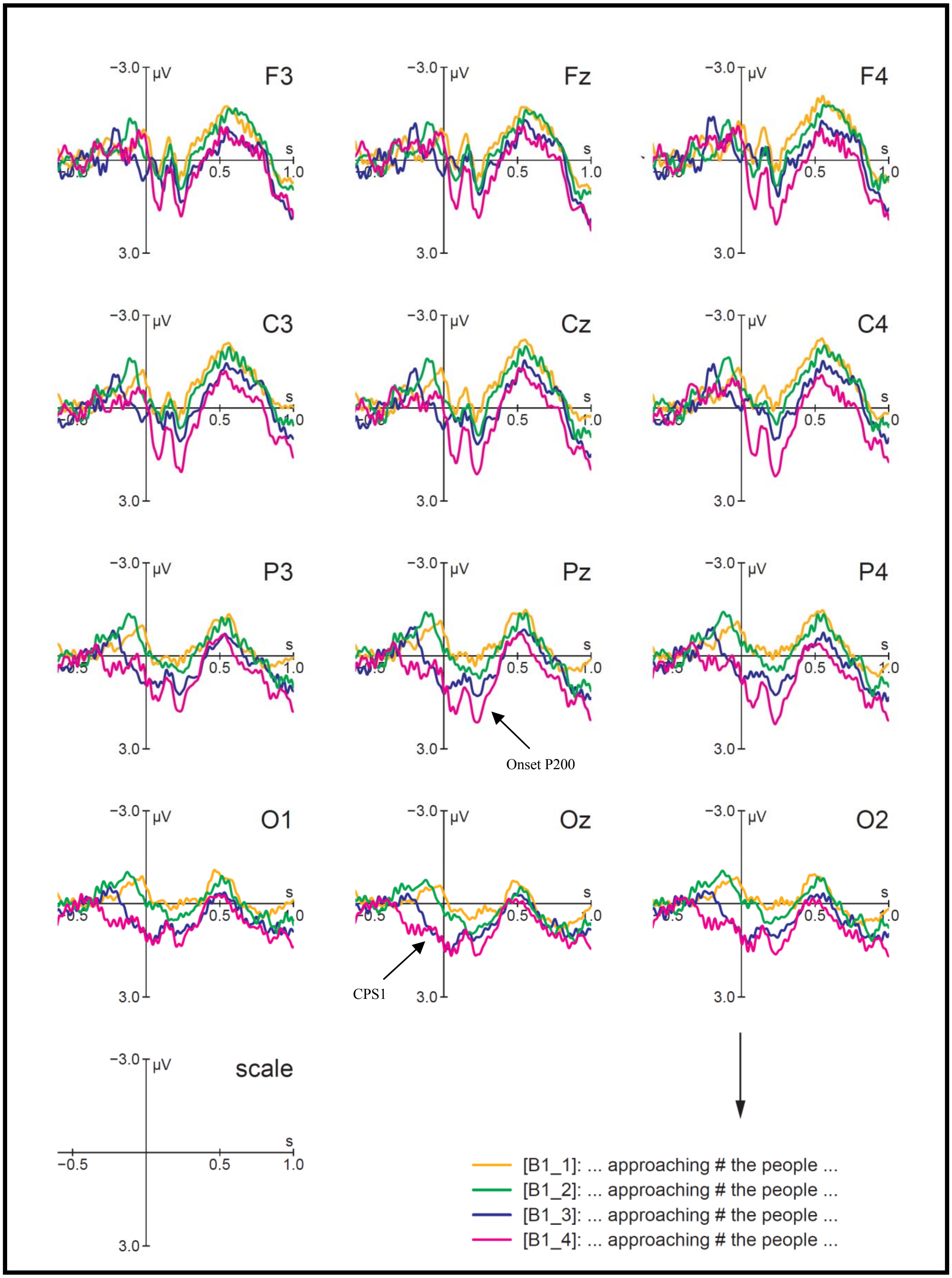


Fig. 3 – CPS1 in conditions [B1_1], [B1_2], [B1_3] and [B1_4] at the first boundary. Grand average ERPs for all four conditions are time-locked to the onset of NP2 ("*the people*"; vertical lines at 0 msec), using a baseline of -520..-320 msec. Condition [B1_4] evokes the largest closure positive shift (CPS) before NP2 onset, with conditions [B1_3] and [B1_2]\[B1_1] eliciting gradually smaller effects. As a result, the onset of the CPS in conditions [B1_2] and [B1_1] occurs with greater proximity to the onset of NP2. In all conditions, the more frontally distributed P200 component, superimposing the CPS, also increases in amplitude with each increasing boundary size.

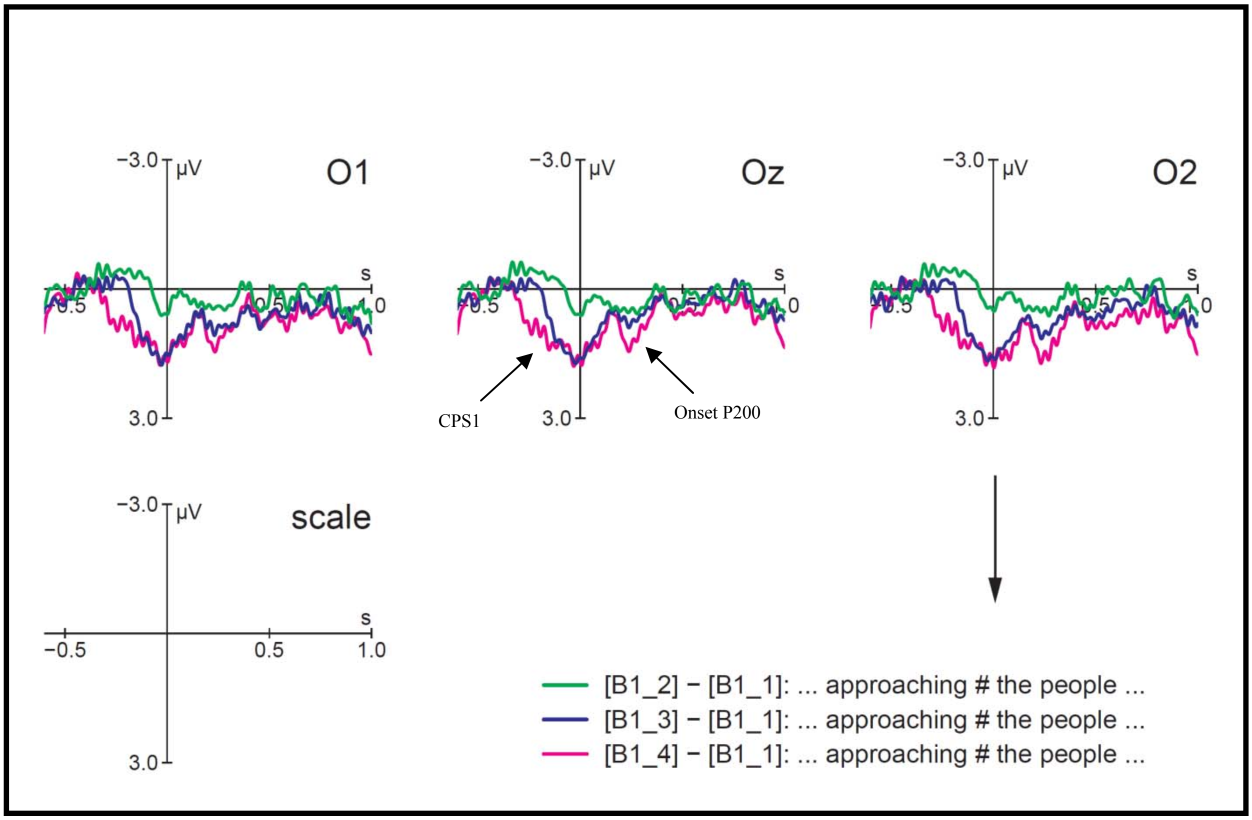


Fig. 4 – CPS1 difference waves at the first boundary between [B1_1] and [B1_4] (pink), [B1_3] (blue), and [B1_2] (green), at posterior electrodes (O1, O2, Oz). Similar to the CPS1 analysis (Figure 3), all conditions are time locked to the onset of NP2 (“*the people”*; vertical lines at 0 msec), using a baseline of -520..-320 msec. This analysis shows that CPS1 is evoked before the onset of NP2, whereas the onset P200 of NP2 (*“the people”*) is elicited around 200 msec, illustrating that the CPS is independent of the P200.

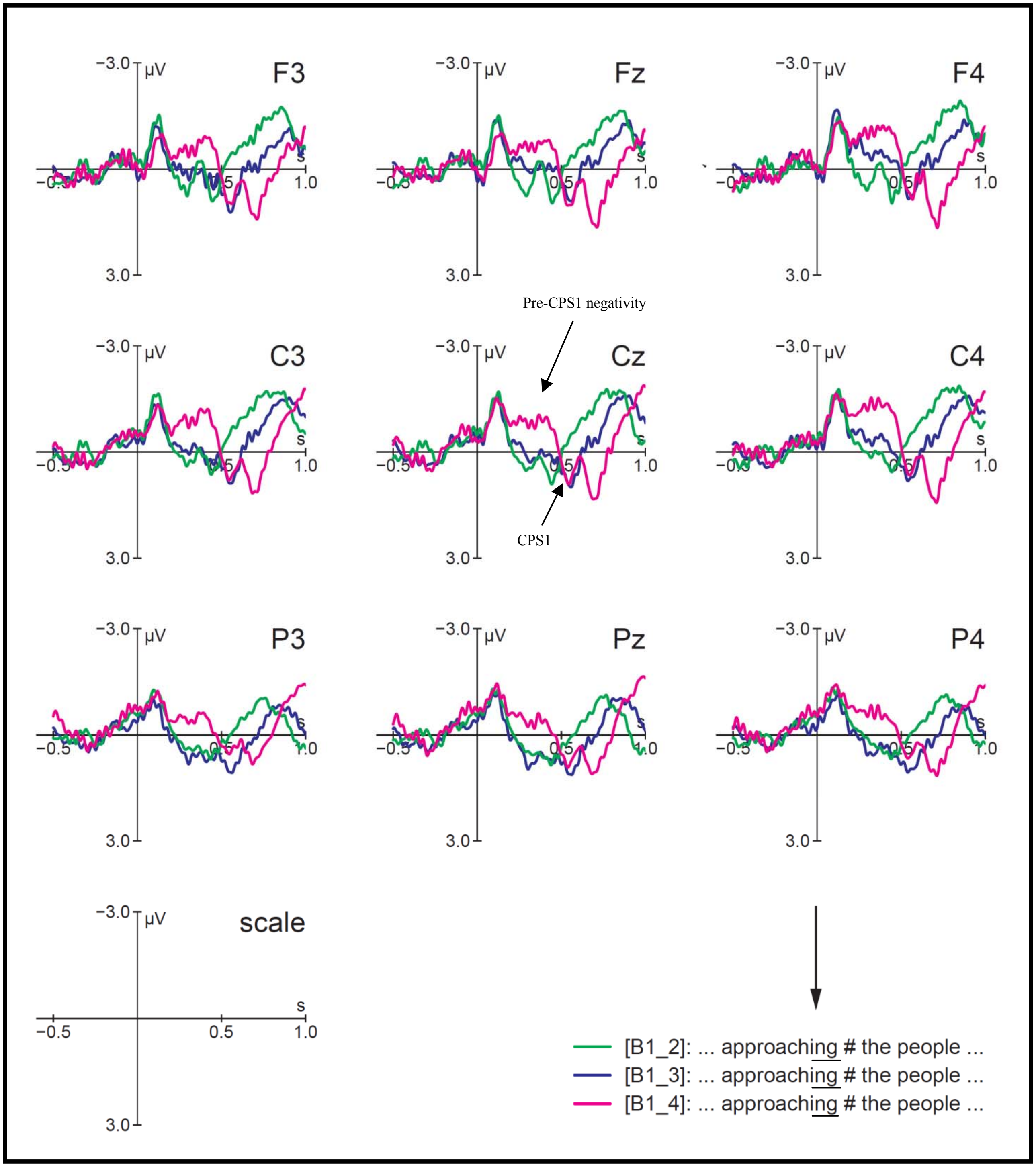


Fig. 5 – pre-CPS1 negativity in conditions [B1_2], [B1_3], and [B1_4] at the first boundary, time-locked to the onset of the pre-final lengthening on the pre-boundary word (PFL1; "approach***ing***"), using a baseline of -300..-100 msec. Condition [B1_4] displays the largest negativity around 150 msec post onset and lasts for about 350 msec before shifting towards the positive range of CPS1. While the effect is broadly distributed, it is most prominent over the frontal and central electrode sites, especially over the right hemisphere. The arrows point to the pre-CPS1 negativity in B1_4 and to the peak of CPS1 in the same condition, which coincides with the P200 in condition B1_3 (due to differences in boundary length).

