Journal of Phonetics 75 (2019) 113–147

Contents lists available at ScienceDirect

Journal of Phonetics

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

Special Issue: Integrating Phonetics and Phonology, eds. Cangemi & Baumann

Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish

Jennifer Cole [a,*], José I. Hualde [b], Caroline L. Smith [c], Christopher Eager [b], Timothy Mahrt [b], Ricardo Napoleão de Souza [c]

[a] Northwestern University, Evanston, IL 60201, USA
[b] University of Illinois, Urbana, IL 61801, USA
[c] University of New Mexico, Albuquerque, NM 87131-0001, USA

ARTICLE INFO

Article history:
Received 4 April 2018
Received in revised form 2 May 2019
Accepted 6 May 2019
Available online 22 June 2019

Keywords:
Prominence
Intonation
Prosody
Pitch accent
Stress
Prosody perception
Prosodic annotation

ABSTRACT

This study tests the influence of acoustic cues and non-acoustic contextual factors on listeners' perception of prominence in three languages whose prominence systems differ in the phonological patterning of prominence and in the association of prominence with information structure—English, French and Spanish. Native speakers of each language performed an auditory rating task to mark prominent words in samples of conversational speech under two instructions: with prominence defined in terms of acoustic or meaning-related criteria. Logistic regression models tested the role of task instruction, acoustic cues and non-acoustic contextual factors in predicting binary prominence ratings of individual listeners. In all three languages we find similar effects of prosodic phrase structure and acoustic cues (F0, intensity, phone-rate) on prominence ratings, and differences in the effect of word frequency and instruction. In English, where phrasal prominence is used to convey meaning related to information structure, acoustic and meaning criteria converge on very similar prominence ratings. In French and Spanish, where prominence plays a lesser role in signaling information structure, phrasal prominence is perceived more narrowly on structural and acoustic grounds. Prominence ratings from untrained listeners correspond with ToBI pitch accent labels for each language. Distinctions in ToBI pitch accent status (nuclear, prenuclear, unaccented) are reflected in empirical and model-predicted prominence ratings. In addition, words with a ToBI pitch accent type that is typically associated with contrastive focus are more likely to be rated as prominent in Spanish and English, but no such effect is found for French. These findings are discussed in relation to probabilistic models of prominence production and perception.

1. Introduction

An important development in 20th-century phonology is the idea that phrasal prominence is assigned to the word that is the structural head (nucleus) of the prosodic phrase (Liberman & Prince, 1977; Liberman, 1975; Selkirk, 1984). This structural prominence can be phonetically marked in various ways. In some languages the prominent word is acoustically enhanced compared to other words in the same prosodic phrase, as measured by one or more acoustic parameters (e.g., duration, energy, F0), or by acoustic measures of hyper-articulation in place or manner features (see Wagner & Watson, 2010; Cole, 2015, and references therein; see also Oh & Byrd, in press). Although all languages appear to have prosodic phrasing (Büring, 2016; Ladd, 2008; Nespor & Vogel, 1986; Selkirk, 1984, 1995), there are differences among languages in the specification of prominence within the prosodic phrase. Some differences concern the density of prominences within a prosodic phrase and the rhythmic patterning of prominence, while other differences arise from the association of prominence with information structure (the status of a word as contrastively focused, new, or given relative to the local discourse context), or the fixed vs. variable location of prominence within the prosodic phrase.

A consequence of these differences in prominence systems is that, across languages, similar acoustic expressions of prominence may serve differently to cue information related to the syntactic, pragmatic and semantic context of a word. Furthermore, F0 excursions and other acoustic features that

* Corresponding author.
E-mail address: jennifer.cole1@northwestern.edu (J. Cole).

https://doi.org/10.1016/j.wocn.2019.05.002
0095-4470/© 2019 The Authors. Published by Elsevier Ltd.

signal phrasal prominence in some languages may be used in other languages to express differences between lexical items, instead. Thus, whereas in English a tonally specified pitch accent can be used to mark prominence related to the discourse meaning of words (as described in Section 2.1), in languages such as Tokyo Japanese and northern Bizkaian Basque, whether a word bears an accent or not in a given syntactic environment is a lexical or morphological property, unrelated to discourse prominence; cf., e.g. Lekeitio Basque *lagunen alabiá* 'the friend's (sg) daughter' (phrase with one accent) vs. *lagúnen alabiá* 'the friends' (pl) daughter' (phrase with two accents) (Hualde, Elordieta, & Elordieta, 1994). Regarding the languages investigated in this paper, Spanish has lexical and grammatical contrasts in the position of the stressed syllable within the word, and word-level stress is phonologically marked via the assignment of pitch accent. Because Spanish lacks the systematic reduction of unstressed vowels and robust consonantal correlates of stress found in English, F0 and duration are important cues for the location of contrastive stress accent. This has implications for our understanding of phrasal prominence. In Spanish, an F0 excursion on a word in a phrasal context may signal nothing more than the location of word-level stress. By comparison, in English, there are robust segmental correlates of word-level stress and therefore ample acoustic cues to word-level stress other than F0, and so F0 is perhaps more available for the encoding of pragmatic meaning than in Spanish. French provides an informative contrast with English and Spanish with its very different treatment of prosodic prominence. At the word level, it lacks lexical stress. At phrase level, prominent syllables are usually in final position in the phrase, so that an F0 excursion signals information about both prominence and phrasal structure.

There may also be differences in the way that non-acoustic factors related to the syntactic, semantic or discourse context of a word trigger listeners' expectations about the prominence status of a word, reflecting the association of prominence with sentence structure, discourse structure and meaning. Expectation-driven processing in prominence perception has been demonstrated in prior work, where factors such as part of speech, information status (new vs. given relative to the discourse context), and word frequency are shown to influence listeners' prominence judgments, independent of acoustic cues present in the signal (Bishop, 2012, 2016; Bishop, Kuo, & Kim, in press; Cole, Mo, & Hasegawa-Johnson, 2010).

The aim of this paper is to understand how phrasal prominence is perceived in relation to acoustic cues and to non-acoustic contextual factors related to the phonological and syntactic properties of a word and to word frequency, all of which may play a role in expectation-driven prominence processing. The broader goal is to understand if there are common factors underlying perceived prominence in languages that differ in the phonological patterning of prominence, or in its syntactic, pragmatic or semantic functions. We investigate three languages—English, French and Spanish—whose prominence systems are reported to differ along these dimensions of form and function. The study presented here compares naïve listeners' word-level ratings of phrasal prominence in each language, under two task conditions: with prominence defined in terms of explicit acoustic criteria, and in terms of speaker-intended meaning. We examine (i) the degree to which prominence ratings can be predicted from acoustic cues and non-acoustic contextual factors in each language, and (ii) the degree to which prominence ratings obtained under acoustic criteria are different from those obtained under meaning-related criteria. We test hypotheses about differences among the languages in the effects of individual acoustic cues (F0, duration, intensity) on prominence ratings obtained under acoustic criteria. We also test hypotheses about differences in the effects of word-level contextual factors on prominence ratings obtained under meaning criteria, including part of speech effects, effects from the local prosodic phrase context, and effects of word frequency. For each language, prominence ratings from naïve listeners are also compared with the pitch accent labels assigned by trained annotators using the Tones and Break Indices system (ToBI) for prosodic annotation (Beckman, Hirschberg, & Shattuck-Hufnagel, 2005), to test whether the presence or type of ToBI pitch accent predicts prominence ratings, and whether there is a difference in prominence ratings for words with nuclear vs. prenuclear prominence.

Section 2 proceeds with a discussion of the properties that differentiate the prominence systems of the three languages investigated here. This rather lengthy section lays out the properties that we consider likely to influence prominence perception, and formulates the specific hypotheses tested in this study. Section 3 introduces the prominence rating study that is the focus of this paper, presenting the experimental task, speech corpus materials, and the statistical methods used. Section 4 presents the results, followed by discussion and conclusion in Sections 5 and 6.

2. Parameters of variation in prominence systems

Previous research reveals similarities and differences in the prominence systems of English, French and Spanish. The differences concern the status of pitch as a correlate of prominence, the location of nuclear prominence and its association with information structure, and the density of prominences within the prosodic phrase. Section 2.1 reviews the distributional patterns of prominence in each language and the role of information structure in prominence assignment. Section 2.2 describes acoustic correlates of prominence in each language, as potential signal-based influences on prominence rating. Seciton 2.3 discusses the effect of word frequency on prominence as a potential expectation-driven influence on prominence rating. The specific hypotheses for the present study are laid out in Section 2.4.

2.1. The distribution of prominence and its relationship to information structure

Common to all three of the languages studied here is the assignment of prominence to the head element of a phrase-level prosodic domain. This structural prominence, termed *nuclear prominence* here, is obligatory and ensures a minimum of one prominence in each prosodic phrase. Additional *prenuclear prominences* are sometimes observed preceding the nuclear prominence in the same prosodic phrase, in each language. English, French and Spanish differ in the distributional properties of nuclear and prenuclear prominences, and in the association of nuclear prominence with contrastive focus and other information status (givenness) distinctions. Below we review the distribution of nuclear and prenuclear prominences in each language.

In Jun's (2014) proposal for a prosodic typology, both English and Spanish are classified as head-prominence stress languages, since word-prominence, manifested as stress in both languages, is assigned to prosodic heads. In both of these languages, the word-level stress prominence licenses the assignment of a pitch accent (Bolinger, 1958)—a tone feature assigned to the stressed syllable of a word that also has phrase-level stress (i.e., that is in a strong position in the phrase-level metrical structure), which is phonetically implemented through a pitch movement (rise or fall) in the vicinity of the word-level stressed syllable. The two languages, however, are said to differ in their macro-rhythm, with Spanish being more macro-rhythmic than English. Jun defines macro-rhythm essentially as a regular alternation of high and low tones in the phrase. The reason why Spanish is taken to be a language with strong macro-rhythm is that it has a high density of pitch-accent marking, with pitch accent serving as a very reliable cue for the location of the lexically stressed syllable (e.g. /kánto/ 'I sing' vs /kantó/ 'she or he sang'). In addition, the most common pitch accent in Spanish is a rising contour with a protracted peak, resulting in a recurring rise-fall pattern on successive word-level stressed syllables across a phrase. English, instead, is classified as a language with medium macro-rhythm, because pitch accents occur comparatively less frequently (a word-level stress is only probabilistically assigned a pitch accent) and its most common pitch accent is a simple high tone, which has a variable phonetic realization with sustained non-low or gently rising pitch.

French, on the other hand, is a head/edge prominence language in Jun's typology, because prominence, which is assigned to word-final syllables, is marked both by a pitch accent associated with the prosodic head and by a boundary tone. Within the head/edge prominence group, French is classified as a language with a strong macro-rhythm. These differences in the association of pitch accent and prominence are further elaborated below.

English. In English (Chafe, 1987; Pierrehumbert, 1980; Selkirk, 1995), the obligatory, nuclear prominence is located by default on the rightmost (content) word in the prosodic phrase (1). Nuclear prominence will be assigned to a word in an earlier position in the phrase if that word is focused, i.e., interpreted relative to a contextually determined set of alternatives (i.e., contrastive focus, see Rooth, 1992; Krifka, 2007; Wagner & McAuliffe, in press), as in (2A), if the phrase-final word is lexically given (3a) or referentially given (3b) (examples adapted from Riester & Baumann, 2017), or under certain syntactic conditions, e.g., with unaccusative verbs (4a) and with indefinites (4b). Speakers may distinguish a prominence that marks contrastive or narrow focus from a "default" prominence used in a broad focus context through scaling and alignment of the pitch contour (Breen, Fedorenko, Wagner, & Gibson, 2010). Such differences are analyzed by some authors as differences in the tonal specification of the pitch accent, with L+H* being the preferred pitch accent for words with contrastive focus (Beckman & Pierrehumbert, 1986; Pierrehumbert & Hirschberg, 1990; Pierrehumbert, 1980), though other authors dispute the status of a categorically distinct, contrast-marking L+H* pitch accent (Bartels & Kingston, 1994; Calhoun, 2006; Ladd & Schepman, 2003; Watson, Tanenhaus, & Gunlogson, 2008). In the following examples, as elsewhere in this paper, prominence is indicated with boldface.