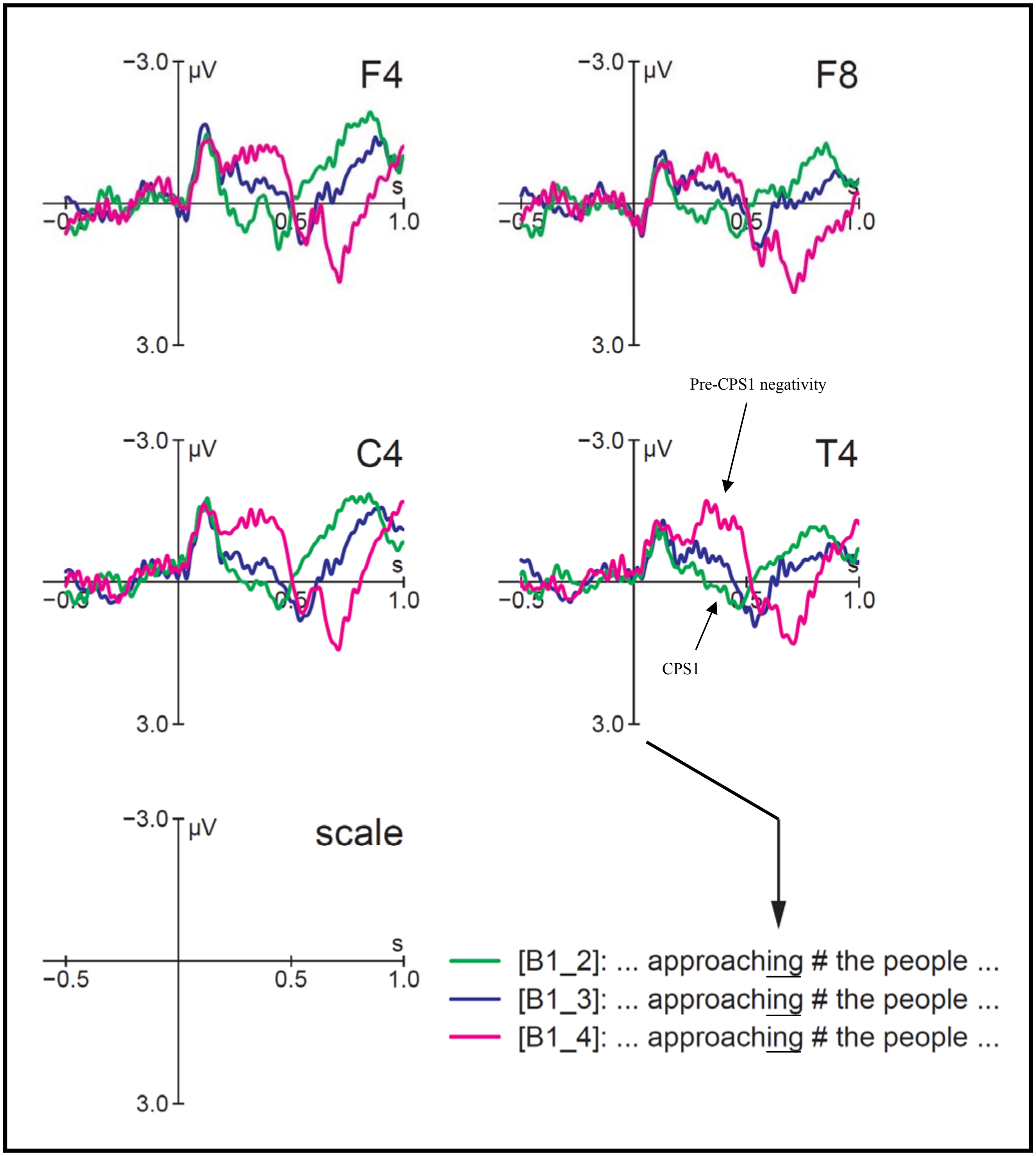


Fig. 6 – pre-CPS1 negativity at selected electrodes F4, F8, C4, and T4 in conditions [B1_2], [B1_3], and [B1_4]. All conditions elicit an early negative deflection ~120 msec post PFL1 onset, after which they begin to diverge. Conditions [B1_2] and [B1_3], with the shorter pre-final lengthening, evoke smaller pre-CPS negativities, whereas condition [B1_4], with the longest pre-final lengthening, evokes the largest negativity. Compared to CPS1, this effect is more anterior, suggesting that the difference between conditions reflects a distinct process related to the amount of pre-final lengthening in each condition, rather than to the delayed onset of CPS1, due to the prosodic manipulation in each condition. The arrows point to the pre-CPS1 negativity in [B1_4] and to the CPS1 in conditions [B1_2] and [B1_3].

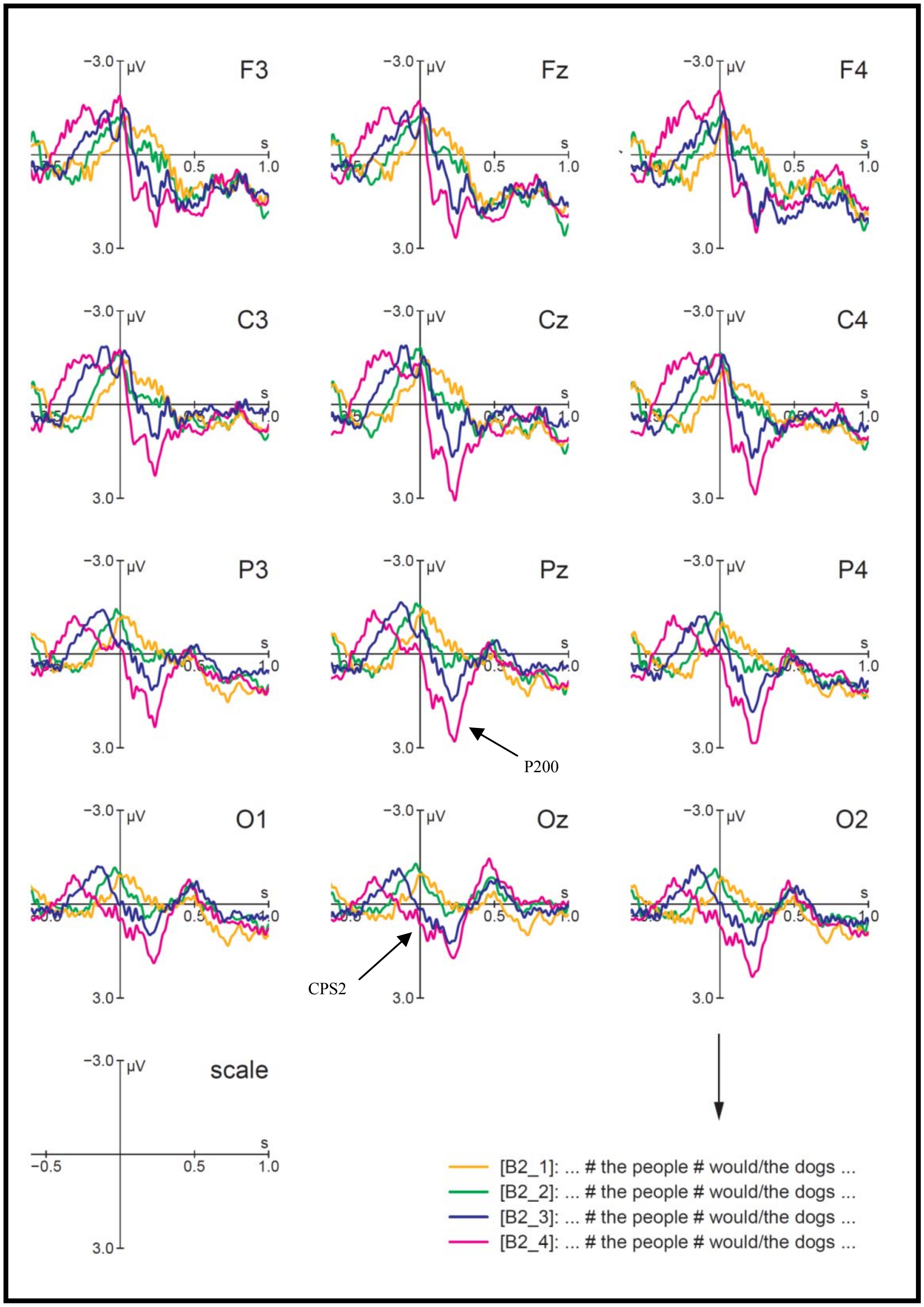


Fig. 7 – CPS2 in conditions [B2_1], [B2_2], [B2_3] and [B2_4] at the second boundary. Grand average ERPs for all four conditions are time-locked to the onset of the post-boundary words ("*the dogs*" in LC, "*would*" in EC; vertical lines at 0 msec), using a distant baseline of -500..-100 msec relative to the onset of PFL2 ("the peo*ple*"). Similarly to CPS1, CPS2 is most prominent over posterior and occipital regions. As in CPS1, the amplitude and latency of CPS2 and the onset P200 gradually increase with boundary size. The enhanced negativity preceding CPS2 may be attributed to an additional N400 component reflecting the lack of a theta role in the 'stranded' NP2 (*"the people"*) which is flanked by prosodic boundaries.

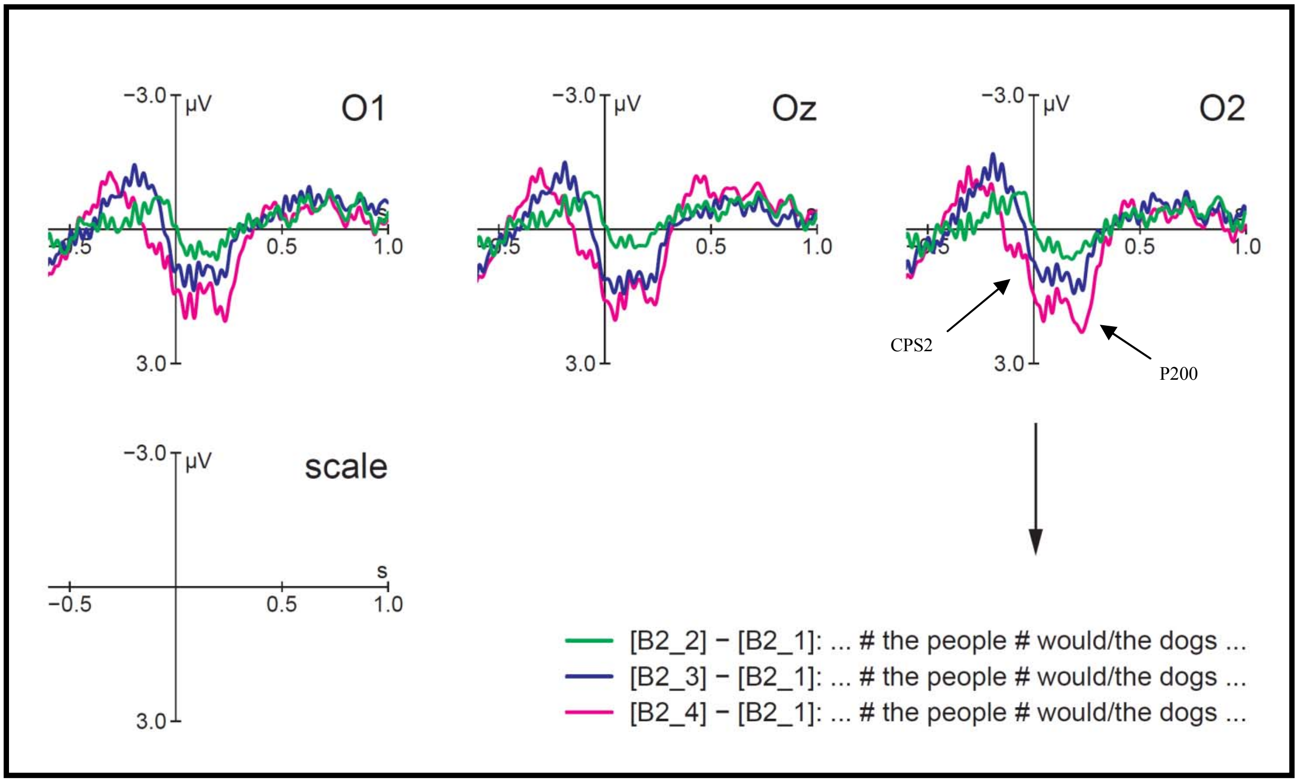


Fig. 8 – CPS2 difference waves between [B2_1] and [B2_4] (pink), [B2_3] (blue), and [B2_2] (green), at posterior electrodes (O1, O2, Oz), using the same baseline and time-locking position as in Figure 7. The plots illustrate that CPS2 is evoked before the onset of the post-boundary word onset, whereas the P200 is evoked some 200 ms after word onset. Thus, similar to CPS1, both components are independent of one another, with the P200 superimposing CPS2.

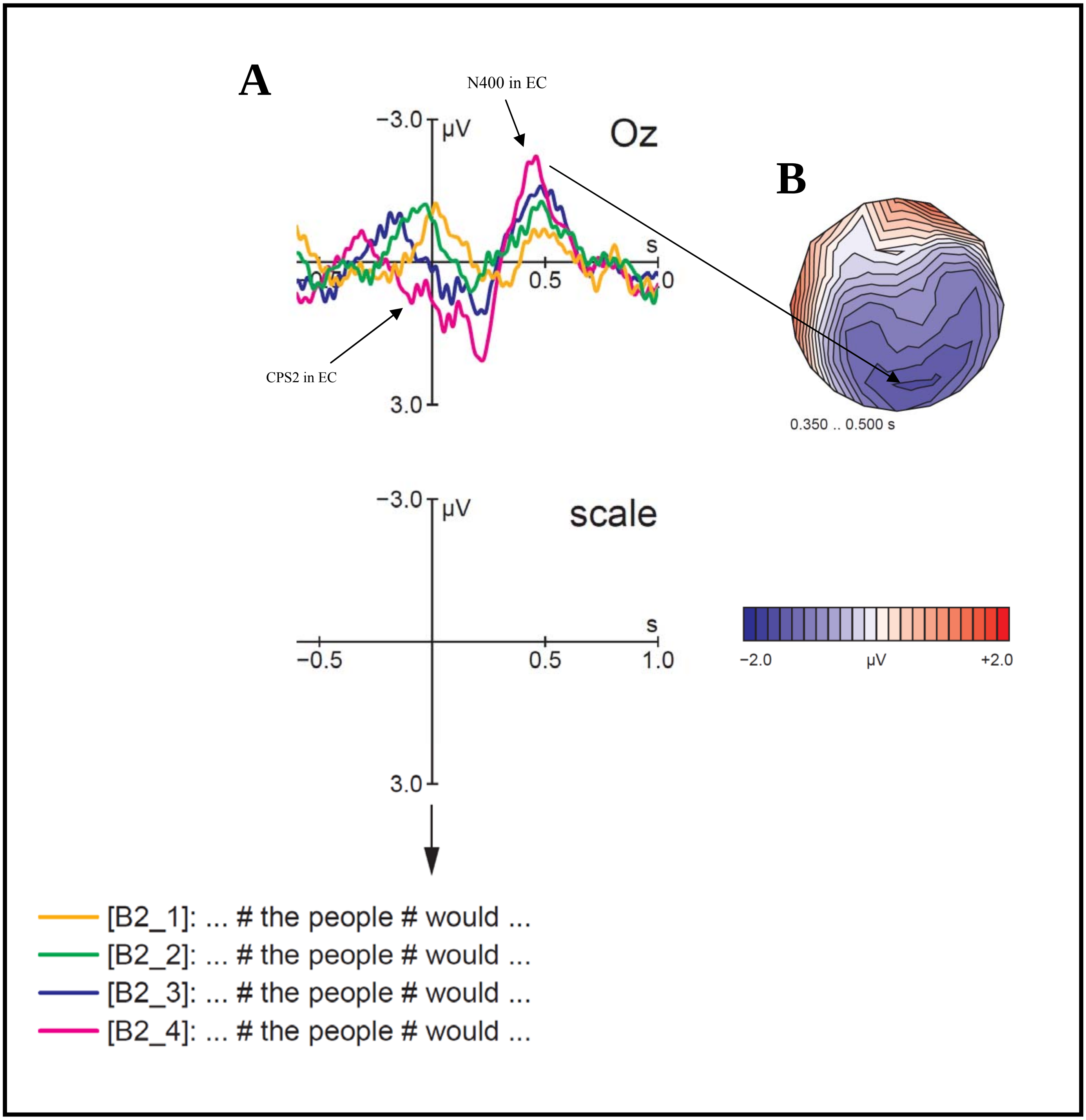


Fig. 9 – Grand-average ERPs in conditions [B2_1]-[B2_4] at Oz, time-locked to the onset of the post-boundary word ("*would*") in EC. The plot illustrates the impact of manipulating the (incompatible) *late* boundary on ERPs at the disambiguating position. **(A)** The prosody-syntax mismatch here elicits an N400-like negativity, which exhibits a graded pattern corresponding to the size of the late boundary across conditions. **(B)** Voltage map of the difference waves ([B2_4] minus [B2_1]) illustrates the posterior scalp distribution of the N400 effect.

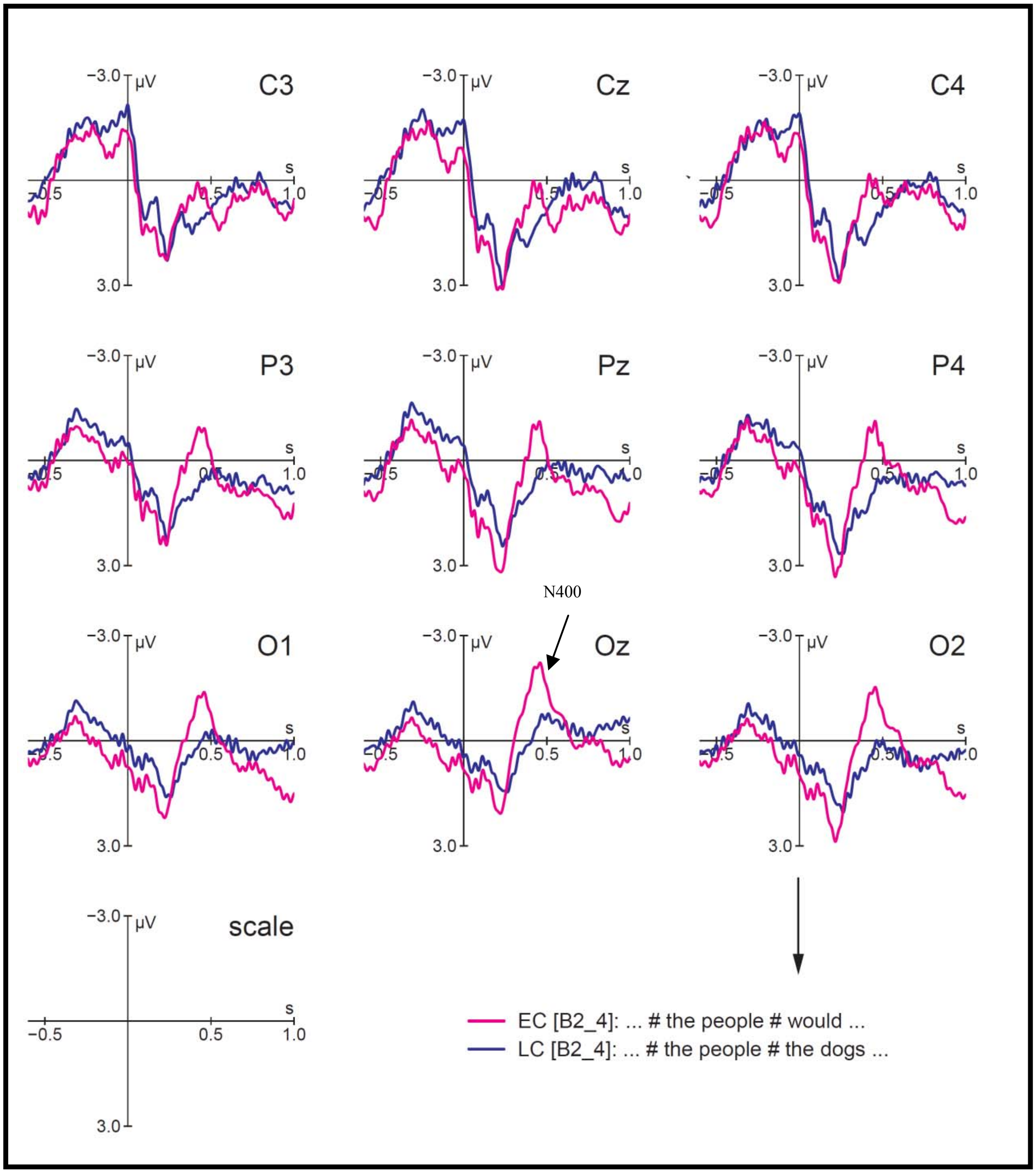


Fig. 10a – N400 in condition [B2_4] in EC (pink) compared to its matched condition [B2_4] in LC (blue). Speech signals in both conditions are identical prior to the disambiguating word (evoking the same CPS2 component), but LC [B2_4] does not elicit a garden-path effect. In this comparison, the N400 effect in EC displays a more typical posterior distribution compared to that in (Figure 9). Note however, that prior to 0 msec, the EC condition is shifted towards the positive amplitude range, which can only be explained as an artifact due to the use of a distant baseline. Figure 10b illustrates the N400 with a corrected baseline.

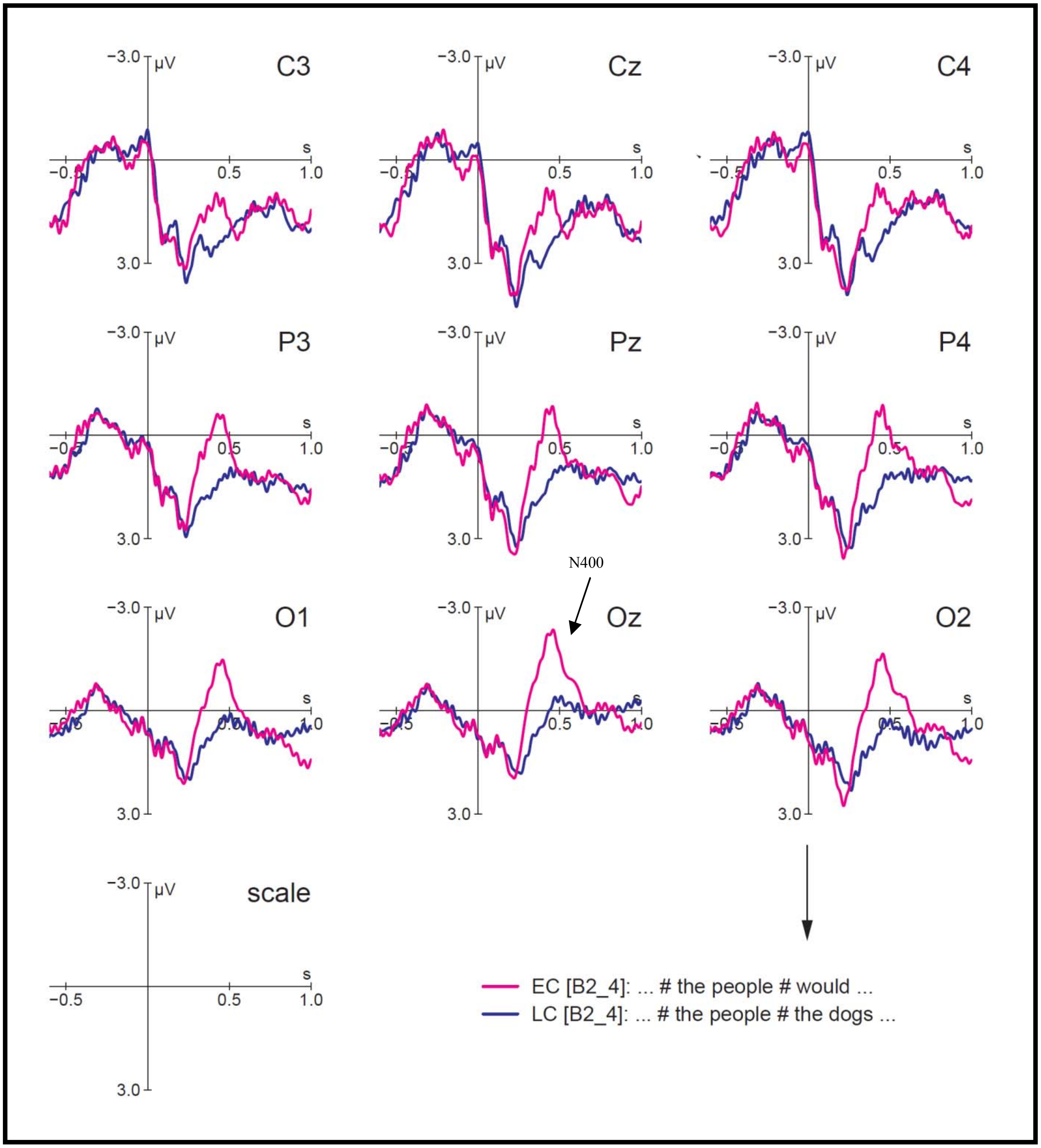


Fig. 10b – N400 in condition [B2_4] in EC (pink) compared to condition [B2_4] in LC (blue). Same plots as in Figure 10a, but now using a -500..0 baseline relative to the onset of the disambiguating words at boundary 2. As expected, with this baseline adjustment the EC and LC conditions are virtually indistinguishable before 0 msec. Moreover, the N400 garden path effect in EC is more prominent than in all previous comparisons, even though its maximal amplitude is still surprisingly posterior (i.e., maximal at OZ).

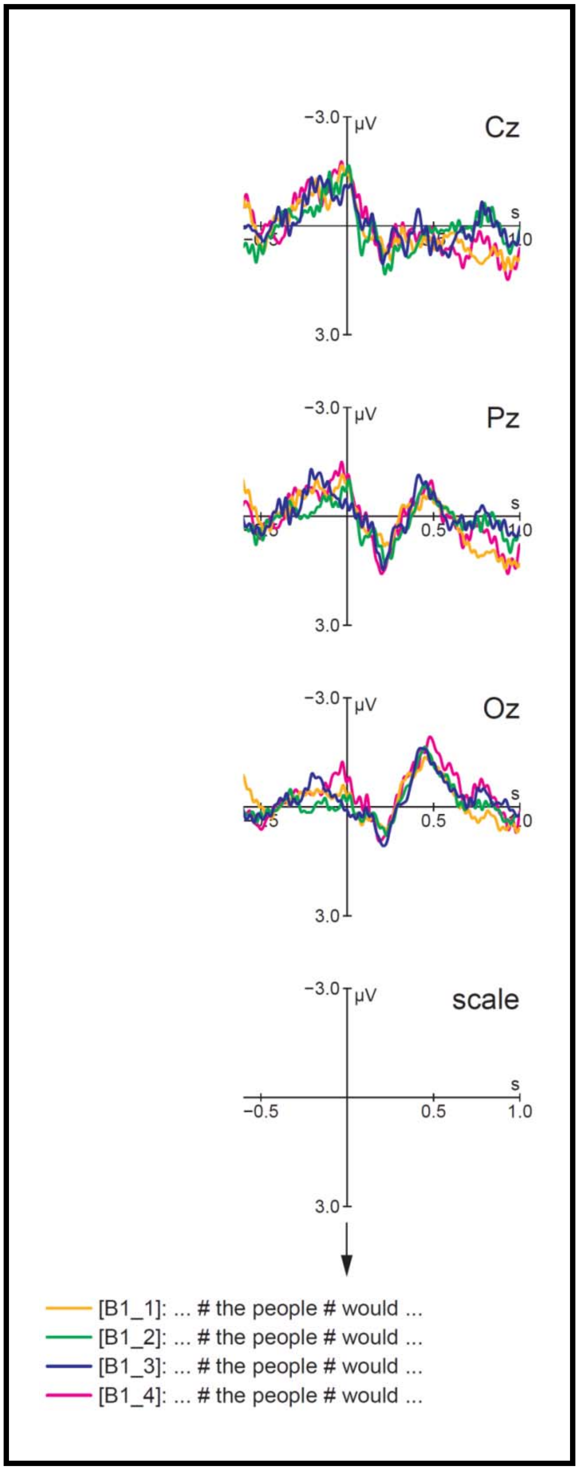


Fig. 11 – A comparison of conditions [B1_1]-[B1_4] in EC (late boundary collapsed), time locked to the onset of "*would*". This comparison illustrates the impact of the (compatible) *early* boundary on the garden path components. Unlike the same comparison in LC (Figure 12), no differences were found between conditions, confirming that the magnitude of the early boundary did not affect the online processing of the EC sentences.