(1) Q: How was the story leaked?
A: Terry phoned the **reporter**.
(2) Q: Did Terry phone the reporter?
A: No, **Jonah** talked to him.
(3) a. Look at that funny dog! I **like** dogs/***dogs**.
b. Ole$_i$ was a brilliant athlete. The local press had nothing but **praise** for the skier$_i$/***skier**$_i$.
(4) a. The **phone** rang.
b. I **saw** something.

Prenuclear prominences also occur in English (Bolinger, 1986; Ladd, 2008), though they appear to be optional. For instance, a speaker may produce a sparse (5a) or dense (5b) pattern of prenuclear prominence for the same sentence. Prenuclear prominences exhibit a tendency towards rhythmic alternation (Calhoun, 2010; Vogel, Bunnell, & Hoskins, 1995) and early placement in the phrase (Shattuck-Hufnagel, 1995; Shattuck-Hufnagel, Ostendorf, & Ross, 1994). Words with prenuclear prominence can be assigned pitch accents of the same type that are assigned to a nuclear prominence, but the nuclear prominence is described as perceptually stronger than prenuclear prominences (Ladd, 2008, sec. 4.1). In the examples in (5), all prominent words are shown in boldface and the extra perceptual salience of the nuclear prominence on *taxes* in (5b) is marked by capital letters.

(5) a. Meg will help you file your **taxes**.
b. **Meg** will **help** you **file** your **TAXES**.

French. Phrasal prominence in French is related to pitch accent and the locations of phrasal boundaries, but these elements pattern quite differently in French than in English. Central to most accounts of French prosody is the notion of accent and its placement in the construction of prosodic phrases. The smallest domain for accent assignment is the Accentual Phrase (AP), a prosodic domain that is intermediate between the word and the intonational phrase, and consists most often of a single lexical word and preceding clitics (Jun & Fougeron, 2000, 2002; Post, 2000). There are typically multiple APs in an intonational phrase, one for each content word, though some authors note that AP parsing is subject to rhythmic "clash" constraints (Dell, 1984; Post, 1999; see also Delais-Roussarie & Rialland, 2007). Each AP is assigned a tone feature as an obligatory final accent on the last full-vowel syllable (6a) (e.g., Di Cristo, 1999; Jun & Fougeron, 2000; Delais-Roussarie et al., 2015), and an optional initial accent whose F0 peak is on the first or second syllable of the content word (Jun & Fougeron, 2002; Welby, 2006). Opinions differ as to whether one, both, or neither of these accents is a pitch accent; we will adopt the view that the final accent is a pitch accent, designated H*, and the initial accent, designated Hi, is not a pitch accent. This interpretation is implicit in the French ToBI annotation system (Delais-Roussarie et al., 2015), which uses these labels, and is argued for more explicitly in Welby (2006), among others.

Within the AP, the position of the accented syllable is fixed. The pitch rise that marks the final accent is described as creating the perception of prominence on that syllable, though these accents are not "prominence-lending" (Welby, 2006: 365), as in English or Spanish. This AP-final prominence is thus largely structure-marking, and does not relate to the word's status in information structure. It is generally considered as perceptually less salient than the nuclear prominence in languages like English (Vaissière, 1997). Prominence location is not used to mark the focus status of a word within the AP or larger prosodic phrase, as it is in English. Instead, contrastive or narrow focus is often achieved by syntactic means: either clefting or dislocation of the focused constituent are possible (Astésano, Espesser, & Rossi-Gensane, 2008; Doetjes, Rebuschi, & Rialland, 2004; Féry, 2013). Example (6a) shows a default sentence structure; (6b) focuses the phrase "le téléphone" by use of a cleft, and (6c) by dislocation, which can move a constituent to either the beginning or end of the sentence. Unlike English, in French lexically or referentially given expressions do not lose their AP-final accent, as shown in (7)

(6) a. [le télé**phone**]$_{AP}$ [a son**né**]$_{AP}$ *the telephone rang*
b. [c'est le télé**phone**]$_{AP}$ [qui a son**né**]$_{AP}$ *it's the telephone that rang.*
c. [le télé**phone**]$_{AP}$ [il a son**né**]$_{AP}$ or [il a son**né**]$_{AP}$ [le télé**phone**]$_{AP}$
the telephone, it rang. it rang, the telephone.
(7) [la chaise **jaune**] $_{AP}$ [et la banane **jaune**] $_{AP}$
the yellow chair and the yellow banana

The syntactic marking of contrastive or narrow focus may be accompanied by prosodic marking through post-focal compression or "deaccenting" (Jun & Fougeron, 2000; Delais-Roussarie & Rialland, 2007; Féry, 2013), although this is more limited than has been described for English (Di Cristo & Jankowski, 1999; Delais-Roussarie et al., 2002; Destruel & Féry, 2015). Two ways of marking contrastive or narrow focus through prosodic means alone, use of an AP-initial accent and the placement of nuclear accent on the right edge of the focused material (Beyssade, Hemforth, Marandin, & Portes, 2009), are probabilistic at best, being found in only slightly more than half the data that Beyssade et al. studied. German and D'Imperio (2016) argue similarly that the initial accent is not a reliable marker of focus, and does not confer prominence, although it may combine with other prominence markers.

In the absence of contrastive or narrow focus, a word may be prominent in an Intonational Phrase by virtue of having the rightmost AP-final accent, which is considered in many analyses to be the default position for the nuclear accent (e.g., Di Cristo, 1999; Delais-Roussarie et al., 2015). Pitch rises in nuclear position are higher with later targets than prenuclear pitch rises, and may be perceived as having slightly different shaped contours (D'Imperio, Bertrand, Di Cristo, & Portes, 2007). These results suggest that the pitch accent in nuclear position is perceived differently than AP-final pitch accents earlier in the utterance, although we are not aware of studies explicitly demonstrating greater perceived prominence of nuclear accents. This may be a promising line for future inquiry, as French listeners do seem to be capable of distinguishing different levels of prominence, as shown by Astésano, Bertrand, Espesser, and Nguyen (2012) who found that accents preceding Intermediate Phrase boundaries were perceived as stronger than accents that were final in an Accentual Phrase but medial in larger phrases.

Spanish. Spanish differs from French in having lexically contrastive stress, but it is similar to French is having a fixed location for nuclear prominence at the end of the prosodic domain (8a), which in Spanish is the prosodic phrase (Vallduví, 1991; Zubizarreta, 1998). Another similarity with French is the use of syntactic means to position contrastive or narrow focused words in phrase-final position (Bolinger, 1954). Spanish allows relatively free rightward displacement of focused words, as in (8b), though this is not always an available option for focus marking due to rigid constraints on the order of words within constituents. Spanish patterns with French, and is different from English, in maintaining accent on lexically or referentially given expressions (Hualde, 2005), as shown in (9).

(8) a. el **teléfono SONÓ** *the telephone rang*
b. **sonó** el **TELÉFONO** *the TELEPHONE rang*
(9) una **silla azul** y una **mesa AZUL**
a blue chair and a blue table

Spanish is described as having a denser distribution of pitch accents than English (Cruttenden, 1993, 2006; Ortiz-Lira, 1994), with prominence typically assigned to each content word in a sentence (as illustrated in the examples above). This is especially true for careful or formal speaking styles, while "deaccented" content words are noted in spontaneous speech (Face, 2003; Rao, 2009). Except in contexts of deaccenting, pitch accents are fairly reliable indicators of lexical stress in Spanish, and pitch is considered to be the primary correlate of word stress in Spanish, which lacks the systematic reduction of unstressed vowels found, for example, in English (Hualde, 2005; Ortega-Llebaria & Prieto, 2011, Nadeu, 2014). Spanish does not exhibit rhythmicity in the distribution of prenuclear prominence, admitting rhythmically irregular prenuclear accent sequences (10a, accented syllables in bold) as well as long sequences of unstressed syllables that arise in constructions with lexically unstressed function words (10b, accented syllable in bold).

(10) a. El **man**go ca**yó** de la estante**rí**a.
'the mango fell from the bookshelf'
b. pero cuando se los en**cuen**tra
'but when one finds them'

Although a prominent pitch accent may be used to convey emphatic stress, there is no consensus that all words bearing a pitch accent in an utterance have greater phrasal prominence than unaccented words in the same utterance. Physically enhancing the stressed syllable of polysyllabic content words may aid in word recognition (by signaling the location of contrastive word-level stress), resulting in high accentual density as compared to English and a weak connection between pitch-accent and discourse meaning. Instead, as already noted, word order can be used to express information structure and thus to a certain extent to convey discourse prominence (Bolinger, 1954; Olarrea, 2014). The last content word in a phrase, bearing nuclear accent, may show a pitch excursion, but in declarative sentences it often shows a falling contour lacking any clear pitch specification. In such cases the primary correlate of nuclear prominence is the lengthening of the lexically stressed syllable (Hualde & Prieto, 2015: 364). The relative perceptual salience of nuclear and prenuclear prominences in Spanish has not been investigated, to our knowledge.

This section has reviewed some of the more substantial and perceptually salient differences among English, French and Spanish in the distribution of prominence within the prosodic phrase and the relationship between prominence and information structure. With respect to information structure, the general picture is that whereas English uses pitch accent to mark contrastive focus, and avoids assigning nuclear accent to a discourse-given word, French and Spanish exhibit comparatively weaker effects of information structure on pitch accent assignment. Neither French nor Spanish have a pitch accent type that is specifically associated with contrastive or narrow focus. With respect to information status (givenness distinctions), Cruttenden (2006) directly compares the effect of information status on pitch accent assignment across languages and argues that while French and Spanish are among the languages most likely to exhibit prominence (“reaccenting” in Cruttenden’s terminology) on words that are discourse-given, English is on the opposite end of the scale as a language where prominence is strongly disfavored (“deaccenting”) on discourse-given words. Besides these differences, there are other differences in pitch accent assignment in English, French and Spanish that we do not consider in this paper. For example, in English, verum focus is expressed by pitch accent assigned to the auxiliary verb (e.g., she **did** buy the book), while in Spanish a common strategy for the same purpose involves the durational enhancement of the lexically-stressed syllable of the last word in the phrase, e.g. *compró el* ***li:bro*** (Escandell-Vidal, Marrero Aguiar, & Pérez Ocón, 2014).

2.2. Acoustic prominence

In both English and Spanish, words that trained annotators assign a prominence-lending pitch accent are likely to show acoustic enhancement of the syllable with primary lexical stress in terms of duration, intensity and the anchoring of an F0 contour. The role of F0 as a correlate of pitch accent is complicated by the fact that pitch-accented syllables may exhibit a number of different F0 contours, including some where the stressed syllable has a lower F0 or a smaller F0 excursion than the preceding or following syllable. For English, there is substantial variation in the F0 contours associated with pitch accents in spontaneous speech, leading some to claim that F0 is not a reliable correlate of accentual prominence (Kochanski, Grabe, Coleman, & Rosner, 2005). Differences among English dialects in the use of acoustic cues of prominence have also been reported (Smith & Rathcke, in press). In this paper ‘English’ refers to Midwest American English, ‘Spanish’ to Castilian Spanish and ‘French’ to Parisian French.