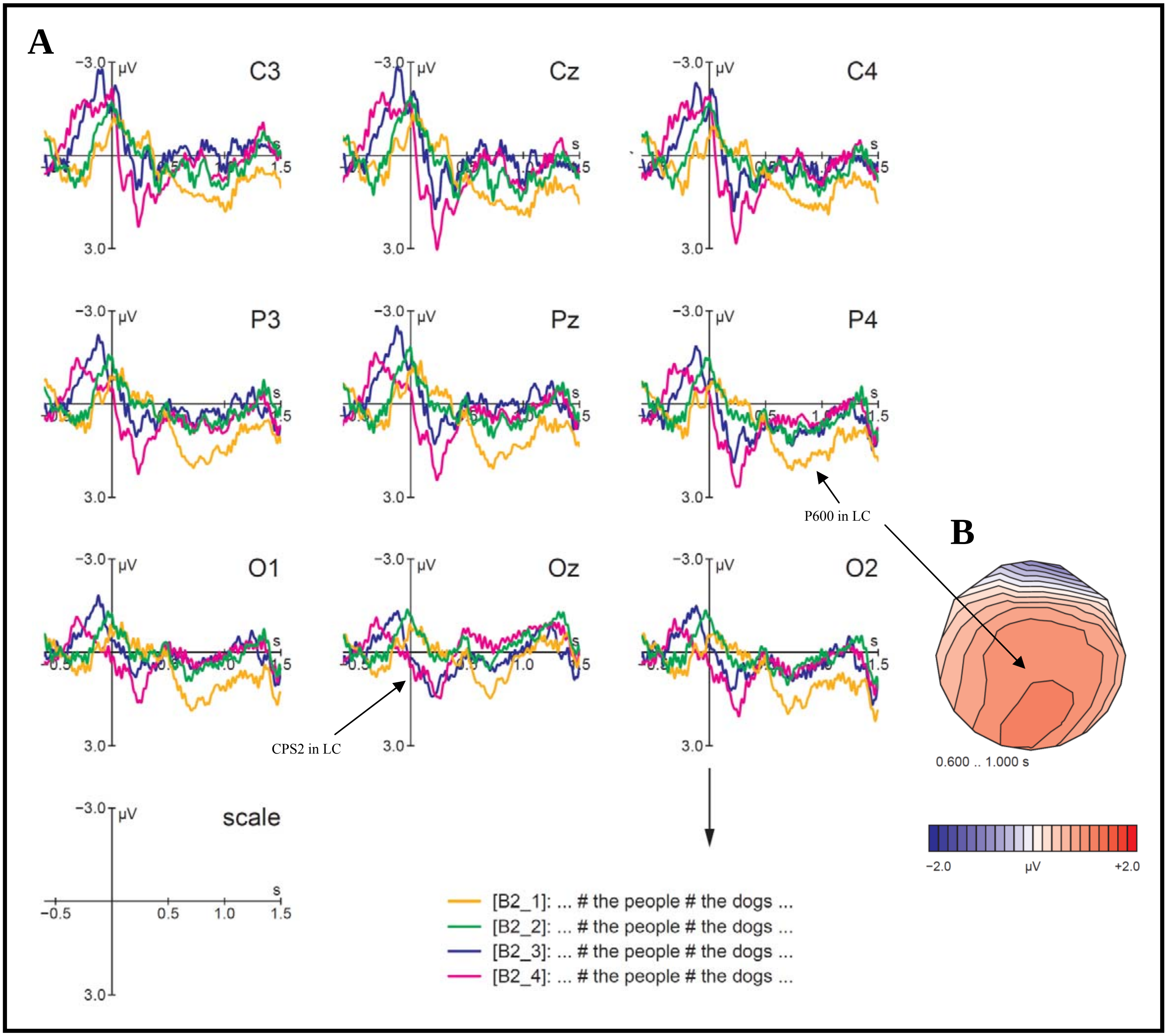


Fig. 12 – Grand-average ERPs in conditions [B2_1]-[B2_4], time-locked to the onset of the disambiguating NP3 immediately following the late boundary ("*the dogs*") in LC. The ERPs illustrate the effect of manipulating the size of the (compatible) *late* boundary **(A)** In condition [B2_1], the prosody-syntax mismatch elicits a broadly distributed P600 with a central maximum compared to the other three conditions. Condition [B2_1] evokes the largest positive deflection. **(B)** Voltage map of the difference waves ([B2_1] minus [B2_4]) suggests a maximum at occipital sites, which is partly due to the local negativity in [B2_4] at Oz in this time window (600-1000 msec).

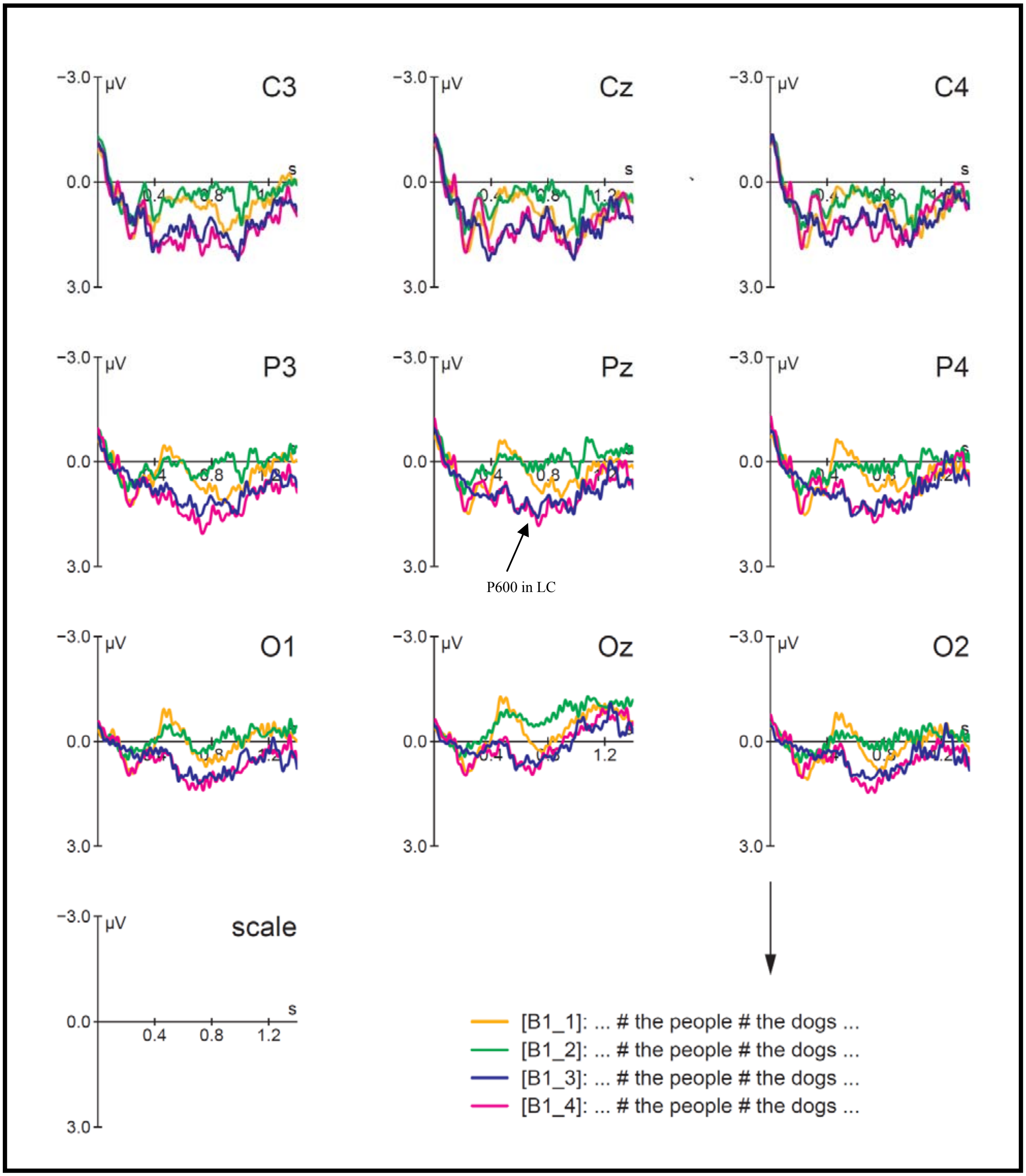


Fig. 13 – A broadly distributed P600 garden path effect in LC in conditions [B1_1]-[B1_4], time-locked to NP3 onset ("*the dogs*"). These ERPs illustrate the impact of the (incompatible) *early* boundary at the disambiguation point. Between 400 and 1200 msec, conditions [B1_4] (pink) and [B1_3] (blue) elicit a larger positivity compared to conditions [B1_2] (green) and [B1_1] (orange).

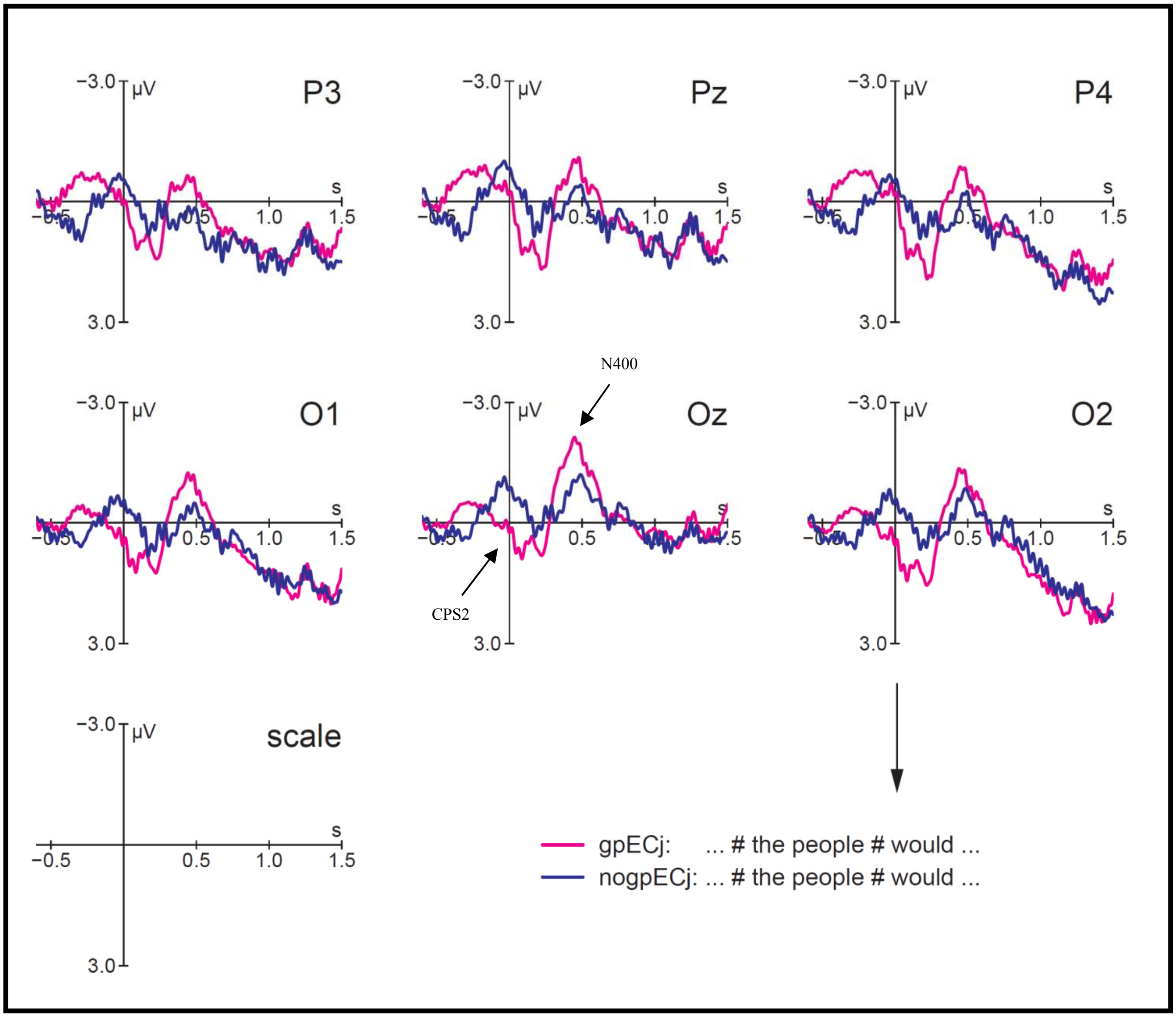


Fig. 14 – A comparison of conditions gpECj ([1_4], [1_3], [2_4] collapsed) and nogpECj ([4_1] [3_1], [4_2] collapsed) in EC, time locked to the onset of "*would*". Condition gpECj displays a posteriorly distributed N400-like negativity between 300 and 500 msec, similar to the one observed in Fig.9.

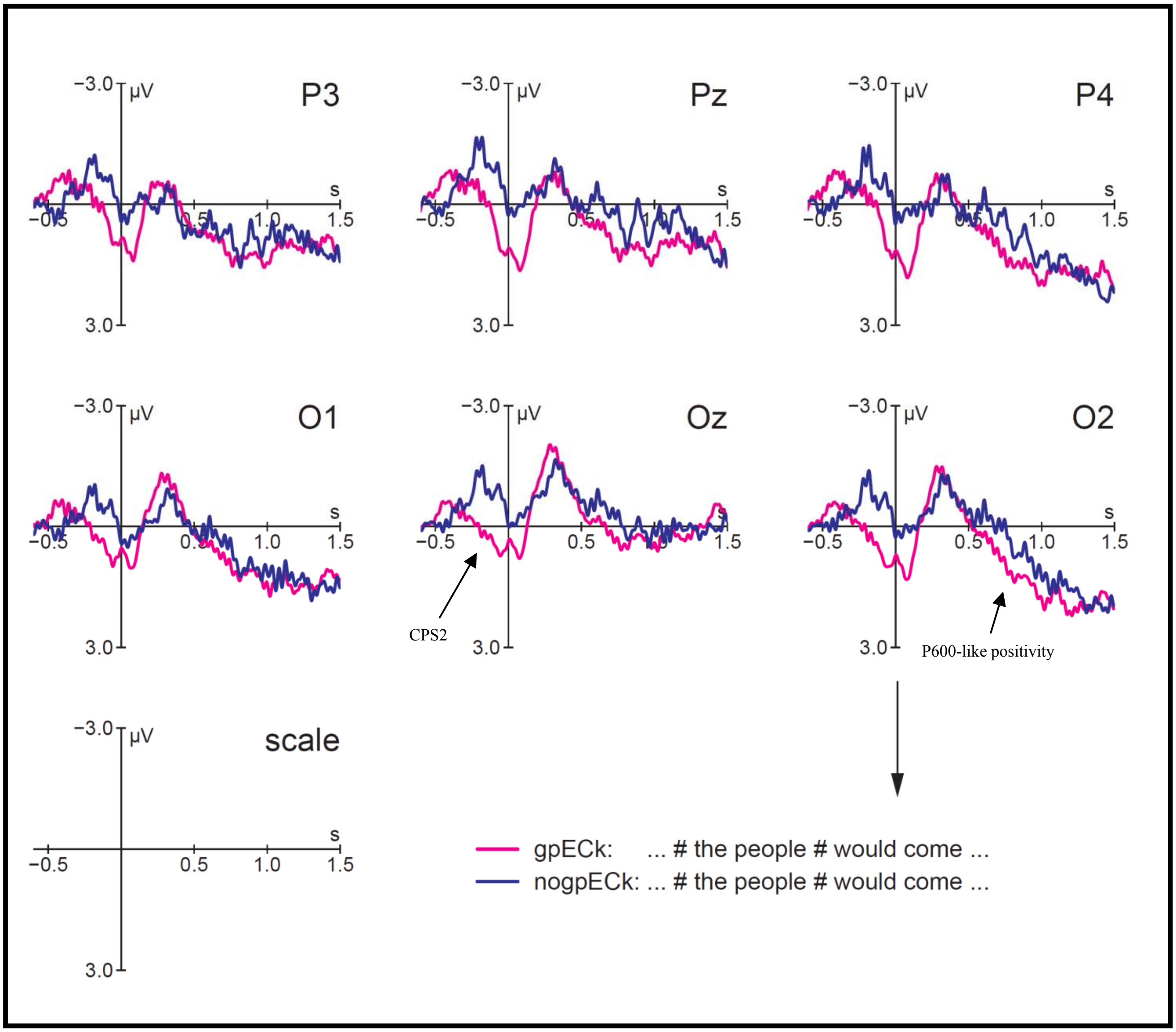


Fig. 15 – A comparison of conditions gpECk ([1_4], [1_3], [2_4] collapsed) and nogpECk ([4_1] [3_1], [4_2] collapsed) in EC, time locked to the onset of VP2 ("*come running*"). Subsequent to a small N400, condition gpECk displays a posteriorly-distributed P600-like positivity after 500 msec, suggesting it partially overlaps the N400 component (Figure 14).