In French, although the AP-final accents do not confer prominence on the word in the way that pitch accents do in English, they do serve to distinguish their associated syllables through a combination of acoustic features that are similar to those that mark pitch accents in English. AP-final accents are manifest in a tonal feature assigned to the final syllable of the accentual phrase (excluding schwa), with an F0 excursion as the primary acoustic correlate and increased duration as a secondary correlate (Di Cristo, 1998; Welby, 2006). A tonal feature can also be assigned to a syllable for rhythmic or pragmatic (emphasis) reasons. These types of accentuation occur on the first or second syllable of a content word (Astésano, 2001; Jun & Fougeron, 2002), but do not mark the syllable as metrically strong (Welby, 2006). Syllables associated with both types of prominence are lengthened, though the durational properties of this accent also differ from those of the default phrase-final accent (Astésano, 2001).

In a task where participants are asked to identify prominent words on the basis of acoustic enhancement, we expect speakers of all three languages to pay attention to duration, F0 and intensity. The degree to which each of these three cues is attended to by native speakers, however, may differ across the languages. In this respect, it is interesting to note that advanced learners of Spanish whose first language is English use variations in duration and pitch to identify lexical stress in Spanish stimuli differently than do native Spanish speakers (Ortega-Llebaria, Gu, & Fan, 2013).

2.3. Word frequency

In Section 2.1 we mentioned the role of information status in the location of nuclear prominence in English; nuclear prominence may be shifted leftward from its default position on the phrase-final content word if that word is discourse-given (3a). Information status distinctions such as *new* and *given* relate to the predictability of a word based on the prior discourse context. Frequency measures also contribute to word predictability, with the usage (token) frequency of a word based on its occurrence in a large sample of spoken or text materials as one such factor. Studies examining English prominence from a production standpoint find that words with high usage frequency are more likely to be phonetically reduced, with shortened duration and hypoarticulation, compared to words with lower usage frequency (Aylett & Turk, 2004, 2006; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009). These effects are similar to what is observed for words that have been recently mentioned in prior discourse (lexically given), or whose referential meaning is accessible from discourse context (referentially given) (Calhoun, 2010; Watson, Arnold, & Tanenhaus, 2008). Similar evidence of word frequency effects on prominence patterns in Spanish spontaneous conversational speech is reported in Rao's (2009) study of Barcelona Spanish, which shows that high frequency words and words that are lexically given are likely to be deaccented (with reduced acoustic prominence) in conversational speech.

We note the similarity between effects of word frequency and information status on acoustic prominence measures (English, Spanish), or on pitch accent assignment (Spanish). While we are interested in the broader class of predictability effects on prominence, here we only examine the effects of word frequency among other possible factors related to predictability.[1] Specifically, building on earlier work showing frequency effects on prominence perception in English (Cole, Mo, & Baek, 2010), we wonder if similar effects arise in French and Spanish. This question is of particular interest given that French and Spanish differ from English in the effect of information structure on prominence (Section 2.1).

2.4. Hypothesized effects on prominence ratings

The preceding sections establish that each of the languages investigated here exhibits prominences that are phonetically expressed through pitch excursions, increased duration, and (for English and Spanish) increased intensity. Differences among the languages arise in the relative strength and consistency of the acoustic cues, in the location and density of prominences within a prosodic phrase, in the association of prominence with contrastive or narrow focus, and possibly in effects of word predictability on prominence assignment and its phonetic realization. We turn now to ask whether and how any of these differences influence the way listeners perceive prominence in each language.

We investigate prominence perception through the analysis of native listeners' perceptual ratings of prominence in heard utterances. Listeners gave binary ratings of prominence, which makes it possible for them to rate prominence in close to real time, while listening to a speech sample (see Section 3 for further details). This annotation method more closely approximates the task of speech perception in everyday conditions than a task requiring gradient ratings (e.g., Astésano et al., 2012), although it does not reflect differences in the degree of prominence that may be perceived by a listener (Buxó-Lugo, Toscano, & Watson, 2018).

We examine acoustic cues, non-acoustic contextual factors, and prosodic factors that influence variation in listeners' prominence ratings in English, French and Spanish, with two statistical models. The first model tests a set of five hypotheses about word-level factors that condition prominence rating, and differences in prominence rating among the studied languages. Hypotheses 1-4 deal with the role of word-level acoustic and contextual cues in prominence perception. Hypothesis 5 addresses differences in the relative effects of acoustic cues and non-acoustic contextual factors depending on whether participants are explicitly instructed to attend to acoustic cues or speaker-intended utterance meaning as it relates to prominence. The second model tests two additional hypotheses about the relationship between the prominence ratings of a word and its intonational features (in the ToBI framework) as identified by trained annotators. Hypothesis 6 relates prominence rating to the status of a word as accented (nuclear vs. prenuclear) or unaccented. Hypothesis 7 is restricted to words that are accented in the ToBI annotation and relates prominence rating to the type of pitch accent, and specifically, to the presence of a contrastive or narrow focus-marking pitch accent. These hypotheses are formulated as follows.

H1. Effects of word-level acoustic prominence. Words with higher peak F0, longer duration and greater intensity are more likely to be rated as prominent compared to words with lower degrees of acoustic prominence for the same measures. Acoustic effects on prominence ratings are expected in all three of the studied languages.[2]

The next two hypotheses have to do with contextual effects due to expectation-driven processing. The basic idea is that a listener may judge prominence based on properties of a word that are often or typically associated with prominence, even if the word does not manifest acoustic prominence. This situation could arise due to variability in the speaker's production of prominence, e.g., when a prominent word is weakly cued as such in its acoustic realization. Non-acoustic factors that may influence perceived prominence include word frequency, given that high-frequency words are often produced with reduced prominence, and sentence position. In languages like English, a structural prominence is licensed in the final position of the prosodic phrase, and yet under certain conditions related to focus, prominence may be assigned to an earlier word and the final word may be realized without acoustic prominence.

H2. Boundary effect. A word that is judged by a listener to be final in a prosodic phrase (specifically, preceding a boundary with the ToBI label ip or IP) is more likely to be rated as prominent by the same listener. Due to the fixed phrase-final location of nuclear prominence in Spanish and the AP-final location of all prominences in French, the boundary effect is expected to be greater for Spanish and French than for English, where nuclear prominence is not always phrase-final.[3]

H3. Word frequency effect. Words with high usage (token) frequency are less likely to be rated as prominent than lower-frequency words. We predict a similar effect of word frequency on prominence ratings in all languages, as prior work cited earlier does not provide a clear basis for predicting differences.

We introduce one more hypothesis about contextual effects on prominence rating, recognizing a possible association between prominence and the part of speech category (POS) of a word. This association, if it is observed, would be indirect, arising due to the fact that some POS categories are more common than others in sentence positions where prominence is more frequently assigned, and/or due to differences in usage frequency of words belonging to different POS categories. There may also be specific POS effects that disfavor accenting

[1] Our decision not to examine effects of information status on prominence perception was made in consideration of the speech materials used in our study. Our participants rated prominence in speech excerpts drawn from the middle of longer dialogues; they did not have access to the discourse context preceding the speech excerpts they heard, so the information status conditions established in the prior discourse would not be reflected in their interpretation of acoustic prosodic cues, or as context informing top-down processing. We refer the reader to Bishop (2012) and Jun and Fougeron (2018); Im, Cole and Baumann (2018), which examine information structure effects on prominence ratings in English, using speech materials more suitable for that purpose.

[2] We examine three acoustic variables measuring peak F0, duration and intensity as possible factors influencing prominence perception, based on a large body of prior work examining acoustic correlates of phrasal prominence or pitch accent, including many of the papers in this special issue (see also Wagner and Watson (2010) and Cole et al. (2015) for an overview and references to this literature). We are not claiming that these are the primary acoustic correlates of prominence in all three languages studied here, nor that these are the most appropriate measures of the acoustic parameters of interest for each language. Notably, we do not have a measure of F0 dynamics, which is clearly relevant in distinguishing among different accent types. See Schweitzer (in press) and Breen et al. (2010) for production studies of German and English that examine other acoustic measures related to prominence.

[3] The picture may be complicated slightly for French in utterances with focus, which could result in a word being perceived as prominent that is final in an Accentual Phrase, but where the Accentual Phrase boundary may not be perceived as very salient (Smith, 2012). But in the vast majority of cases, the occurrence of prominence and phrasal boundaries are tightly linked in French.

of certain POS categories, such as the avoidance of accentual prominence on verbs in English (Büring, 2016; Gussenhoven, 1983).

H4. Part of speech (POS) effect. The POS category of a word may influence the likelihood of prominence rating, reflecting differences among POS categories in how often the category occurs in the position of obligatory (nuclear) prominence, or differences in the usage frequency of words belonging to different POS categories.

The specific POS effects are predicted to vary across the studied languages. For English, we expect nouns will be more often rated as prominent than verbs because nouns are frequent in sentence-final position (the default location for nuclear prominence), while many verbs are typically non-final (for instance, transitive verbs), or ineligible for pitch accent due to other factors, as illustrated in examples (3) and (4) above. For Spanish, there may be a similar difference in prominence ratings for nouns and verbs, given that verbs and adverbs are reported to be more often deaccented than nouns and adjectives (Face, 2003; Rao, 2009). We are not aware of any basis for a predicting a difference in the likelihood of prominence rating for nouns vs. verbs in French. All three languages may exhibit a lower likelihood of prominence rating for adverbs as a class, due to the frequent presence of semantically bleached adverbs like English *really* and *very,* which may be used rhetorically or stylistically without emphasis in conversational speech.

As reviewed above (Section 2.1), an important difference between English, French and Spanish has to do with the relationship between prominence and information structure, including focus. In English, where the relationship is strongest, acoustic prominence should largely coincide with prominence that relates to speaker-intended meaning, such that a listener may arrive at the same prominence rating for a word regardless of whether they apply acoustic or meaning-related criteria. In other words, a word may be rated as prominent due to its acoustic prominence or due to a discourse context that specifies contrastive or narrow focus, and these criteria should often converge on the same rating. The same prediction does not hold for Spanish or French, where syntactic rather than prosodic means are more typically used to mark contrastive or narrow focus. To test the relationship between prominence defined acoustically and prominence defined in terms of its meaning function, we compare prominence ratings obtained using two task instructions (see details in Section 3).

H5. Instruction effect. Within each language, prominence ratings will differ depending on whether listeners are directed to attend to acoustic criteria (signal-based processing) or criteria related to utterance meaning (expectation-based processing). Instruction-dependent rating differences will be greater for French and Spanish than for English.

Interactions. There are some rather obvious ways that the task manipulation might interact with the effects in Hypotheses 1–4. First, in each of the languages the acoustic effect on prominence rating should be greater when listeners are explicitly instructed to listen for such effects, though the interaction of task and acoustic effects on prominence ratings should be smaller for English than for French and Spanish. Second, all of the effects of non-acoustic contextual factors (boundary, frequency, POS) should be greater for ratings done when listeners are guided to consider the meaning of an utterance, but again, this difference should be smaller for English.

The final two hypotheses below address the relationship between the prominence rating of a word and its pitch accent status and type in terms of ToBI pitch accent features. A relationship between naïve prominence ratings and ToBI pitch accents is expected for two reasons. First, the ToBI system is based on the Autosegmental-Metrical (AM) theory of intonation (Beckman & Pierrehumbert, 1986; Ladd, 2008), in which pitch accents are assigned to metrically strong positions. Therefore, a pitch-accented word is also expected to manifest phonetic exponents of phrasal prominence, to the extent that phrasal prominence is marked at all in the language. Second, a pitch accent is often (though not always) acoustically realized in a perceptually salient pitch excursion. Combining phrasal stress and a pitch excursion, we expect a pitch-accented word to be perceptually salient, and rated as prominent. That pitch accent status may correlate with perceived prominence is supported by evidence from English in Bishop et al. (in press) and from Baumann and Winter's (2018) study of German.

H6. Effect of accent status. A word that is assigned a pitch accent in a ToBI annotation will be more likely to be rated as prominent than an unaccented word, all else equal, and a nuclear accented word will be more likely to be rated as prominent than a pre-nuclear accented word.

H7. Effect of contrastive or narrow focus-marking accent. A word that bears the type of ToBI pitch accent most commonly associated with contrastive or narrow focus in a language will be more likely to be rated as prominent compared to words with other types of pitch accents. This effect is expected to be stronger for English than for Spanish, while for French the prediction is less clear given the uncertain status of a focus-marking accent.

3. Methods

Separate prominence rating experiments were conducted in the USA (Illinois), France (Lyon) and Spain (Valladolid), with native speakers of English, French and Spanish, respectively, using the same experimental procedure and analysis methods.

3.1. Participants & materials

30 native speakers of each language were recruited from university student communities in each location (English: 18 f, 12 m; French: 29 f, 1 m; Spanish: 11 f, 19 m). Participants self-reported as having no deficits in speech, hearing or reading, and normal or corrected-to-normal vision. Participants listened to excerpts of spontaneous speech in their native language, which were drawn from publicly available corpora of spontaneous, conversational speech (Table 1). Transcriptions and source file names for the complete excerpts can be found in Appendix D (online supplement).

Participants rated prominence for the full set of excerpts as described in Table 1 (2907 words total). 94 words from the full dataset were excluded from analysis due to coding and measurement errors, leaving 2813 words. Due to the low rate of prominence marking on function words, function words were also excluded. (See Appendix A Fig. F in the online

Table 1
Sources of speech materials used for prominence rating experiments. Each excerpt was taken from a different speaker within the source corpus. Shown are the number of speakers (spkr), the total number of words over all excerpts (words) and the mean duration of the excerpts in seconds (dur).

Language	Corpus name	Corpus citation	# spkr	# words	dur (s)
English	Buckeye Corpus	Pitt, Johnson, Hume, Kiesling, and Raymond (2005)	16	931	17.5
French	Corpus du Français Parlé Parisien	Branca-Rosoff et al. (2012)	14	1079	24.2
Spanish	Glissando Corpus	Garrido et al. (2013)	16	897	15.5

supplement for a graph of the distribution of average by-item prominence ratings, pooled across participants.) A total of 1667 function words were removed from the analysis dataset (509 English, 630 French, 547 Spanish). An additional 24 words with undefined F0 (e.g., due to creaky voice), and 13 words with outlier F0 measures were also removed, leaving 1109 words submitted for statistical analysis (345 English, 428 French, and 336 Spanish). Criteria for data exclusion are detailed in the Methods supplement in Appendix A.4 (online supplement).

3.2. Prominence rating task

The experiments were conducted in a phonetics lab or quiet room in each location. Prominence ratings were obtained using the method of Rapid Prosody Transcription (Cole, Mo, & Hasegawa-Johnson, 2010; Cole, Mo, & Baek, 2010; Roy, Cole, & Mahrt, 2017; Cole, Mahrt, & Roy, 2017), and administered using LMEDS (Mahrt, 2016a), a web platform for prosodic rating and annotation experiments.[4] Speech excerpts were presented auditorily one at a time, through headphones, and a transcript of the excerpt with punctuation and capitalization removed was simultaneously displayed on a computer screen. Although our focus in this paper is on prominence ratings, participants also rated boundaries for the same speech excerpts. Boundary ratings were needed to identify for each prominence rating whether the prominent word was perceived as final in a prosodic phrase ("chunk"). The participant first listened to the excerpt and clicked on every word they heard as preceding a boundary between "chunks of speech". Then, listening again to the same excerpt, the participant marked words they heard as "prominent". A participant's boundary marks were visible to him/her during the prominence marking task for the same excerpt. Participants proceeded through the excerpts at their own pace, rating prominences or boundaries in real time while listening to the audio recording, based on auditory impression alone. Participants could listen to an entire excerpt twice for each rating task, but they could not stop or restart the audio recording during playback.

Each participant performed boundary and prominence marking in two blocks with the same materials, and with block order counter-balanced across participants. The blocks differed only in the criteria the participants were instructed to use in rating prominences and boundaries. In the *acoustic instructions* block, participants were told to mark a boundary where they heard a "break, discontinuity or disconnection in the speech stream, strong or subtle" and to mark a prominence where they heard a word stand out by "being louder, longer, more extreme in pitch, or more crisply articulated" (with Spanish and French translations for Spanish and French experiments, respectively). In the *meaning instructions* block, subjects were asked to mark boundaries where the audio could be "segmented with minimal disruption of the speaker's intended meaning" and were asked to mark as prominent the words that "convey the main points of information as you think the speaker intended." The full instructions in each language are provided in Appendix C (online supplement).

[4] See also Wagner et al. (in press) and Bishop et al. (in press) for other studies of prominence perception using Rapid Prosody Transciption.

3.3. Acoustic measures

All acoustic measures were extracted from WAV files and manually corrected TextGrids obtained through forced alignment of the audio file with written transcripts, with Praat scripts written by the authors (Mahrt, 2016b). Acoustic correlates of prominence were examined through three measures taken at the word level, for content words only. Details of the acoustic measurement methods are reported in Appendix A.2 (online supplement).

The durational effect of prominence was measured with the **Word Phonerate** variable, which is a measure of word duration that is normalized for the speech rate of the utterance using Pfitzinger's (1998) RateLR formula. Word Phone-rate values are z-scored values of phones/second and indicate how much the word is lengthened or shortened relative to the speech rate of the utterance (including function words) and taking into account the number of phones in the word.

Intensity is measured as the mean intensity in dB of the stressed vowel in each content word for English and Spanish, and mean intensity of the last non-schwa vowel for French.

F0 as a correlate of prominence is measured in terms of the **Maximum F0** (in semitones) in the stressed vowel of each content word for English and Spanish, and in the final non-schwa vowel for French. F0 values were extracted using the Praat autocorrelation function, with gender-specific pitch ranges of 70–250 Hz for males and 100–300 Hz for females (Vogel, Maruff, Snyder, & Mundt, 2009) that were manually checked and adjusted in the event of files with clipped pitch excursions. Pitch halving and pitch doubling errors were automatically checked and hand-corrected, and median filtering was used to smooth micro-perturbations at consonant–vowel junctures. Raw Intensity and Max F0 values were normalized using the mean and standard deviation of intensity of stressed vowels in a 5-word window centered on the target word. For the first two and last two words in each speech excerpt, the window was shortened (e.g., the first word is normalized using a 3-word window consisting of the first three words in the stimulus utterance). The result is a measure of the degree to which a word stands out from its neighbors in terms of intensity and peak F0.

3.4. Non-acoustic contextual factors

Participants rated each word for prominence and for the presence of a boundary following the word. These ratings were binary coded (0, 1) for each word, separately for each participant. The **Boundary** feature of a word is thus a binary feature specific to a participant, which is used to test an association between prominence rating and boundary rating for each participant. **Word frequency** is based on word counts from the Buckeye corpus for English, from the Glissando Spanish informal-dialogue sub-corpus for Spanish, and from the Lexique 3.8 database (New, Pallier, Ferrand, & Matos, 2001) for French. For each item, the number of occurrences of that word in its corpus was divided by the total number of words in that corpus, and the natural logarithm of the resulting value was taken as the word frequency measure. **POS** labels were assigned manually to each word for English and Spanish, using only the following labels: adjective, adverb, complementizer, conjunction, subject-verb contraction, negative contraction, polarity, determiner, discourse marker, disfluency, noun, preposition, pronoun, auxiliary verb, main verb. For French, POS labels were created using the Stuttgart TreeTagger (Schmid, 1994), hand-corrected, and then simplified to the system used for English and Spanish in order to have a common set of POS labels for all three languages.

ToBI annotation. For each language, the speech excerpts were prosodically labeled by two of the authors working together to assign labels by consensus following the standard ToBI conventions for English (MAE_ToBI: Beckman & Hirschberg, 1993; Beckman et al., 2005), French (Fr_ToBI: Delais-Roussarie et al., 2015), and Spanish (Sp_ToBI: Estebas-Vilaplana & Prieto, 2008; Gutiérrez González & Aguilar Cuevas, 2015; Hualde & Prieto, 2015). This was a full ToBI labelling that specified pitch accents and boundary tones. The last pitch accent in each phrase (preceding an ip or IP boundary) was identified as nuclear. ToBI labels were compared with average prominence ratings (p-scores, see Section 3.5) to establish the relationship between prominence as judged by expert and non-expert listeners and to determine possible differences in perceived prominence among ToBI accent labels.

Since pitch-accent labels do not always have the same interpretation across ToBI systems for different languages, and languages also differ in the pragmatic value of accentual contours, a brief explanation may be useful. In particular, notice that, in Sp_ToBI, L+>H* is a rising contour with a valley towards the beginning of the lexically-stressed syllable and a peak in the next syllable. This contour is considered to lend less prominence than L+H*, a rise with a peak within the stressed syllable. Typically L+>H* is found in prenuclear position and L+H* in nuclear position. The French ToBI system (Delais-Roussarie et al., 2015) has a smaller set of labels than English or Spanish. Of these, two (H* and the less common L*) mark the prominent syllable at the end of an Accentual Phrase. These are considered to be pitch accents although they do not share all the properties of pitch accents in Germanic languages (see Welby, 2006 for discussion). An Accentual Phrase may optionally have an early high tone labeled as Hi; opinions differ as to whether these are pitch accents (e.g., Post, 2011), or phrase accents (e.g., Jun & Fougeron, 2002). The ToBI guidelines also allow for a couple of bitonal accents but these occur only in very specific contexts or dialects which were not found in our data.

3.5. Statistical models

Prominence ratings were submitted to mixed effects logistic regression in R using glmer() in the lme4 package (Bates et al., 2014). The dependent variable was the prominence rating (0 or 1) from each participant for each content word in the analysis dataset. In the combined analysis dataset, including prominence ratings for the English, French and Spanish materials, there were a total of 66,540 observations (1109 content words $\times$ 30 transcribers $\times$ 2 instructions). Each observation was coded for the categorical factors of *Language* (3 levels: English, French and Spanish), *Instruction* (2 levels: Acoustic, Meaning), *Boundary* (0 or 1), and *POS* (4 levels: adjective, adverb, noun, or verb), and for the continuous variables of *(Word) Frequency, Phonerate, Intensity,* and *Max F0*.

The model fixed effects included the eight factors just listed, the interaction of *Language* $\times$ *Instruction,* two-way interactions between *Language/Instruction* and the six word-level predictors (*Boundary, Frequency, POS, Phonerate, Intensity,* and *Max F0*), and the three-way interactions between *Language, Instruction* and the six word-level predictors (*Boundary, Frequency, POS, Phonerate, Intensity,* and *Max F0*), for a total of 27 fixed effects in addition to the intercept, which represents the grand mean of the response variable, for a total of 54 fixed parameters. All factors were coded using sum contrasts, and all continuous variables were scaled in R prior to regression analysis, which allows straightforward interpretation of regression coefficients. Random intercepts for transcriber and word (=item) were also included.[5] The likelihood ratio test was used to obtain *p*-values with the mixed() function in the afex package (Singmann, Bolker, & Westfall, 2015). Additional details on the statistical methods are presented in Appendix A.2 (online supplement).