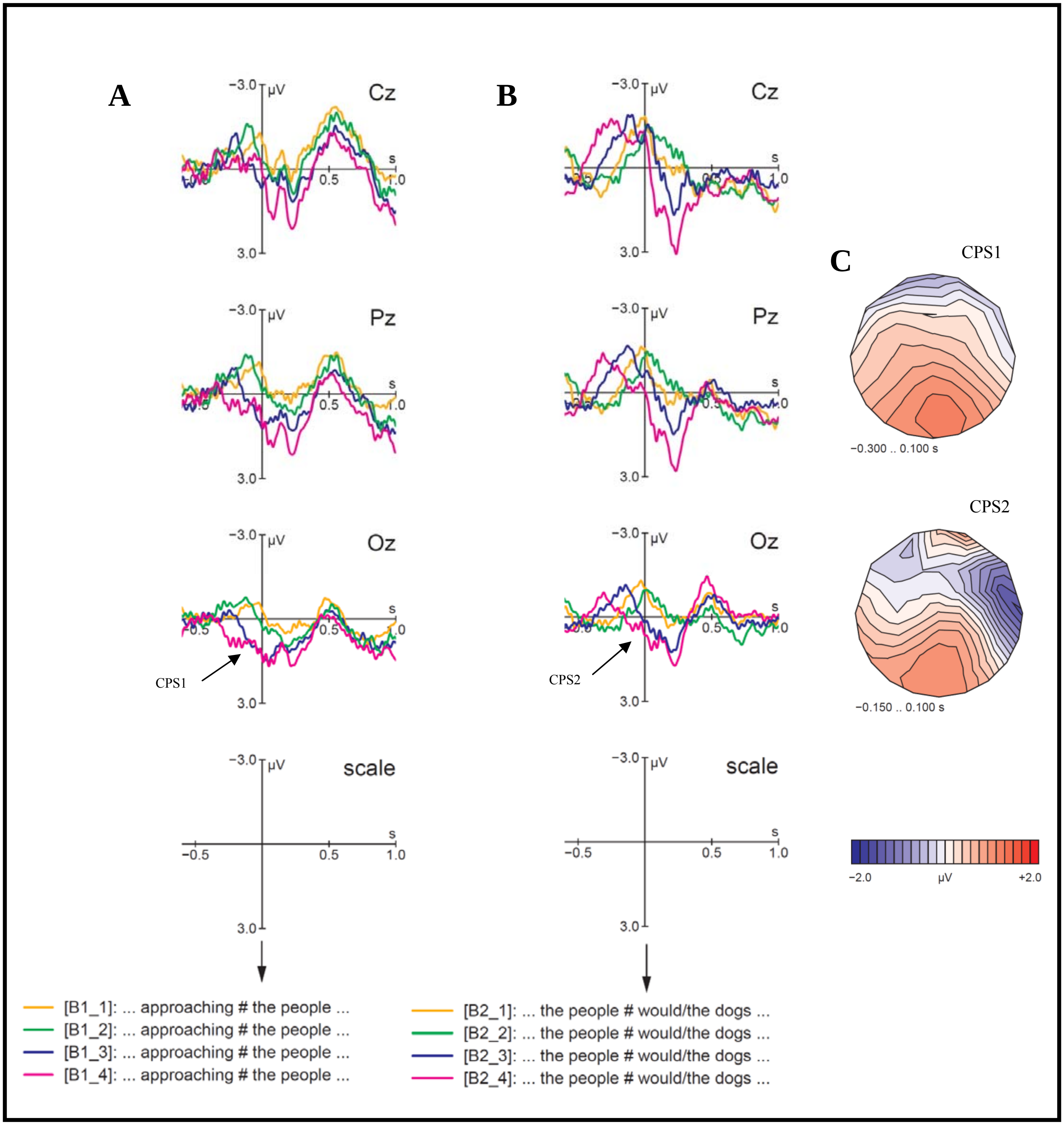


Fig. 16 – Direct comparison between **(A)** CPS1 and **(B)** CPS2 at midline electrodes. At both boundary positions, a posteriorly-distributed CPS is evoked, starting before the onset of the post-boundary word and lasting for some 500 msec. Both CPS components are similar in terms of morphology and scalp distribution. Note, however, that CPS2 is partially superimposed by an enhanced preceding negativity, which may reduce its amplitude as a result. **(C)** Voltage maps of the difference waves illustrate a similar posterior distribution of the CPS effects at both boundary positions.

CHAPTER 4: General Discussion

The current project investigated the impact of two prosodic boundaries of varying sizes on the processing of ambiguous garden-path sentences in English. In particular, we aimed to test: (1) the predictions of the two leading theories, the Anti-Attachment Hypothesis (AAH; Watson & Gibson, 2004a, 2005) and the Informative Boundary Hypothesis (IBH; Carlson et al., 2001; Clifton et al., 2002), assuming either a local or global impact of boundaries on comprehension, respectively; (2) whether gradient quantitative differences among prosodic boundaries exhibit a gradient or a categorical effect on listeners' parsing decisions, both behaviorally and electrophysiologically.

4.1. Testing the predictions of the IBH and AAH, and the extended BDH

Study 1 (Experiments 1 and 2) focused on the predictions made by the two theories (although the behavioral task in Experiment 3 can also be used for direct comparison, as it employed the same stimuli as Experiment 2); in particular, the Study tested the AAH's assumption that large (IPh), but not small (ip), boundaries serve as probabilistic cues to closure, and the IBH's assumption that the relative difference between boundary sizes (e.g., IPh preceding an ip = Early Closure) determines the syntactic preference. Although these predictions seem in conflict at first, when comparing them using the two phrase-level boundary types (ip and IPh), postulated by the categorical view assumed by the theories, we reached the conclusion that both the AAH and IBH are virtually indistinguishable (see General Introduction). This also holds true for their assumptions concerning incremental processing, whereby each boundary is immediately integrated when it is encountered (Lee & Watson, Unpublished manuscript; Frazier, Carlson, & Clifton, personal communication, 2011), and the "canceling-out" effect of two same-size boundaries, resulting in the default syntactic preference (LC in this case). Thus, according to both theories, boundaries are expected to be processed locally and evaluated globally.

With only two possible phrase-level categories, it had previously been impossible to tease the theories apart using two boundaries. Most recently, and in contrast to their 2006 paper (Frazier et al., 2006), the authors of the IBH have admitted that their hypothesis is a more nuanced version of the AAH, rather than a truly different approach (Carlson, 2012, personal communication), and the authors of the AAH have claimed that "theories arguing that boundaries serve as cues to local syntactic structure or serve as points of local processing are not inconsistent with the global prosodic structure influencing syntactic parsing" (Lee & Watson, Unpublished manuscript, p. 39). Despite recent awareness of some common ground, in previous years, the absence of a boundary had been used to differentiate between the theories. For example, when comparing the following prosodic patterns: [0_IPh] vs. [0_ip], the IBH would predict similar outcomes (i.e., late closure/high attachment) because both late boundaries are larger than "no boundary", whereas the AAH would predict the outcomes should differ because only large boundaries can signal closure. However, as we were interested in the interplay between two audible phrase-level boundaries, rather than in the replication of the well-studied influence of a single boundary on processing, we hypothesized it could be possible to differentiate between the predictions of the theories using a continuum of digitally-manipulated boundary sizes. In theory, since the IBH argues that relative differences between boundaries should influence perception, whereas the AAH limits closure effects to only a single category, differences between smaller-size boundaries could, in principle, only be accommodated by a global account of processing (allowing meaningful within/between categories variations).

Another important difference between the current and previous studies was our choice to employ local, rather than global ambiguities. This was motivated by two considerations. First, previous research has shown a strong pre-existing syntactic bias in the material utilized. For example, Carlson et al. (2001) reported an overwhelming low-attachment preference for global ambiguities with adverbial adjuncts, resulting in a rather weak effect of prosodic phrasing, even in conditions with maximal prosodic differences. In order to test a range of boundary sizes, it was imperative to use a structure that is strongly influenced by prosodic phrasing in production as well as perception (Schafer et al., 2000; Wagner & Crivellaro, 2010; Watson et al., 2006); therefore we chose to use EC/LC sentences, which require an obligatory boundary at the end of an initial subordinate clause (either early or late; see Frazier et al., 2006). Second, unlike global ambiguities, temporary ambiguities allow lexical disambiguation during sentence presentation, which can be used to test the influence of the initial, prosody-induced, structural expectation. The combination of these factors proved instrumental in both the behavioral and ERP studies, as reflected by participants' acceptability ratings and brain responses.

One of the most remarkable aspects of the outcomes of all three behavioral experiments was the identical pattern of results. This was achieved with differently manipulated boundary sizes in Experiment 1 compared to Experiments 2 and 3. This pattern was preserved and extended even after having introduced a fourth condition (i.e., "no-boundary"), which followed the same order of gradience in both structures.

The results of Experiments 1 and 2, in particular the unexpected differentiated scoring pattern in each structure, displaying an effect of the size of the incompatible boundary only (e.g., boundary #1 in LC) rather than being symmetrical, were unsupported by either hypothesis, but were accounted for by the BDH (Pauker et al., 2011; Steinhauer & Friederici, 2001). Because the BDH was never tested using a balanced design, where both structures contained a superfluous boundary, this was a first opportunity to examine the influence of an interfering boundary on both structures. We thus incorporated the insights gained from these experiments and extended the BDH such that it can account for the early boundary advantage (EC preference), the gradient processing difficulty, and the perceptual difference between the early and late boundary sizes required to overturn the initial EC preference. We refer to this version of the hypothesis as the extended BDH (or the eBDH). The behavioral results of Experiment 3 replicated and validated these assumptions. Table 1 below summarizes the main findings of the behavioral studies, and compares the results to the predictions of the IBH, AAH, and BDH.

Although the ERP study was not designed to directly test the predictions of the IBH and AAH, due to the need to collapse across conditions in order to assess the CPS at both boundary sites, we did find support for the predictions of the extended BDH. Results showed: (i) gradient differences in the magnitude of the CPS (i.e., amplitude, duration and latency) between the conditions at both boundary locations, (ii) an overall EC preference resulting in a more severe garden-path effect in LC (P600) compared to EC (N400), and (iii) partly graded differences between the conditions within each garden-path effect. Analyses also revealed a consistent impact of the structurally *incompatible* (but not of the compatible) boundary on garden path effects in both LC (early boundary) and EC (late boundary), although these comparisons were partly affected by noise due to the use of a distant baseline. A follow-up study with fewer conditions (to increase the signal-to-noise ratio) would be required to validate this observation.

Table 1. Comparison between the predictions of the IBH and AAH and the extended BDH in light of the behavioral results of all three experiments

Prosodic pattern	*Behavioral findings*	*IBH + AAH predictions*	*Extended BDH predictions*
#1 > #2	EC preference	EC preference	EC preference (#1 advantage)
#1 < #2	EC preference (except when #2 was 2 steps larger than #1 – Exp.2, 3: [14], [13], [24])	LC preference	EC preference (#1 advantage), except when #1 and #2 are maximally different
#1 = #2	EC preference	2 IPh boundaries cancel each other out resulting in baseline preference (LC)	EC preference (#1 advantage)
Categories vs. Gradient	Gradient effect of boundaries between all prosodic levels	Categorical effects. No differences should be found between variants of the same category	Processing difficulty should gradually increase with increasing superfluous boundary size
EC and LC acceptability	Differentiated effect of boundary location on structure rating (EC - #2; LC - #1)	Symmetrical pattern of results (e.g., #1 should increase acceptability for EC and decrease acceptability for LC to the same extent, and vice versa - X% EC, 100-X% LC)	Processing difficulty is associated directly with the magnitude of the superfluous boundary: EC scores decrease when #2 is larger; LC scores decrease when #1 is larger

4.2. The processing of prosodic boundaries

Perhaps the most central question that has been motivating the research on prosodic phrasing, especially in the context of multiple boundaries of varying sizes, is: how are boundaries processed? In the literature, the effect of boundaries is described using words like: “chunking” and “grouping”, but also “dividing” and “separating”. These terminologies often reflect the underlying assumptions regarding the effect of boundaries on parsing. Despite showing great similarity with respect to their overall predictions, the IBH and AAH differ in their view on this particular issue. Consistent with the view of many researchers, the IBH assumes that boundaries serve to perceptually group syntactic constituents together (see also Frazier & Clifton, 1997; Kjelgaard & Speer, 1999; Pynte & Prieur, 1996; Schafer, 1997; Speer et al., 1996). One of the main types of evidence used to support this claim is the finding of increased processing difficulty (in terms of RT and acceptability) when structurally ambiguous constituents have to be attached to a different, as compared to the same, prosodic domain, thereby crossing a prosodic boundary (Schafer, 1997). Since prosodic domains are delimited by prosodic boundaries, it is postulated that boundaries create a facilitating processing environment for the grouped constituents. The AAH, on the other hand, argues that boundaries serve to decrease the likelihood/strength of a potential attachment site rather than strengthen it (Lee & Watson, unpublished manuscript; Watson & Gibson, 2005). As described by Watson and Gibson (2005; p. 286): "Central to the AAH is the idea that intonational boundaries cause a break between constituents. That is, boundaries create biases that force constituents apart during the parsing process. This contrasts with many of the theories in the literature that claim that intonational phrasing's primary role in comprehension is to group relevant constituents together". Using the same type of evidence mentioned above, Watson and Gibson (2005) claim that IPh boundaries mark the ends of constituents, thus signaling to the parser that no further attachments are expected.

Since previous data was mainly based on a binary response (forced choice) paradigm, this rather subtle distinction was difficult to evaluate. By increasing the scoring range within each structure (using a rating scale), the present studies allowed an in-depth inspection of the influence of boundaries on comprehension. The most salient effect observed in all three behavioral studies is the interfering, rather than supporting, effect of prosodic boundaries on sentence acceptability. That is, instead of increasing the scores of their compatible structures (e.g., early boundary in EC), we observed a decrease in acceptability triggered exclusively by the incompatible (superfluous) boundary, whereas the size of the supporting boundary made no difference for acceptability. Thus, in keeping with both the AAH and BDH, the outcomes indicate that boundaries serve to weaken (or tax), rather than strengthen (or aid), syntactic preference, with larger boundaries posing greater difficulties to comprehension (i.e., scores pattern: 1 > 2 > 3 > 4). Note, however, that unlike the BDH, the AAH limits this effect to only IPh boundaries, while the present findings reveal that smaller boundary sizes influence parsing in a similar, albeit gradient, manner. This latter observation is – in principle – in accord with the spirit of the IBH; however, the current version of the IBH has adopted the categorical view of boundary perception and cannot account for gradient differences beyond these categories.