We examine the logistic regression model output in two ways. First, the model tells us which of the six word-level factors (and interactions among factors) significantly predict individual participants' prominence ratings, the relative strength of these effects for each language, for each instruction, and for combinations of language and instruction. We also assess the overall fit of the model to determine how much of the variation in prominence ratings the model accounts for. Model fit was assessed by using the fixed effects from the model to predict the log-odds of each item being marked as prominent under each instruction, ignoring the random effects associated with transcriber and word, resulting in 2218 predictions. These results from the full model are presented in Section 4.1.

Second, a by-items analysis was carried out to compare the model predictions to the actual prominence ratings for each word under each language-instruction condition. This comparison allows us to assess the accuracy of the model in predicting prominence ratings for each pairing of language and instruction. To test for differences in model accuracy among the language-instruction conditions, the log-odds prediction

[5] The maximal random effects structure (Barr et al., 2013) was computationally infeasible given the size of the dataset and the number of random effects parameters each slope introduces.

for each word was converted to a p-score, which is the average prominence rating across participants (using the inverse of the logit transformation), and subtracted from the true p-score for that word. The comparison of predicted and actual p-scores is presented in Section 4.2.

Participants' binary prominence ratings were compared with ToBI pitch accent labels in a separate analysis for each language. For each language a mixed effects logistic regression was run in R using the lme4 package (Bates et al., 2014) with each transcriber's response to each item coded as 0 (not prominent) or 1 (prominent) as the dependent variable. The interaction between ToBI labeling (Unaccented, H*, etc.) and accent location (prenuclear or nuclear) was included as a fixed effect with sum contrast coding, with random intercepts for transcriber and for word token as item. Planned contrasts were performed using the lsmeans package (Lenth, 2016) to obtain the estimated difference in log-odds for each contrast, as well as a test of statistical significance. The grand mean of each category for comparison was used to control for imbalances in the frequency of occurrence of different pitch accents in prenuclear and nuclear position. The comparison of the participants' prominence ratings and ToBI pitch accent labels is presented in Section 4.3.

4. Results

4.1. Regression model likelihood ratio tests, effect estimates and model fit

Table 2 presents the results of likelihood ratio tests for the full model predicting English, French and Spanish prominence ratings, for each of the fixed effects and their interactions. For details of the model output, see Appendix B (online supplement). The likelihood ratio tests show that all main effects are significant. In particular, with all other factors are held constant, there are significant differences in the overall rate of prominence rating among the three languages, and significant differences based on whether prominence is rated based on acoustic or meaning instructions. There is also a significant interaction of Language × Instruction, which tells us that the effect of instruction on prominence ratings differs among the three languages. There are significant main effects of the six word-level factors on prominence ratings pooled across the three languages and two instructions. The interaction of Language × Frequency is significant, but there are no other significant interactions with Language, which tells us that with the exception of Frequency, the word-level predictors have very similar effects across the three studied languages. A different pattern of results is seen in the two-way interactions with Instruction. The effects of word-level factors other than Intensity differ depending on whether participants are instructed to use acoustic or meaning criteria. Looking at the three-way interactions, we find that the effects of four out of six word-level factors, Frequency, POS, Boundary and Phonerate, differ significantly among the language-instruction conditions. The other two word-level factors have a more stable pattern of effects across the conditions. Max F0 differs by instruction but not by language, and the effect of Intensity does not differ between either by language or instruction.

Table 2
Likelihood ratio tests for the full model (as described in Section 3.4), including English, French and Spanish prominence ratings. Gray shading marks non-significant effects.

Overall model fit: r^2 = .45		
Fixed Effect	χ^2	p
Language	$\chi^2(2) = 12.18$	.002
Instruction	$\chi^2(1) = 335.08$	< .001
Word Frequency	$\chi^2(1) = 190.63$	< .001
Part of Speech (POS)	$\chi^2(3) = 95.84$	< .001
Boundary Marking	$\chi^2(1) = 119.46$	< .001
Phonerate	$\chi^2(1) = 75.14$	< .001
Intensity	$\chi^2(1) = 11.46$	.001
Max F0	$\chi^2(1) = 14.81$	< .001
Language : Instruction	$\chi^2(2) = 49.98$	< .001
Language : Word Frequency	$\chi^2(2) = 13.61$	.001
Language : POS	$\chi^2(6) = 9.27$	.159
Language : Boundary Marking	$\chi^2(2) = 1.29$	.525
Language : Phonerate	$\chi^2(2) = 0.08$	.962
Language : Intensity	$\chi^2(2) = 3.59$	.166
Language : Max F0	$\chi^2(2) = 3.54$	.170
Instruction : Word Frequency	$\chi^2(1) = 74.38$	< .001
Instruction : POS	$\chi^2(3) = 109.04$	< .001
Instruction : Boundary Marking	$\chi^2(1) = 68.04$	< .001
Instruction : Phonerate	$\chi^2(1) = 121.32$	< .001
Instruction : Intensity	$\chi^2(1) = 1.44$	.230
Instruction : Max F0	$\chi^2(1) = 92.08$	< .001
Language : Instruction : Word Frequency	$\chi^2(2) = 16.38$	< .001
Language : Instruction : POS	$\chi^2(6) = 45.67$	< .001
Language : Instruction : Boundary	$\chi^2(2) = 8.58$	.014
Language : Instruction : Phonerate	$\chi^2(2) = 15.80$	< .001
Language : Instruction : Intensity	$\chi^2(2) = 1.89$	.389
Language : Instruction : Max F0	$\chi^2(2) = 3.28$	.194

The significant main effects from the full model were used to predict the log-odds of prominence marking for each word in the analysis dataset.[6] These model predictions are correlated with the actual p-score for each word (average prominence rating across participants), with $r^2 = 0.45$. In other words, the fixed effects from our model account for roughly 45% percent of the variation in average prominence ratings overall. We also ran separate regression models with only acoustic factors (Phonerate, Max F0, Intensity), and with only non-acoustic contextual factors (Boundary, Word frequency, POS) to assess the sufficiency of each factor type in predicting prominence ratings pooled over language and instruction. The correlation analysis of predicted and actual prominence ratings for the model with only acoustic factors returns $r^2 = 0.22$, while the model with only non-acoustic contextual factors returns $r^2 = 0.38$, indicating that neither of these captures as much variance as the two sets of predictors combined.

Table 3 provides estimates of effect size for the main effects in the model, measured in terms of the increase/decrease in the likelihood of a word being rated as prominent. The grand means in Table 3 represent the log-odds that a content word will be rated as prominent in a given language under a given instruction, holding the effects from all other predictors at their average values (e.g., the average boundary rating across participants is 0.5, the scaled word-frequency, phonerate, intensity and maximum F0 values are equal to 0, and prominence ratings are

[6] See Appendix B.3 (online supplement) for the density plot of by-item errors and results of Shapiro-Wilk test showing that errors are normally distributed.

Table 3
Regression model estimates of the overall rate of prominence marking (log-odds) for each language-instruction pair.

		English		French		Spanish	
		Acoustic	Meaning	Acoustic	Meaning	Acoustic	Meaning
Intercept (Grand Mean)		−1.036	−0.734	−1.552	−0.733	−0.622	−0.130
Word Frequency		−0.893	−1.060	−0.470	−0.647	−0.476	−0.921
Part of Speech	Adjective	0.485	0.601	0.511	1.077	0.245	0.518
	Adverb	−0.225	−0.355	−0.059	−0.846	0.256	−0.213
	Noun	0.162	0.253	0.092	0.146	0.169	0.297
	Verb	−0.422	−0.499	−0.544	−0.377	−0.670	−0.601
Boundary Marked		0.540	0.269	0.647	−0.019	0.584	0.163
Phonerate		−0.380	−0.251	−0.513	−0.165	−0.484	−0.159
Intensity		0.231	0.178	0.063	0.013	0.150	0.167
Maximum F0		0.300	0.130	0.338	0.071	0.158	−0.077

averaged over the different parts of speech).[7] All of the grand mean values are negative, which means that in all language-instruction conditions a word has a lower-than-chance likelihood of prominence rating. In other words, participants are to some degree conservative in prominence marking in each of the studied languages. Fig. 1 illustrates the comparison of grand means across languages, showing a higher likelihood of prominence rating (i.e., smaller negative grand mean) for Spanish than for English or French, indicating that Spanish participants were less conservative in prominence rating than were English and French participants, and this difference holds for prominence ratings by both acoustic and meaning instructions. Fig. 1 also shows that for all three languages, words are more likely to be rated as prominent under meaning instructions than under acoustic instructions. The effect of instruction is greatest for French, due to the much lower likelihood of prominence rating under acoustic instruction in French, compared to English or Spanish.

Turning to the effects of individual word-level factors on prominence rating in Table 3, we see that each factor has a similar effect of increasing or decreasing the likelihood of prominence ratings over all three languages, but the strength of that effect varies by language and instruction. Differences in the overall effect size for each factor are displayed graphically in Fig. 2 for factors other than POS (for which see Fig. 3).[8] The strongest effect is that of word frequency: more frequent words are much less likely to be rated as prominent, as are words with lower phonerate (i.e., spoken more quickly). A word that is marked as preceding a boundary is more likely to be rated as prominent (by the same rater), as are words that have greater Intensity or higher Max F0 (with the exception of Spanish prominence rated under meaning instructions).

Also shown in Fig. 2, the effect of instruction on acoustic factors goes in the expected direction for all the languages; when participants are rating prominence under acoustic instructions, the acoustic factors (Phonerate, Intensity, Max F0) have a greater effect than they do for ratings assigned under meaning instructions. The sole exception is in the effect of Intensity on prominence ratings in Spanish, which has a very

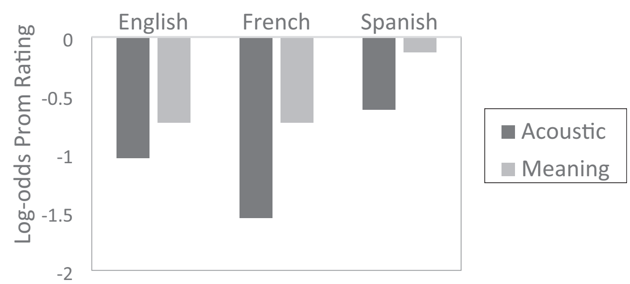

Fig. 1. The effect of instruction and language on the overall likelihood (log-odds) of a word being rated as prominent. Likelihoods are estimated from regression model predictions (see Section 3.4). The *y*-axis is the difference in log-odds in relation to chance (50%, at zero on the *y*-axis). All effects are negative, indicating decreased likelihood of prominence rating relative to chance.

slightly greater effect under meaning instructions. The effect of instruction is in the opposite direction for Word Frequency, as expected, with a stronger effect on prominence rating under meaning instructions than under acoustic instructions, for all three languages.

The interaction of Instruction and POS is more variable across the languages. As shown in Fig. 3, adjectives and nouns have a higher likelihood of prominence rating in all three languages, and this effect is strongest for ratings based on meaning in all languages. On the other hand, verbs have a lower likelihood of prominence rating an all languages, with little difference in effect size due to instruction for English and Spanish, but with an unexpectedly stronger effect in the acoustic instruction condition for French. In all but one condition, adverbs have a decreased likelihood of prominence rating, especially for ratings under meaning instructions. The exception is with Spanish prominence ratings under acoustic instructions, where adverbs actually have an increased likelihood of prominence.