Another central finding, which cannot be accounted for by either the IBH or the AAH, is the primacy effect of early boundaries, which seemed critical to the processing of the entire sentence (at least in EC/LC structures), and was extremely difficult to overturn. Only a much larger (maximally distant) late boundary was sufficient to reverse the effect of even the smallest of the early boundaries. Even then, the influence of the early boundary lingered, modifying the extent of this reversal, with larger earlier boundaries lowering acceptability to a greater degree than smaller boundaries (i.e., [1_4] > [2_4]). In addition, when the late boundary was sufficiently large to reverse the EC bias, a larger magnitude did not change the acceptability score, which seemed to be mainly driven by the size of the early boundary (i.e., [1_4] = [1_3]).

There could be several reasons as to why EC was so strongly preferred. One possibility, which we have argued against in Study 1, is the likelihood of EC being the baseline structure due to the progressive aspect of the optionally transitive verb in the subordinate clause (approaching; see Frazier, Carminati, et al., 2006). First, these verbs were also used in their progressive form in an earlier study (Pauker et al., 2011) specifically in order to negate the general bias towards transitive use, based on a corpus analysis (dervied from Lapata et al., 2001), and showed no differences in preference between well-formed EC and LC sentences. More importantly, an EC condition without any boundaries turned out to be more difficult than that with the compatible boundary, suggesting a potential LC preference *despite* the use of progressive verb forms. Second, if the first verb already guided interpretation towards EC, no differences in acceptability between structures should have been found at any of the Early Boundary levels, as the material was identical at this stage. Nevertheless, we found significant differences in acceptability scores between EC and LC in Early Boundary size levels 3 and 4, but not in 1 and 2 (see Chapter 2, Discussion sections). These results indicate that the initial syntactic preference was primarily driven by the prosodic structure rather than a pre-existing lexical bias. This is in line with a study by Itzhak et al. (2010), showing that overt prosodic phrasing overrides such structural biases. However, the latter study also showed that biases independent of prosodic patterns exist. For example, it is conceivable that the Late Closure principle (Frazier, 1987) itself is a prosody- rather than syntax-driven strategy ('avoid assuming boundaries unless you encounter positive evidence'), where 'positive evidence' (e.g., measured in terms of boundary strength) may well be modulated by factors such as transitivity biases or aspect. Another possibility is that a processing strategy was employed by the participants, due to the repetitive nature of the sentence material. Since all sentences required a boundary to indicate the closure of the subordinate clause, even a small indication of one was adequate to trigger closure. Another option is that the lower scores for LC reflected the low likelihood for another boundary in a sentence requiring one obligatory boundary (Clifton et al., 2006; Frazier et al., 2004). Alternatively, our findings may in fact reflect prosody-driven processing as hypothesized by the BDH. In the following section we outline an account concerning the manner by which boundaries are processed throughout the sentence, and how they affect parsing over time.

Given our sentence material consisted exclusively of EC/LC structures, we propose that listeners may have become aware of this dichotomy (at least to a certain extent).[22] Therefore, each sentence began with both structures at a maximal activation level. During sentence presentation, each boundary gradually lowered the activation level for the competing structure. The early boundary, however, presented a considerably larger penalizing effect compared to the late boundary. Thus, at sentence offset, the activation level of the EC structure was generally higher than that of the LC structure, resulting in an overall EC preference, as reflected by the main effect of Prosody found in all studies. Since each sentence contained a conflicting boundary by default, the activation of each structure was lowered in each trial, as illustrated by the overall low scores found across the board (compared to previous studies using similar material with a single boundary; e.g., Pauker et al., 2011)[23]. We illustrate this hypothesis using the processing of both EC and LC structures in three different prosodic conditions from our study: [2_2], [4_2], and [2_4], in Figures 1A, 1B, and 1C below (acceptability percentage is taken from the results of Experiment 2).

[22] The initial assumption that participants in our study developed some awareness of the possible structures may be realistic, but does not seem critical at all. The subsequent processing steps would work in a very similar manner, even if the 'initial state' were characterized by a slight bias towards LC (be it due to the LC principle, lexical biases of the verb, or other factors), or towards EC (e.g., due to the progressive verb forms).

[23] Note that the assumption of a lower level of acceptability ratings is largely based on speculations, and may not be true. That is, a average score of '3' on a scale from 0 to 4 could be viewed as 75% acceptability, however, if a '3' translates into 'quite acceptable' this might actually correspond to 100% acceptability in a binary forced choice task. Another problem with this comparison is that subjects may have avoided extreme scores in order to be prepared for 'much better' or 'much worse' sentences (that were not included in the experiment).

Figure 1.

(A) Activation levels and corresponding acceptability ratings for EC and LC in condition [2_2].

(B) Activation levels and corresponding acceptability ratings for EC and LC in condition [4_2].

(C) Activation levels and corresponding acceptability ratings for EC and LC in condition [2_4].

In all conditions, the early boundary reduces the activation level of LC to a greater amount than the late boundary reduces the activation of EC. This is best illustrated by the higher rating percentage for EC in condition [2_2] (Figure 1A), where both boundaries are of the same size. When the early boundary is larger (Figure 1B), the difference between the two structures is further increased, lowering the scores for LC considerably. However, the opposite prosodic pattern (Figure 1C) shows that the largest late boundary decreased the activation of EC to a lesser amount, as shown by the similar acceptability scores in both structures. Note that as the activation level of EC is only modulated by the late boundary, it receives the same scores in conditions [2_2] and [4_2] (67% - Figures 1A and 1B). By contrast, LC, whose activation level is only modulated by the early boundary, receives the same scores in conditions [2_2] and [2_4] (~52% - Figures 1A and 1C).[24] The last processing stage in this scenario takes place when the disambiguating word is being encountered. According to the eBDH (and largely in line with our ERP data), the incompatible boundary has to be deleted in order to recover the required structure.

4.3. Phonological representations of prosodic boundaries

The debate between the categorical and gradient approaches to boundary size stems from an assumed underlying architecture of prosodic structure. The categorical approach is based on theoretical accounts hypothesizing a hierarchical organization with two phrase-level prosodic categories (Nespor & Vogel, 1986; Selkirk, 1984; although see Selkirk, 1986, for 3 levels; see Shattuck-Hufnagel & Turk, 1986, for review), whereas the gradient approach is based on empirical evidence showing naïve listeners' sensitivity to a range of prosodic boundary strengths, when given the option to rate them on a scale (de Pijper & Sanderman, 1994; Sanderman & Collier, 1996, 1997; Wagner, 2005; Wagner & Crivellaro, 2010). Our data show

[24] If we assume that the 'initial state' of the sentence parser is not 'neutral' but slightly biased towards LC (be it for a prosodic or syntactic 'Late Closure' principle or for a lexical bias, or both), then the primacy effect (i.e., the advantage of the early boundary in deactivating LC) may have to do with the inhibition and dis-inhibition of mental operations. That is, once an initial preference has been abandoned, it may be more difficult to reactivate this option (requiring a stronger inhibitor for its competitor).

that similar gradient boundary perception can also be observed when listeners' attention is attuned to comprehension. The fact that the graded pattern of results was consistent across all three experiments supports a non-categorical architecture that incorporates a larger range of boundary sizes. Recall that much criticism was directed towards ToBI's low inter-transcriber agreement with respect to the labels assigned to pitch accents and edge tones (Hirst, 2005; Syrdal & McGory, 2000; Wightman, 2002; Wightman et al., 1992). In addition, labelers are instructed to use the spectrogram as visual aid for categorization, instead of labeling the perceived input, and to combine specific acoustic cues, even if they are not actually perceived as such, because the category should always be defined by them (ip – break index 3. No boundary tone). Although these guidelines were designed to decrease labeling time and increase efficiency, as ToBI labeling takes up to 100-200 times the actual speech segment length (Syrdal, Hirschberg, Beckman, & McGory, 2001), it farther distances the labeled categories from their acoustic reality. Wightman (2002) raises the question: If prosody is a natural element of speech, why can only expert labelers transcribe it? The present studies provide clear evidence that listeners show high levels of agreement concerning acceptability when attending only to the acoustic stimulus, even when presented with a large number of prosodic conditions (9 in Experiment 1, and 16 in Experiments 2 and 3).

Taken together, our findings suggest that prosodic categories, if they exist, most likely take another form than assumed thus far by the categorical view. One possibility is that many instances of the phonological phrase exist, as hypothesized by Selkirk (1984), with each representing a different magnitude within the category (p.29):

"… language may exhibit more than one level of phonological phrase, in which case finer terminological distinctions can be made: PhP^1, PhP^2, . . . , PhP^n. With this terminology then, an intonational phrase is a special case of a phonological phrase, one that is associated with a characteristic tonal contour and that has an important function in representing the "information structure" of the sentence. The unit utterance, if it existed, would also be a phonological phrase in this sense".

Selkirk's claim may tap into the present results, as the manipulated boundaries were all derived from the ToBI-defined ip category. Since the boundaries differed with respect to their manipulated duration, but not tonal contours, the categorical view may treat them as variants of the same category, which should not exhibit different effects (Carlson et al., 2001). Alternatively, a continuum of boundary sizes exists, with each boundary size meaningful to processing. A final possibility is that categorical distinctions do exist, but are far more numerous than previously assumed. To test which of these representations is better accounted for, two main strategies can be employed. First, one could vary the increments used to manipulate boundary sizes in order to reach floor and ceiling effects. That is, one might investigate the minimal and maximal size difference between boundaries that can significantly affect comprehension or stop being meaningful to comprehension. This kind of manipulation may also help us better understand what the exact requirements are for a late boundary to be strong enough to override the effect of a preceding boundary. How is this relationship defined? Which factors (and acoustic cues) play a role? Is it possible to calculate the acoustic (rather than phonological) characteristics of a late boundary that can override the early one, *based on* the acoustic characteristics of the early one? And is such a relationship specific to a given syntactic ambiguity? Second, one might examine a larger array of boundaries to evaluate whether categorical distinctions in fact exist. This can be done by increasing the boundary size grid, and by examining what is the minimal distance between each level that is still informative to processing. Speech rate and constituent length may also serve as influencing factors.

4.4. Limitations and future directions

A potentially important issue that should be addressed is the lower-than-expected acceptability ratings in all three studies (but see footnote 2). In contrast to previous studies, where natural productions of IPh boundaries received high acceptability ratings (90%) in both EC and LC structures (Itzhak et al., 2010; Pauker et al., 2011), here the overall scores were considerably lower across the board (not higher than 60-70%). Various reasons may underlie this inconsistency. The most obvious explanation is that each of our sentences contained two prosodic boundaries, which has been previously found to induce garden-path effects in such structures, where a single disambiguating boundary is expected (Pauker et al., 2011). It could also be argued that while the use of a scale increased the response range, it possibly centralized the responses, as participants may have tended to avoid the extremes (although they were explicitly asked to make use of the entire scale). Another explanation concerns the ecological validity of the boundary types employed here. That is, speakers, admittedly, are not likely to naturally produce these sentences with the same boundaries used in the present project, both in terms of size and intonation. The pitch contour selected for the sentence material (H*L-) was adopted from previous research (Kjelgaard & Speer, 1999; baseline conditions), where it was originally designed to be ambiguous, albeit equally acceptable, in both EC and LC structures. This was a necessary compromise made to ensure all sentences contained the same exact prosodic information while being sufficiently acceptable. However, it also sacrificed the degree to which these boundaries sounded natural (and maybe even informative) to listeners.

Nevertheless, our findings show a completely graded acceptability pattern, mirroring the parametric manipulation of both boundaries, which reflects the manner in which listeners processed these boundaries.

A second, related issue concerns both the (potentially) low acceptability ratings as well as the global EC preference. What options may exist to increase the acceptability of LC structures compared to EC structures, especially when the former contain an early boundary? Given the various factors contributing to the EC preference, using different boundary combinations and replacing progressive verb forms with simple present or past tense would likely make a difference. For the greater challenge of increasing the acceptability for LC sentences that contain early (i.e., incompatible) boundaries, the introduction of alternative 'motivators' for boundaries may provide a solution. According to the Rational Speaker Hypothesis as well as the IBH, it should be possible to make boundaries 'less informative' to the syntactic parser. That is, prosodic boundaries are not exclusively motivated by syntactic constraints, but also by length and symmetry constraints.

Gee and Grosjean (1983; p.416), for example, found that when reading written sentences aloud, speakers tended to divide the utterances by creating phrases that are "more or less symmetrical (or balanced). That is, the main pause break is located close to the middle of the sentence; then, each segment on either side of the break is itself broken up into more or less equal parts and so on". Relating to this observation, Fodor (1998) argued that the reason balanced phrasing is important is the need to reconcile the recursive nature of syntactic information with the non-recursive, or flat organization of phonological information (but see Wagner, 2005, for an argument regarding recursion in prosodic representations), to allow an interface between the two structures. This is enabled by constructing symmetrical syntactic representations (or syntactic sisters). Fodor (1998; p. 302) claims that the reason for the abundant evidence for balanced phrasing preference in the literature, concerns the nature of phonological processing:

> "I suggest, therefore, that the same-size-sister constraint is prosodic in origin, and that the packaging mechanism is in fact the prosodic processor. It shuttles through the sentence in tandem with the syntactic processor, constructing phonological phrases on the basis of fairly low-level lexical and syntactic information together with whatever suprasegmental cues the input contains. The phrasing that it imposes then influences the higher-level decisions of the syntactic parser. In some cases the grammar insists that syntactic boundaries coincide with phonological ones; these are the alignment constraints of prosodic phonology"

In light of this argument, a future direction would be to test whether it is possible to modify constituent length such that both structures, and LC in particular, would exhibit higher acceptability scores, even when they contain 'syntactically implausible' boundaries. One way to test this assumption is to construct EC/LC structures containing syntactic sisters (and two corresponding boundaries), such that one boundary is syntactically motivated, whereas the other one is motivated by length and symmetry considerations. In example (1) below, the modified version of one of our sentences is expected to be more acceptable as LC structure (1b) than the original version.

(1) Whenever the scary brown bear was approaching $_{\#1}$ either the old park ranger or his young apprentice $_{\#2}$ …
   a. … would come running (EC)
   b. … the dogs would run away (LC)

Previous experiments have reported that boundary length also modulates the perception of prosodic boundaries, with longer constituents motivating the appearance of boundaries (Clifton et al., 2006; Hwang & Schafer, 2009). Since speakers are assumed to produce prosodic boundaries intentionally, rather than arbitrarily (Frazier et al., 2006), increasing the length of syntactic constituents should signal that not only syntactic constraints, but also constituent length motivated the upcoming boundaries. And since both phrases would be equally lengthened, both boundaries should be more highly (and possibly equally) motivated, and thus more highly accepted. It could also be that in the lack of clear disambiguating effect, this manipulation may maintain the structural ambiguity until the disambiguating region.

Regarding the ERP study, since we used 16 conditions in order to evaluate the CPS at both boundary sites at each prosodic level (4 × 4), fewer trials were used in each condition. As a result, it was necessary to collapse conditions together to allow a good signal to noise ratio. As a result, we could not examine the influence of both boundaries on comprehension at the same time, as we did in the behavioral studies. For the same reason, it was impossible to compare the predictions of the AAH and IBH. Finally, this also influenced the observed magnitude of the garden-path components, which were smaller than expected. Nevertheless, the fact that these effects were significant, despite the limitations of collapsing data across conditions, indicated that fewer conditions with a larger number of trials would most likely reveal much larger effects. Since this is the first ERP study in this field of research, this first step was necessary to chart the waters, so to speak, and enable these insights for future studies.

4.5. Conclusion

The current dissertation provides initial evidence for the effects of multiple prosodic boundaries of various sizes on the interpretation of English EC/LC garden-path sentences, both behaviorally and electrophysiologically. Prosodic boundaries exhibited gradient effects on acceptability. However, only the incompatible boundaries (e.g., early boundary in LC) drove acceptability in each structure. Moreover, boundaries exhibited a primacy effect, whereby the early boundary had a stronger impact on interpretation. The findings suggest that prosodic boundaries gradually lower the acceptability of the competing structure, with the early boundary exhibiting a stronger influence than the late boundary. As a result, acceptability for EC was generally higher, because the early boundary lowered the acceptability for LC to a greater extent than the late boundary lowered EC acceptability. ERP data showed that subtle differences between prosodic boundaries are detected by the brain and affect the degree of processing difficulty. The amplitudes of both CPS and garden-path components were influenced by the boundary size in a largely gradient manner. Mirroring the offline data, the garden-path effect in the less preferred structure, LC (P600), was also larger than the garden-path effect in EC (N400). These outcomes cannot be explained by a purely categorical account. Moreover, they cast serious doubts on most current models of prosodic online processing and strongly support the extended BDH.

REFERENCES

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of experimental psychology: learning, memory, and cognition, 22*(3), 714-735.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30*(6), 644-663.

Beckman, M. E., & Elam, G. A. (1997). Guidelines for ToBI labelling, version 3. The Ohio State University Research Foundation. *World Wide Web address: http://www. ling. ohio-state. edU/phonetics/ToBI/ToBI. 0. html*.

Beckman, M. E., & Hirschberg, J. (1994). The ToBI annotation conventions. *Ohio State University*.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. *Prosodic typology: The phonology of intonation and phrasing*, 9-54.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology yearbook, 3*(1), 5-70.

Blodgett, A. R. (2004). *The interaction of prosodic phrasing, verb bias, and plausability during spoken sentence comprehension.* (Doctoral Dissertation), The Ohio State University.

Boersma, P., & Weenink, D. (1996). PRAAT, a system for doing phonetics by computer. report 132. institute of Phonetic Sciences of the University of Amsterdam.

Bögels, S., Schriefers, H., Vonk, W., & Chwilla, D. J. (2011). Prosodic Breaks in Sentence Processing Investigated by Event-Related Potentials. *Language and Linguistics Compass, 5*(7), 424-440.

Bögels, S., Schriefers, H., Vonk, W., & Chwilla, D. J. (2011). The role of prosodic breaks and pitch accents in grouping words during on-line sentence processing. *Journal of cognitive neuroscience, 23*(9), 2447-2467.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2010). The interplay between prosody and syntax in sentence processing: The case of subject-and object-control verbs. *Journal of cognitive neuroscience, 22*(5), 1036-1053.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2013). Processing consequences of superfluous and missing prosodic breaks in auditory sentence comprehension. *Neuropsychologia, 51*(13), 2715-2728.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological review, 113*(4), 787.

Carlson, K. (2009). How prosody influences sentence comprehension. *Language and Linguistics Compass, 3*(5), 1188-1200.

Carlson, K., Clifton, C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language, 45*(1), 58-81.

Carlson, K., Clifton, C., & Frazier, L. (2009). Nonlocal effects of prosodic boundaries. *Memory & cognition, 37*(7), 1014-1025.

Carlson, K., Frazier, L., & Clifton, C. (2009). How prosody constrains comprehension: A limited effect of prosodic packaging. *Lingua, 119*(7), 1066-1082.

Carroll, P. J., & Slowiaczek, M. L. (1987). Modes and modules: Multiple pathways to the

language processor. In J. Garfield (Ed.), *Modularity in Knowledge Representation and Natural Language Understanding* (pp. 221-247). NewYork: Academic Press.
Clifton, C., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and speech, 45*(2), 87-114.
Clifton, C., Carlson, K., & Frazier, L. (2006). Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic bulletin & review, 13*(5), 854-861.
Cooper, W. E., Paccia, J. M., & Lapointe, S. G. (1978). Hierarchical coding in speech timing. *Cognitive Psychology, 10*(2), 154-177.
Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press.
Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech, 40*(2), 141-201.
de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America, 96*, 2037-2047.
Dilley, L. C., Breen, M., Bolivar, M., Kraemer, J., & Gibson, E. (2006). *A comparison of inter-transcriber reliability for two systems of prosodic annotation: rap (rhythm and pitch) and toBI (tones and break indices).* Paper presented at the INTERSPEECH.
Dilley, L. C., & Brown, M. (2005). The RaP (Rhythm and Pitch) Labeling System. *Unpublished manuscript*.
Donchin, E. (1981). Surprise!… surprise? *Psychophysiology, 18*(5), 493-513.
Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 547-619.
Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*(2), 210-233.
Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological review, 100*(2), 233-253.
Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11-15.
Ferreira, F., & Patson, N. D. (2007). The 'good enough'approach to language comprehension. *Language and Linguistics Compass, 1*(1-2), 71-83.
Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27*(2), 285-319.
Fodor, J. D. (2002). *Psycholinguistics cannot escape prosody.* Paper presented at the Speech Prosody 2002, International Conference.
Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and Performance XII* (pp. 559-586). Hillsdale, NJ: Lawrence Erlbaum Associates.
Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*(6), 244-249.
Frazier, L., Carminati, M. N., Cook, A. E., Majewski, H., & Rayner, K. (2006). Semantic evaluation of syntactic structure: Evidence from eye movements. *Cognition, 99*(2), B53-B62.
Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.
Frazier, L., & Clifton, C. (1997). Construal: Overview, motivation, and some new evidence. *Journal of Psycholinguistic Research, 26*(3), 277-295.
Frazier, L., Clifton, C., & Carlson, K. (2004). Don't break, or do: prosodic boundary preferences. *Lingua, 114*(1), 3-27.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*(4), 291-325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178-210.

Friederici, A. D., & Frisch, S. (2000). Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language, 43*(3), 476-507.

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: a spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research, 11*(2), 305-323.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*(3), 183-192.
Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*(3), 183-192.
Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax–semantic interface. *Brain research, 1146*, 50-58.

Frisch, S., Hahne, A., & Friederici, A. D. (2004). Word category and verb–argument structure information in the dynamics of parsing. *Cognition, 91*(3), 191-219.

Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *NeuroReport, 12*(15), 3391-3394.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*(1), 58-93.

Gorrell, P. (1995). *Syntax and parsing*: Cambridge University Press Cambridge.

Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics, 21*(3), 501-530.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38*(11), 1531-1549.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of cognitive neuroscience, 11*(2), 194-205.

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13*(3), 339-356.

Hasting, A. S., & Kotz, S. A. (2008). Speeding up syntax: On the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. *Journal of cognitive neuroscience, 20*(7), 1207-1219.

Hayes, B. (1989). The prosodic hierarchy in meter. *Phonetics and phonology, 1*, 201-260.

Hirst, D. J. (2005). Form and function in the representation of speech prosody. *Speech Communication, 46*(3), 334-347.

Hwang, H., & Schafer, A. J. (2006). Prosodic effects in parsing early vs. late closure sentences by second language learners and native speakers. *Proceedings of the SP*.

Hwang, H., & Steinhauer, K. (2011). Phrase Length Matters: The Interplay between Implicit

Prosody and Syntax in Korean "Garden Path" Sentences. *Journal of cognitive neuroscience, 23*(11), 3555-3575.
Isel, F., Alter, K., & Friederici, A. D. (2005). Influence of prosodic information on the processing of split particles: ERP evidence from spoken German. *Journal of cognitive neuroscience, 17*(1), 154-167.
Itzhak, I., Pauker, E., Drury, J. E., Baum, S. R., & Steinhauer, K. (2010). Event-related potentials show online influence of lexical biases on prosodic processing. *NeuroReport, 21*(1), 8-13.
Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology, 23*(4), 367-384.
Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D. J. (2007). Discourse, syntax, and prosody: The brain reveals an immediate interaction. *Journal of cognitive neuroscience, 19*(9), 1421-1434.
Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D. J. (2008). Sentence processing in the visual and auditory modality: Do comma and prosodic break have parallel functions? *Brain research, 1224*, 102-118.
Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40*(2), 153-194.
Kjelgaard, M. M., Titone, D. A., & Wingfield, A. (1999). The influence of prosodic structure on the interpretation of temporary syntactic ambiguity by young and elderly listeners. *Experimental Aging Research, 25*(3), 187-207.
Klatt, D. H. (1975). Vowel Lengthening is Syntactically Determined in a Connected Discourse. *Journal of phonetics, 3*(3), 129-140.
Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*, 194-231.
Kutas, M., & Federmeier, D. K. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology, 62*, 621-647.
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.
Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161-163.
Ladd, D. R. (1988). Declination ''reset''and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America, 84*, 530.
Ladd, R. D. (1986). Intonational phrasing: The case for recursive prosodic structure *Phonology yearbook* (Vol. 3, pp. 311-340).
Ladd, R. D. (2008). *Intonational phonology*: Cambridge University Press.
Lapata, M., Keller, F., Schulte, I. M., & Walde, S. (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research, 30*(4), 419-435.
Lee, E. K., & Watson, D. (Unpublished manuscript). Incremental Interpretation of Intonational Boundaries.
Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *The Journal of the Acoustical Society of America, 53*, 380.
Lehiste, I., Olive, J. P., & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *The Journal of the Acoustical Society of America, 60*, 1199-1202.
Li, W. J., & Yang, Y. F. (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience, 158*(4), 1416-1425.

Liberman, M. Y. (1975). *The intonational system of English.* (PhD Thesis), Massachusetts Institute of Technology.
Liberman, M. Y., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8*(2), 249-336.
Männel, C., & Friederici, A. D. (2009). Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. *Journal of cognitive neuroscience, 21*(10), 1988-2006.
Marcus, M., & Hindle, D. (1990). Description theory and intonation boundaries. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 483-512). Cambridge, MA: MIT Press.
Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of experimental psychology, 45*(1), 73-87.
Mietz, A., Toepel, U., Ischebeck, A., & Alter, K. (2008). Inadequate and infrequent are not alike: ERPs to deviant prosodic patterns in spoken sentence comprehension. *Brain and language, 104*(2), 159-169.
Millotte, S., René, A., Wales, R., & Christophe, A. (2008). Phonological phrase boundaries constrain the online syntactic analysis of spoken sentences. *Journal of experimental psychology: learning, memory, and cognition, 34*(4), 874.
Millotte, S., Wales, R., & Christophe, A. (2007). Phrasal prosody disambiguates syntax. *Language and Cognitive Processes, 22*(6), 898-909.
Mixdorff, H. (2002). *Speech technology, ToBI, and making sense of prosody.* Paper presented at the Speech Prosody 2002, International Conference.
Mueller, J. L., Hahne, A., Fujii, Y., & Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of cognitive neuroscience, 17*(8), 1229-1244.
Nagel, H. N., Shapiro, L. P., Tuller, B., & Nawy, R. (1996). Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. *Journal of Psycholinguistic Research, 25*(2), 319-344.
Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.
Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.
Ostendorf, M. (2000). Prosodic boundary detection *Prosody: Theory and Experiment* (pp. 263-279): Springer.
Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785-806.
Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes, 8*(4), 413-437.
Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of experimental psychology: learning, memory, and cognition, 20*(4), 786.
Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of cognitive neuroscience, 22*(12), 2728-2744.
Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven

sentence processing: An event-related brain potential study. *Journal of cognitive neuroscience, 17*(3), 407-421.
Pauker, E., Itzhak, I., Baum, S. R., & Steinhauer, K. (2011). Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the Boundary Deletion Hypothesis. *Journal of cognitive neuroscience, 23*(10), 2731-2751.
Pauker, E., Wagner, M., & Steinhauer, K. (In preparation). Rethinking the role of prosodic boundaries in ambiguous sentence resolution.
Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language, 43*(3), 447-475.
Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* (Doctoral Dissertation), Massachusetts Institute of Technology.
Pitrelli, J. F., Beckman, M. E., & Hirschberg, J. (1994). *Evaluation of prosodic transcription labeling reliability in the tobi framework.* Paper presented at the ICSLP.
Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America, 90*(6), 2956-2970.
Primus, B. (1999). *Cases and thematic roles: Ergative, accusative and active* (Vol. 393): Walter de Gruyter.
Pynte, J. (2006). Phrasing effects in comprehending PP constructions. *Journal of Psycholinguistic Research, 35*(3), 245-265.
Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes, 11*(1-2), 165-192.
Rösler, F., Pütz, P., Friederici, A. D., & Hahne, A. (1993). Event-related brain potentials while encountering semantic and syntactic constraint violations. *Journal of cognitive neuroscience, 5*(3), 345-362.
Rossi, M. (1997). *Is syntactic structure prosodically retrievable.* Paper presented at the Proceedings of 5th European Conference of Speech Communication and Technology.
Sanderman, A. A., & Collier, R. (1996). Prosodic rules for the implementation of phrase boundaries in synthetic speech. *The Journal of the Acoustical Society of America, 100*, 3390-3397.
Sanderman, A. A., & Collier, R. (1997). Prosodic phrasing and comprehension. *Language and speech, 40*(4), 391-409.
Schafer, A. J. (1997). *Prosodic parsing: The role of prosody in sentence comprehension.* University of Massachusetts Amherst.
Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29*(2), 169-182.
Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. *Approaches to studying world situated language use: Psycholinguistic, linguistic and computational perspectives on bridging the product and action tradition*, 209-225.
Schlesewsky, M., & Bornkessel, I. (2004). On incremental interpretation: degrees of meaning accessed during sentence comprehension. *Lingua, 114*(9), 1213-1234.
Selkirk, E. O. (1980). Prosodic domains in phonology: Sanskrit revisited. In M. Aronoff & M. L. Kean (Eds.), *Juncture* (Vol. 7, pp. 107-129). Saratoga, CA: Anma Libri.

Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry, 11*(3), 563-605.
Selkirk, E. O. (1984). *Phonology and Syntax: The Relation between Sound and Structure* Cambridge, MA: MIT Press.
Selkirk, E. O. (1986). On derived domains in sentence phonology *Phonology yearbook* (Vol. 3, pp. 371-405).
Selkirk, E. O. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. Goldsmith (Ed.), *Handbook of Phonological Theory* (pp. 550-569). New-York: Blackwell.
Selkirk, E. O. (2000). The interaction of constraints on prosodic phrasing *Prosody: Theory and experiment* (pp. 231-262). Dordrecht, the Netherlands: Kluwer.
Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193-247.
Silverman, K. E. A., Beckman, M. E., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P. J., . . . Hirschberg, J. (1992). *TOBI: a standard for labeling English prosody.* Paper presented at the ICSLP.
Snedeker, J., & Casserly, E. (2010). Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes, 25*(7-9), 1234-1264.
Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*(1), 103-130.
Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language, 58*(2), 574-608.
Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research, 25*(2), 249-271.
Speer, S. R., Warren, P., & Schafer, A. J. (2003). *Intonation and sentence processing.* Paper presented at the Proceedings of the Fifteenth International Congress of Phonetic Sciences, Barcelona, Spain.
Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and language, 86*(1), 142-164.
Steinhauer, K., Abada, S. H., Pauker, E., Itzhak, I., & Baum, S. R. (2010). Prosody–syntax interactions in aging: Event-related potentials reveal dissociations between on-line and off-line measures. *Neuroscience letters, 472*(2), 133-138.
Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature neuroscience, 2*(2), 191-196.
Steinhauer, K., & Connolly, J. F. (2008). Event-related potentials in the study of language. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of the neuroscience of language* (pp. 91-104). New York: Elsevier.
Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and language, 120*(2), 135-162.
Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research, 30*(3), 267-295.
Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *The Journal of the Acoustical Society of America, 64*, 1582-1592.
Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *The Journal of*

the Acoustical Society of America, 101, 514-521.
Syrdal, A. K., Hirschberg, J., McGory, J. T., & Beckman, M. E. (2001). Automatic ToBI prediction and alignment to speed manual labeling of prosody. *Speech Communication, 33*(1), 135-151.
Syrdal, A. K., & McGory, J. T. (2000). *Inter-transcriber reliability of toBI prosodic labeling.* Paper presented at the INTERSPEECH.
Toepel, U., Pannekamp, A., & Alter, K. (2007). Catching the news: Processing strategies in listening to dialogs as measured by ERPs. *Behavioral and Brain Functions, 3*(1), 53-66.
Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry, 30*(2), 219-255.
Van Valin, R. D., & LaPolla, R. J. (1997). *Syntax: Structure, meaning, and function*: Cambridge University Press.
Wagner, M. (2005). *Prosody and recursion.* Massachusetts Institute of Technology.
Wagner, M., & Crivellaro, S. (2010). *Relative prosodic boundary strength and prior bias in disambiguation.* Paper presented at the Speech Prosody.
Wagner, M., & Watson, D. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*(7-9), 905-945.
Walker, J. P., Fongemie, K., & Daigle, T. (2001). Prosodic facilitation in the resolution of syntactic ambiguities in subjects with left and right hemisphere damage. *Brain and language, 78*(2), 169-196.
Warren, P. (1985). *The temporal organization and perception of speech.* (Doctoral Dissertation), Cambridge University.
Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology and parsing in closure ambiguities. *Language and Cognitive Processes, 10*(5), 457-486.
Watson, D. (2002). *Intonational phrasing in language production and comprehension.* Massachusetts Institute of Technology.
Watson, D., Breen, M., & Gibson, E. (2006). The role of syntactic obligatoriness in the production of intonational boundaries. *Journal of experimental psychology: learning, memory, and cognition, 32*(5), 1045-1056.
Watson, D., & Gibson, E. (2004). Making sense of the sense unit condition. *Linguistic Inquiry, 35*(3), 508-517.
Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19*(6), 713-755.
Watson, D., & Gibson, E. (2005). Intonational phrasing and constituency in language production and comprehension*. *Studia Linguistica, 59*(2-3), 279-300.
Wightman, C. W. (2002). *ToBI or not ToBI?* Paper presented at the Speech Prosody 2002, International Conference.
Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America, 91*, 1707-1717.
Ye, Z., Luo, Y., Friederici, A. D., & Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials. *Brain research, 1071*(1), 186-196.

APPENDIX 1

The list of stimuli used in Study 1 and Study 2.

Item	Syntax	Sentence
1	EC	Whenever the woman was browsing # the journal # would fall off the table
	LC	Whenever the woman was browsing # the journal # the game would fall off the table
2	EC	Whenever the dog was chasing # the cats # would run away
	LC	Whenever the dog was chasing # the cats # the mice would run away
3	EC	Whenever the player was batting # the ball # would reach the street
	LC	Whenever the player was batting # the ball # the cheers would reach the street
4	EC	Whenever the man was parking # the cars # would wait
	LC	Whenever the man was parking # the cars # the bikes would wait
5	EC	Whenever the puppy was licking # the baby # would laugh
	LC	Whenever the puppy was licking # the baby # the nanny would laugh
6	EC	Whenever the snakes were eating # the rats # would hide
	LC	Whenever the snakes were eating # the rats # the frogs would hide
7	EC	Whenever the bear was approaching # the people # would run away
	LC	Whenever the bear was approaching # the people # the dogs would run away
8	EC	Whenever the girl was baking # the cake # would smell good
	LC	Whenever the girl was baking # the cake # the house would smell good
9	EC	Whenever the artist was drawing # the kids # would smile
	LC	Whenever the artist was drawing # the kids # the parents would smile
10	EC	Whenever the cat was climbing # the tree # would shake
	LC	Whenever the cat was climbing # the tree # the leaves would shake
11	EC	Whenever the student was studying # the notes # would become clearer
	LC	Whenever the student was studying # the notes # the lesson would become clearer
12	EC	Whenever the professor was teaching # the students # would come late
	LC	Whenever the professor was teaching # the students # the TA would come late
13	EC	Whenever the actress was performing # the play # would amuse the audience
	LC	Whenever the actress was performing # the play # the costumes would amuse the audience
14	EC	Whenever the star was singing # the jingle # would become a hit
	LC	Whenever the star was singing # the jingle # the product would become a hit
15	EC	Whenever the man was swimming # the channel # would feel cold
	LC	Whenever the man was swimming # the channel # the air would feel cold
16	EC	Whenever the man was leaving # the club # would close early
	LC	Whenever the man was leaving # the club # the bar would close early
17	EC	Whenever the clown was hosting # the show # would attract many people
	LC	Whenever the clown was hosting # the show # the music would attract many people
18	EC	Whenever the patient was phoning # the doctor # would put him on hold
	LC	Whenever the patient was phoning # the doctor # the nurse would put him on hold
19	EC	Whenever the enemy was striking # the army # would retaliate
	LC	Whenever the enemy was striking # the army # the navy would retaliate
20	EC	Whenever the driver was starting # the race # would captivate the crowd
	LC	Whenever the driver was starting # the race # the action would captivate the crowd
21	EC	Whenever the critic was buying # the meal # would taste very good
	LC	Whenever the critic was buying # the meal # the wine would taste very good
22	EC	Whenever the carpenter was checking # the door # would break
	LC	Whenever the carpenter was checking # the door # the lock would break
23	EC	Whenever the man was programming # the website # would crash
	LC	Whenever the man was programming # the website # the software would crash

24	EC	Whenever the man was watching # the sky # would inspire him
	LC	Whenever the man was watching # the sky # the stars would inspire him
25	EC	Whenever the woman was skiing # the slope # would amaze her
	LC	Whenever the woman was skiing # the slope # the view would amaze her
26	EC	Whenever the neighbor was renovating # the house # would get messy
	LC	Whenever the neighbor was renovating # the house # the driveway would get messy
27	EC	Whenever the mouse was sniffing # the cheese # would fall on the floor
	LC	Whenever the mouse was sniffing # the cheese # the plate would fall on the floor
28	EC	Whenever the man was driving # the car # would stop working
	LC	Whenever the man was driving # the car # the radio would stop working
29	EC	Whenever the mom was parking # the truck # would hit the curb
	LC	Whenever the mom was parking # the truck # the wheels would hit the curb
30	EC	Whenever the rider was stopping # the bike # would fall down
	LC	Whenever the rider was stopping # the bike # the basket would fall down
31	EC	Whenever the woman was calling # the boss # would pick up the phone
	LC	Whenever the woman was calling # the boss # the secretary would pick up the phone
32	EC	Whenever the couple was dancing # the Salsa # would look great
	LC	Whenever the couple was dancing # the Salsa # the outfits would look great
33	EC	Whenever the grocer was closing # the store # would need cleaning
	LC	Whenever the grocer was closing # the store # the floor would need cleaning
34	EC	Whenever the boy was scrubbing # the cups # would get soapy
	LC	Whenever the boy was scrubbing # the cups # the sink would get soapy
35	EC	Whenever the author was reading # the book # would come alive
	LC	Whenever the author was reading # the book # the tale would come alive
36	EC	Whenever the artist was painting # the portrait # would look peaceful
	LC	Whenever the artist was painting # the portrait # the colors would look peaceful
37	EC	Whenever the boy was swinging # the rope # would break
	LC	Whenever the boy was swinging # the rope # the branch would break
38	EC	Whenever the man was digging # the trench # would widen
	LC	Whenever the man was digging # the trench # the cracks would widen
39	EC	Whenever the parents were playing # the game # would seem simple
	LC	Whenever the parents were playing # the game # the rules would seem simple
40	EC	Whenever the coach was helping # the team # would feel hopeful
	LC	Whenever the coach was helping # the team # the crowd would feel hopeful
41	EC	Whenever the friends were playing # the game # would last for hours §
	LC	Whenever the friends were playing # the game # the food would last for hours §
42	EC	Whenever the man was clicking # the button # would freeze §
	LC	Whenever the man was clicking # the button # the remote would freeze §
43	EC	Whenever the boy was playing # the game # would stop working †§
	LC	Whenever the boy was playing # the game # the joystick would stop working †§
44	EC	Whenever the boy was visiting # the dentist # would give him candy †§
	LC	Whenever the boy was visiting # the dentist # the hygienist would give him candy †§
45	EC	Whenever the girl was humming # the tune # would sound nice †§
	LC	Whenever the girl was humming # the tune # the melody would sound nice †§

Notes: (1) # marks the location of the prosodic boundaries in the sentence.
(2) † = Sentence used in the practice block in Experiment 1 (omitted from the experimental list).
(3) § = Sentence used in the practice block in Experiments 2 and 3 (omitted from the experimental list).

APPENDIX 2

Table 1. Average number of trials for each condition (EC+LC) across experimental versions as a function of boundary position

		Boundary 2				
		1	*2*	*3*	*4*	***Total***
Boundary 1	*4*	16	16	16	16	**64**
	3	18	18	18	18	**72**
	2	18	18	18	30	**84**
	1	20	20	30	30	**100**
	Total	**72**	**72**	**82**	**94**	**320**

Table 2. Number of trials for each structure in each condition of experimental version 1 as a function of boundary position

		Boundary 2				
		1	*2*	*3*	*4*	***Total***
Boundary 1	*4*	8 EC 8 LC	8 EC 8 LC	8 EC 8 LC	8 EC 8 LC	**64**
	3	9 EC 9 LC	8 EC 9 LC	9 EC 9 LC	10 EC 9 LC	**72**
	2	9 EC 9 LC	8 EC 9 LC	10 EC 7 LC	16 EC 16 LC	**84**
	1	10 EC 10 LC	10 EC 10 LC	15 EC 15 LC	15 EC 15 LC	**100**
	Total	**72**	**70**	**81**	**97**	**320**

Table 3. Number of trials for each structure in each condition of experimental version 2 as a function of boundary position

		Boundary 2				
		1	*2*	*3*	*4*	***Total***
Boundary 1	*4*	8 EC 8 LC	8 EC 8 LC	8 EC 7 LC	8 EC 9 LC	**64**
	3	9 EC 9 LC	9 EC 9 LC	9 EC 10 LC	9 EC 8 LC	**72**
	2	9 EC 9 LC	9 EC 9 LC	9 EC 9 LC	15 EC 15 LC	**84**
	1	10 EC 10 LC	10 EC 10 LC	15 EC 15 LC	15 EC 15 LC	**100**
	Total	**72**	**72**	**82**	**94**	**320**

Table 4. Number of trials for each structure in each condition of experimental version 3 as a function of boundary position

		Boundary 2				
		1	*2*	*3*	*4*	***Total***
Boundary 1	*4*	8 EC 8 LC	8 EC 8 LC	8 EC 9 LC	8 EC 7 LC	**64**
	3	9 EC 9 LC	10 EC 9 LC	9 EC 8 LC	8 EC 10 LC	**72**
	2	10 EC 8 LC	10 EC 9 LC	9 EC 9 LC	14 EC 15 LC	**84**
	1	10 EC 10 LC	10 EC 10 LC	15 EC 16 LC	15 EC 14 LC	**100**
	Total	**72**	**74**	**83**	**91**	**320**

Table 5. Number of trials for each structure in each condition of experimental version 4 as a function of boundary position

		Boundary 2				
		1	*2*	*3*	*4*	***Total***
Boundary 1	*4*	8 EC 8 LC	8 EC 8 LC	8 EC 8 LC	8 EC 8 LC	**64**
	3	9 EC 9 LC	9 EC 9 LC	9 EC 9 LC	9 EC 9 LC	**72**
	2	9 EC 9 LC	9 EC 9 LC	9 EC 10 LC	15 EC 14 LC	**84**
	1	10 EC 10 LC	10 EC 10 LC	15 EC 14 LC	15 EC 16 LC	**100**
	Total	**72**	**72**	**82**	**94**	**320**