To aid in comparing the size of the Instruction effect among the word-level factors, Fig. 4 graphs the effect size measured as the absolute value of the difference in estimated effects (Acoustic minus Meaning). The higher the bar, the greater the effect of instruction. This graph clearly shows that the greatest effects of Instruction were in modulating the effect of Word Frequency, POS (Adj. and Adv.), Boundary, Phonerate and Max F0 on prominence ratings in French and Spanish.

4.2. Model predictions by item and language

Next we look at model predictions (log-odds prominence rating) for each word compared to actual p-scores (average prominence ratings). Fig. 5 shows this relationship for each

[7] The POS estimates represent the difference between each POS level's log-odds of being marked and the grand mean for that language-instruction pair. For boundary estimates, the effect size is the difference in log-odds between a word that is marked as preceding a boundary and a word which does not precede a boundary. The continuous predictor estimates (word frequency, phonerate, intensity, and maximum F0) represent the change in the log-odds of being marked as prominent for a one standard deviation increase in the predictor. All effect sizes assume that only a single factor is being changed.

[8] Graphs showing in more detail the relationship between predicted prominence (log-odds) and each word-level factor over the range of variation of the word-level factor are included in Appendix B.2 (online supplement).

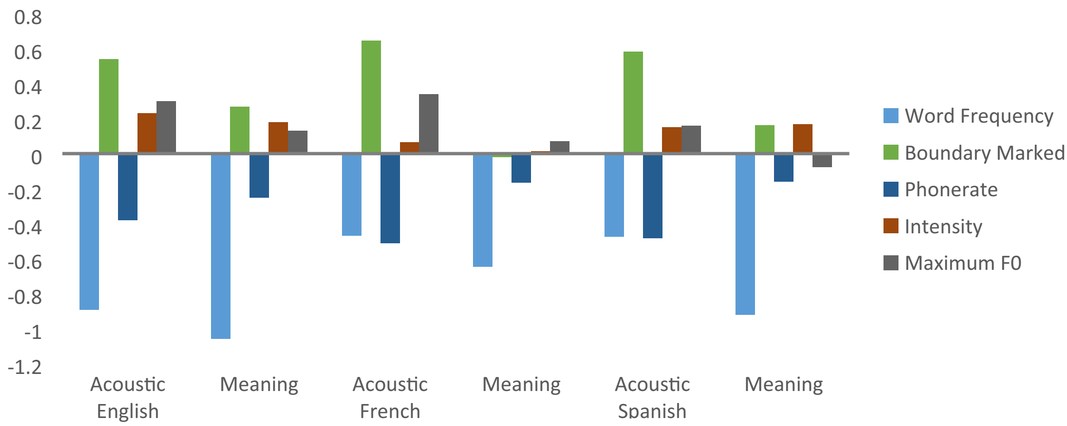

Fig. 2. The effect of word-level factors (acoustic factors, word frequency and boundary) on the overall likelihood (log-odds) of a word being rated as prominent, grouped by instruction and language (*x*-axis). Likelihoods are estimated from regression model predictions (see Section 3.4). The *y*-axis is the difference in log-odds from chance likelihood (at zero on the *y*-axis). Bars extending below the horizontal line are effects that decrease the likelihood of prominence rating, and bars extending above the line are effects that increase the likelihood of prominence rating.

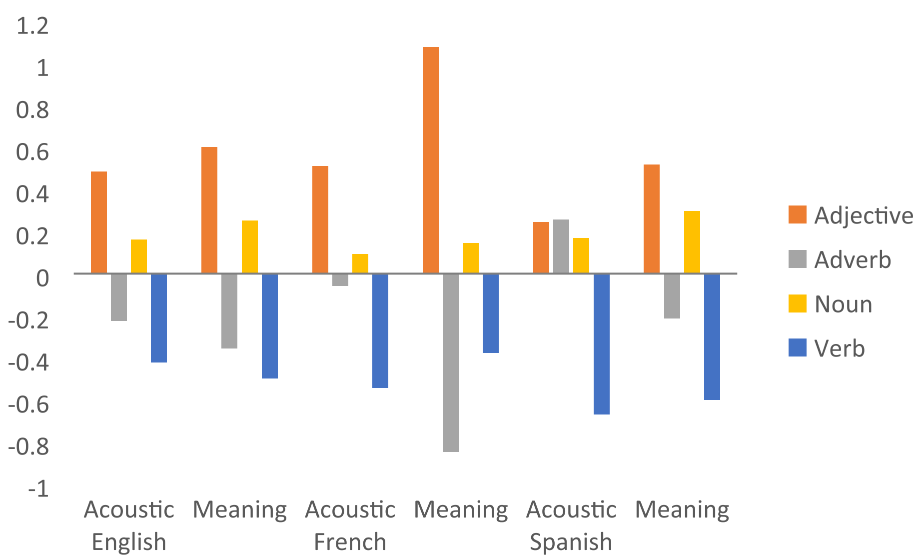

Fig. 3. The effect of POS on the overall likelihood (log-odds) of a word being rated as prominent, grouped by instruction and language (*x*-axis). Layout as in Fig. 2.

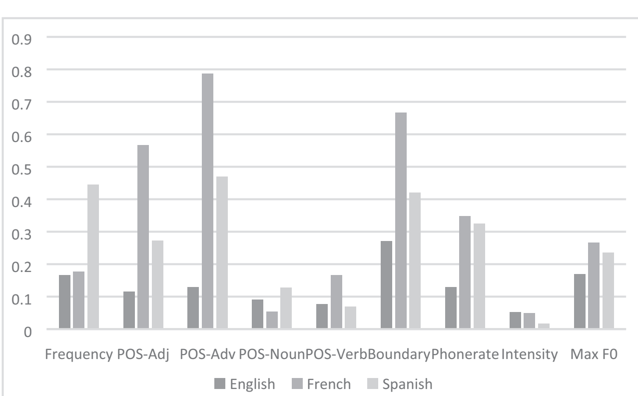

Fig. 4. The absolute value of the difference in effect estimates (log-odds) for prominences rated under acoustic and meaning instructions, for each word-level factor, by language. Effect estimates from Table 3. By taking the absolute value of the effects, this graph ignores differences in the direction of the effect on increasing or decreasing the likelihood of prominence rating, focusing only on the strength of the effect.

language-instruction pairing. On the basis of visual inspection alone, we see that the model predictions rather poorly match actual p-scores for French and Spanish prominence ratings under meaning instructions, compared to prominence ratings under acoustic instructions for these two languages. For English, on the other hand, the relationship between predicted an actual prominence ratings is similar under acoustic and meaning instructions.[9]

[9] The lower accuracy of the model for predicting prominence ratings in French and Spanish under meaning instruction is confirmed by results from a linear model predicting squared errors (predicted – actual p-scores on the basis of language-instruction condition). The model shows significant differences between errors in the French-meaning condition vs. all other conditions, and between errors in the Spanish-meaning condition vs. all other conditions. Pairwise comparisons with Tukey-adjusted p-values shows that the model performs significantly worse on French and Spanish prominence ratings under meaning instructions, and performs similarly well on prominence ratings in the other four conditions.

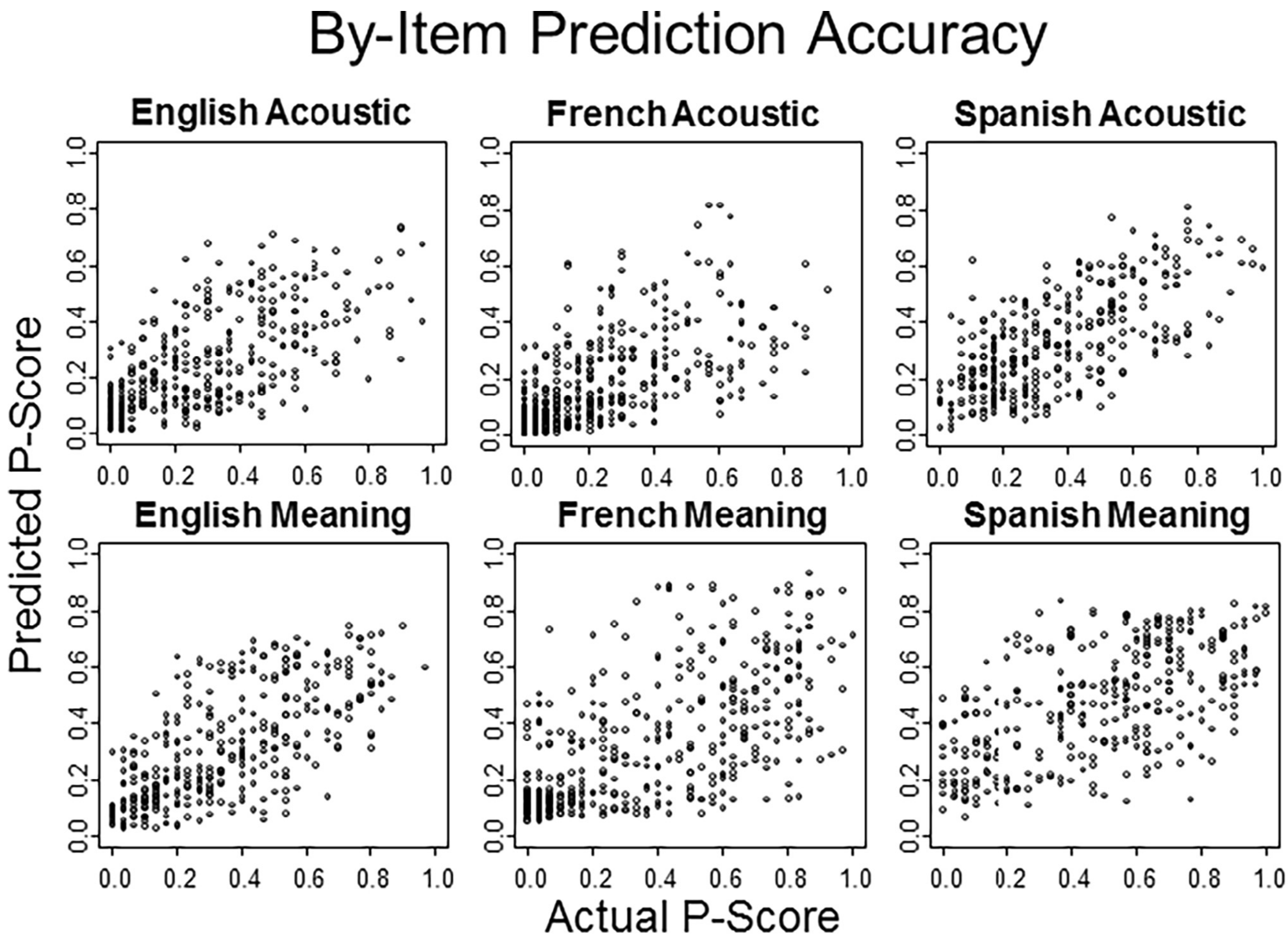

Fig. 5. Relationship between actual prominence ratings (p-score, or average of participant ratings) and predicted prominence rating (log-odds converted to p-score) for each pairing of language and instruction.

4.3. Prominence ratings and ToBI pitch accent labels.

The distribution of ToBI pitch accent labels is shown in Table 4.[10] In the English annotations, the * label without a tonal specification corresponds to the "X*?" label in the ToBI conventions (Beckman & Hirschberg, 1993), and is used when the annotator is unable to decide which among two or more pitch accent labels to assign to a particular word. Words assigned the * label are perceived to have phrasal prominence, but with pitch evidence that is ambiguous between two or more pitch accent types.

Fig. 6 compares the actual p-scores of accented and unaccented words, separating also words with a nuclear vs. prenuclear accent, where nuclear was defined as the rightmost accent in the prosodic phrase (intermediate or intonational), which for Spanish and French was on the phrase-final content word. This comparison reveals that accented words were more often perceived as prominent than unaccented words and nuclear accents are more frequently perceived as prominent than prenuclear accents in all three languages.

Table 4
Counts of ToBI pitch accent labels (including "Unaccented") by language.

English		French		Spanish	
Label	Count	Label	Count	Label	Count
Unacc.	603	Unacc.	712	Unacc.	517
H*	122	H*	217	H*	157
L*	10	L*	42	L*	23
!H*	35	Hi	79	!H*	18
L+H*	55			L+>H*	65
H+!H*	11			L+H*	82
*	25				

To test the significance of the accented/unaccented contrast in each language, the log-odds estimate of prominence marking for unaccented words obtained from the mixed effects logistic regression model was subtracted from the grand mean of the nuclear words and prenuclear words' log-odds estimates and tested for significance. Similarly, to test the nuclear/prenuclear contrast in each language, the grand mean of the prenuclear estimates was subtracted from the grand mean of the nuclear estimates. All differences are significant, as shown in Table 5.

Fig. 7 shows the distribution of empirical p-scores for different ToBI pitch accents in each language. Planned compar-

[10] For some additional accent types, fewer than ten tokens were obtained (English: L*+H = 2, ^H* = 1; Spanish: * = 5, ^H* = 3, H*+L = 1, H+L* = 7, L+^H* = 5, L+>^H* = 1), and so these items were excluded from further analysis. In the French labeling, phrase-initial aL tones and phrase-medial L tones were grouped with unaccented words, as these tones do not confer prominence and are not considered accents. Additionally, H (without a diacritic) was used in the French labeling to indicate a high tone that was difficult to classify as either Hi or H* because of its location. We exclude those tokens (=12) from further analysis.

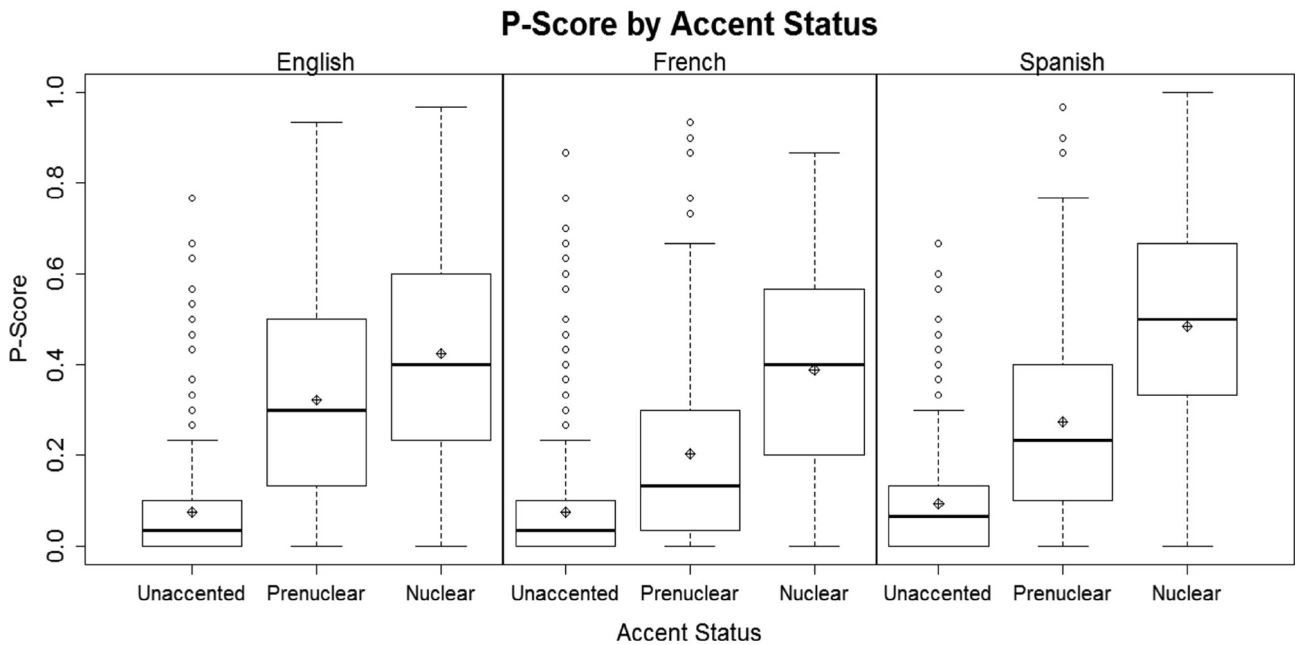

Fig. 6. Boxplots of empirical p-scores (average prominence ratings across participants and items, values range from 0-1) for words based on ToBI pitch accent status: Unaccented, Prenuclear accent, nuclear accent, by language.

Table 5
Contrast estimates (log-odds of prominence marking) for accent status as accented or unaccented, and for accented words, as nuclear or prenuclear.

Language	Accented – Unaccented				Nuclear – Prenuclear			
	Est.	SE	*z*	*p*	Est.	SE	*z*	*p*
English	2.70	0.18	14.7	<0.001	0.79	0.33	2.4	0.015
French	2.09	0.14	15.4	<0.001	0.80	0.28	2.9	0.004
Spanish	2.31	0.12	18.7	<0.001	1.30	0.21	6.3	<0.001

isons tested predicted prominence ratings for the pitch accent that is most likely to be associated with contrastive or narrow focus in each language against other pitch accents and showed significant differences in each language. For English, the L+H* pitch accent typically associated with contrastive or narrow focus has a higher predicted prominence rating than H* (log-odds diff. 0.8, $z = 3.2$, $p = 0.001$). For French, the High initial accent (Hi), probabilistically associated with contrastive or narrow focus, has lower predicted prominence than non-focus marking L*, H* (log-odds diff. −0.8, $z = -2.9$, $p = 0.003$). For Spanish, the early peak accent L+H* is associated with contrastive or narrow focus only indirectly as the accent that is typical in nuclear position, and has higher predicted prominence than the non-focal, prenuclear accent L+>H* (log-odds diff. 0.6, $z = 2.6$, $p = 0.01$).

5. Discussion

A real-time prominence rating task was performed by speakers of English, French and Spanish, using the method of Rapid Prosody Transcription with speech excerpts from conversational speech in the participants' native language. The resulting binary (0, 1) prominence ratings for content words were submitted to logistic mixed-effects regression to assess the role of acoustic and non-acoustic contextual factors on participants' prominence ratings. The speech excerpts were also prosodically labeled by trained annotators (the authors) using the ToBI systems for English, French and Spanish, and these labels were compared with the average prominence ratings of each word from all participants, for each language separately.

Eight factors were entered as predictors of prominence ratings. Two factors were explicitly manipulated in the design of the study. The **Language** factor refers to the language of the speech materials, which was always matched to the native language of the participant (English, French, or Spanish). **Instruction** refers to the criteria the participants were instructed to use when rating prominence (acoustic or meaning-related). The remaining six factors related to properties of the words that were rated for prominence: **Word Frequency**, **POS**, prosodic **Boundary**, **Phonerate**, **Intensity** and **Max F0**. As the speech materials were selected from existing databases of spontaneous, conversational speech in each language, and include samples from multiple speakers, the six word-level factors were not controlled for the purpose of this study. The model includes two-way interactions to test for differences in the effect of instruction across the three languages, and to test for differences in the six word-level factors across the languages and instructions. There are also three-way interactions

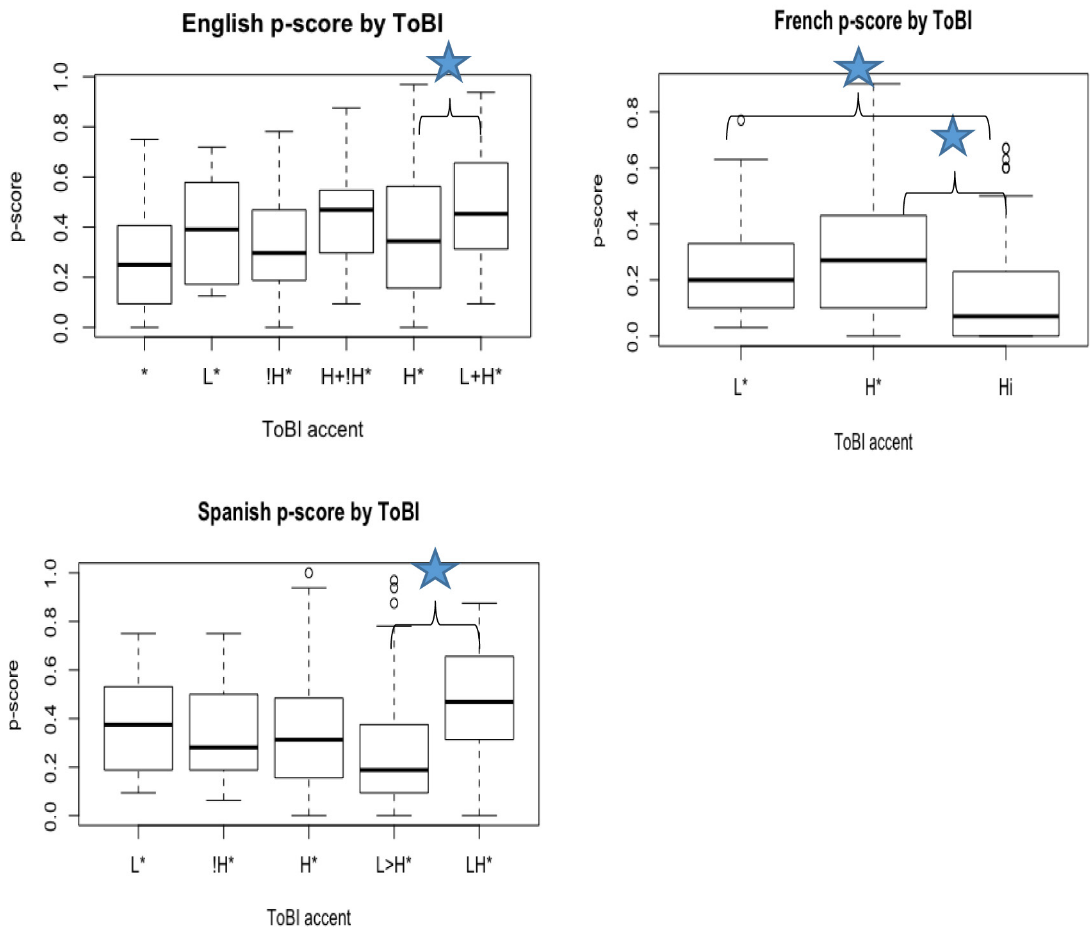

Fig. 7. Distribution of empirical p-scores for words grouped by their ToBI pitch accent label, by language. The accents most commonly associated with contrastive or narrow focus are L+H* for English and LH* for Spanish, and possibly Hi for French (see discussion in Section 2.1).

to test for differences in the effects of the word-level factors on specific pairings of language and instruction. Random effects are included for participant and item (=word). The full model accounts for 45% of the variance in participants' rating of prominence, and each of the main effects are significant. This measure of model fit indicates that we have captured some of the important factors that influence prominence rating in this real-time, auditory rating task, but also suggests that there may be additional factors not yet taken into account.

A primary objective of this study is to determine if there are differences in native listeners' perception of prominence across three languages whose prominence systems, as described in prior research, differ phonologically, phonetically, and in their relationship to information structure. The results are clear on this point, as the model shows a significant main effect of language ($\chi^2(2) = 12.18$, $p = 0.002$). Participants in all three language conditions are selective in rating words as prominent; the estimated intercept for the grand mean prominence rating is a negative value (−0.8012), indicating that words have a lower-than-chance likelihood of being rated prominent, where chance is 50%. See full model output in Appendix B (online supplement). Prominence rating is more liberal in Spanish, where model estimates show the highest likelihood of prominence, while prominence rating is most conservative in French, with lower model estimates. Beyond the main effect of Language, there is also a significant effect of Instruction ($\chi^2(1) = 335.08$, $p < 0.001$), which means that overall, participants rated prominence differently depending on the criteria they were instructed to attend to. Here also we find differences among the three languages. The significant interaction of Language and Instruction ($\chi^2(2) = 49.98$, $p < 0.001$) points to differences in how the task instruction influenced prominence ratings across the languages. These differences, shown in Fig. 1, show that prominence ratings in French and Spanish differ markedly for ratings based on acoustic vs. meaning criteria, while in English, there is much less of a difference. In other words, acoustic prominence and meaning-related prominence appear to converge on the same prominence rating for many more English words compared to French or Spanish words.

Other differences in prominence ratings among the three languages are seen in the effects of the word-level factors, and their interaction with the rating instruction. These effects are discussed below in relation to the experimental hypotheses from Section 2.4.

Hypothesis 1, concerning effects of word-level acoustic prominence, is confirmed in the significant main effects of the acoustic predictors: Phonerate ($\chi^2(1) = 75.14$, $p < 0.001$), Intensity ($\chi^2(1) = 11.46$, $p = 0.001$), and Max F0 ($\chi^2(1) = 14.81$, $p < 0.001$). The model estimated effects (Table 3) are uniform in the predicted direction in each language: Words are more likely to be rated as prominent if they have lower phonerate (i.e., are slower, with longer duration), higher intensity or higher peak F0. Among the acoustic factors, the effect of phonerate is the strongest in all three languages, based on the model-estimated effects (see Fig. 2). There were no significant differences between the languages in the effects of acoustic factors; the two-way interactions of acoustic factors with Language were not significant. This finding suggests that, setting aside effects due to task instruction and non-acoustic factors, acoustically defined prominence is perceived similarly in English, French and Spanish.

From **Hypothesis 2** we predict that the perception of a prosodic boundary following a word will boost the likelihood that the same listener will rate that word as prominent. Recall that this hypothesis is based on the fixed (French, Spanish) or frequent (English) occurrence of the obligatory nuclear prominence in phrase-final position. This hypothesis is confirmed in the main effect of Boundary ($\chi^2(1) = 119.46$, $p < 0.001$). Here also we fail to observe predicted differences between the languages. Despite the more consistent occurrence of phrase-final nuclear prominence in Spanish and French, the effect of boundary on prominence rating in those languages is not stronger than in English. Instead, we observe a remarkably similar and strong effect of perceived boundary on prominence ratings across the three languages, pointing to a common, structurally determined perceptual prominence: The phrase-final word is likely to be perceived as prominent in English, French and Spanish, alike. This effect holds independently of the effect of phonerate, so it can't be attributed as an effect of final lengthening. Rather, the boundary effect is evidence of top-down, expectation-driven processing in prominence perception.

Hypothesis 3 states an effect of usage (token) frequency in lowering the likelihood of prominence rating, which is confirmed by the model in the significant main effect of Word Frequency ($\chi^2(1) = 190.63$, $p < 0.001$). This effect is independent of the expected acoustic correlates of word frequency, and is another example of top-down, expectation-driven processing in prominence rating. In other words, listeners expect a low-frequency word to be prominent, perhaps because of its relatively high information value, and are biased to perceive such as word as prominent independent of its acoustic prominence. Word frequency effects on prominence ratings were predicted to be common across languages, reflecting shared mechanisms of speech processing and lexical encoding, yet there was a significant interaction of Language with Word Frequency ($\chi^2(2) = 13.61$, $p < 0.001$). This interaction effect is due to a much stronger effect of word frequency on prominence ratings for English. Though this finding of a stronger word frequency effect for English is not anticipated based on prior work, it is perhaps not surprising given the greater role of prominence in signaling information structure in the language, compared to French and Spanish.

Hypothesis 4 predicts an effect of POS on prominence ratings, which is confirmed as a main effect ($\chi^2(3) = 95.84$, $p < 0.001$).[11] We predicted POS effects to be similar among the three languages, which is confirmed by the absence of a significant interaction of Language and POS. The estimated effects of POS categories by language, as shown in Fig. 3, show that nouns and adjectives are more likely to be rated as prominent in each language, when all other factors are held constant, while verbs have a uniformly lower likelihood of prominence rating. Adverbs are more variable in prominence ratings across the languages.

The prominence rating task was administered under two instructions, one defining prominence in relation to its acoustic correlates, and the other defining prominence in terms of a word's contribution to utterance meaning. By manipulating the task instruction we hoped to test the strength of association between acoustic prominence and prominence related to word and discourse meaning. **Hypothesis 5**, stating the effect of task instruction on prominence rating, is confirmed by the main effect of Instruction, already noted. We predicted that the effect of instruction would be weaker for English, where information structure plays a strong role in the assignment of prominence and pitch accent, compared to French and Spanish, where prominence is only weakly associated with information structure. This prediction is borne out in the significant interaction of Language and Instruction ($\chi^2(2) = 49.98$, $p < 0.001$), and in the difference in the model estimated grand means of prominence rating (log-odds) in the two instruction conditions, as shown in Fig. 1. The difference in prominence rating under acoustic vs. meaning instruction in English is much smaller than in French or Spanish. The different effects of Instruction among the three languages are also reflected in how Instruction modulates the effects of the individual word-level factors on prominence ratings. The effects of the word-level factors on prominence ratings vary depending on the instruction, with the biggest difference (Acoustic-Meaning) found for prominence ratings in French and Spanish (Fig. 4).

Further evidence of the effect of task instruction is seen in the analysis of model fit, comparing the prominence ratings predicted by the model and the empirical mean of those ratings (p-scores). As shown in Fig. 5, there is only a weak relationship between the predicted p-score and actual p-score of each word for prominence ratings in French and Spanish under the meaning instruction. The relationship is stronger for prominence ratings under acoustic instructions for all three languages, and English stands out for the similarity of the model predictions under acoustic and meaning instructions. These findings indicate that while the factors included in our model are doing fairly well in modeling prominence as an acoustic (and auditory) phenomenon, the model is less successful with meaning-based prominence in French and Spanish. The model's relatively greater accuracy in predicting meaning-based prominence in English is evidence that prominence ratings are very often the same under both acoustic and meaning criteria in that language. Simply put, a word that sounds prominent in English is also very often prominent in relation to utterance meaning, and vice-versa. The visualizations in Fig. 5 reveal that this correspondence is much weaker for Spanish, and appears tenuous at best for French. The connection between acoustic- and meaning-based criteria for prominence may differ between these languages in production as well. The L1 English speakers studied by Smith, Erickson, and Savariaux (in press) produced articulatory and acoustic

[11] A similar, strong effect of POS on prominence ratings from an RPT task in English is reported by Wagner et al. (in press), and by Goldman, Auchlin, Roekhaut, Simon, and Avanzi (2010) for French.

cues to prominence other than f0 less consistently than did the L2 English speakers (whose L1 was French). The authors suggest that this may be because of the tighter link between acoustic- and meaning-based prominence in English than in French.

The statistical model gives some insight into which factors are responsible for the disparity in model predictions disfavoring the French-Meaning and Spanish-Meaning conditions. There were significant three-way interactions between Language, Instruction and the word-level factors Word Frequency, POS, Boundary, and Phonerate (Table 3). Comparing the effect estimates displayed in Fig. 2, it is evident that for French, acoustic factors contributed less overall compared to English and Spanish, and in the French-Meaning condition in particular, several factors played a smaller role than in English-Meaning or Spanish-Meaning conditions. To summarize these findings, while many word-level factors have a similar influence on prominence ratings in English, French and Spanish, the effects of these factors varies depending on whether prominence is defined acoustically or in relation to utterance meaning. The instructions governing criteria for prominence rating have a bigger impact on Spanish and especially French, and this difference rests largely on the weaker role of word-level factors in prominence ratings under meaning conditions. These results for the effect of Instruction, and its interaction with word-level factors, are in line with the predictions from Hypothesis 5.

We turn next to the comparison of prominence ratings from participants, and ToBI pitch accent labels. **Hypothesis 6** predicts that ToBI pitch accent distinctions (nuclear, prenuclear, unaccented) will be reflected in the RPT prominence ratings from untrained participants in each language. This prediction is confirmed in two ways. First, the empirical mean prominence ratings (p-scores) are higher for accented words than for unaccented words, and are higher for nuclear accented words than for prenuclear accented words (Fig. 6). Second, model-predicted prominence ratings are also significantly different according to the ToBI accent label of a word, and the difference goes in the expected direction: Nuclear > Prenuclear > Unaccented (Table 5).

Hypothesis 7 concerns the relationship between participants' prominence ratings and the type of ToBI pitch accent, and predicts that words with pitch accents typically associated with contrastive or narrow focus are more likely to be rated as prominent. This hypothesis was confirmed for English L+H* and also for Spanish LH* as pitch accents associated with contrastive or narrow focus (Fig. 7). Comparing the contrastive or narrow focus-marking rising pitch accents in English and Spanish with other rising accents (English H* and Spanish L>H*), we find a seemingly surprising result. The contrastive or narrow focus-marking pitch accent appears to boost prominence rating in Spanish more than it does in English. We predicted the contrastive or narrow focus-marking pitch accent to have a stronger effect on prominence ratings for English, due to the fact that pitch accent may be the sole mark of focus in English, whereas Spanish (and French) typically employ syntactic means to express focus. An explanation could be that, given the possibility of rearranging word order in Spanish, the use of LH* on a non-phrase-final word may make this word particularly prominent. The relatively weak effect of L+H* on prominence ratings in English calls for a closer examination of the production evidence in future research. We had a weaker prediction for an effect of focus-marking pitch accent on prominence ratings in French. German and D'Imperio (2016) reported a probabilistic association between the AP-initial Hi accent and contrastive focus, which was the basis for comparing the predicted prominence of Hi with H*and L* (the other pitch accents in the French ToBI inventory). But the AP-initial Hi accent is not restricted to occur on contrastively focused words, so at best there is only a subset of Hi accents associated with contrastive focus. Our results show a clear difference in French prominence ratings by ToBI accent type, but with the AP-initial accent (Hi) as the least likely accent type to be rated as prominent. It appears that the probabilistic association of Hi with contrastive focus was not sufficient to confer extra perceptual salience to the accented word.

Our findings from the English prominence ratings can be compared to those reported by Bishop et al. (in press), who analyze prominence ratings from an RPT task similar to our acoustic rating task. Overall, there are striking similarities in the role of acoustic and non-acoustic factors on prominence ratings between these two studies. The same acoustic parameters (F0, intensity, duration) were significant predictors of variation in the likelihood of prominence rating in both studies, although with differences in the relative strength of these three acoustic parameters, which may relate to differences in the speech samples (from 16 speakers for our study, and 1 speaker for Bishop et al.'s study). Bishop et al. also report similar effects of word frequency and (prosodic) boundary on prominence ratings, and a distinction in prominence ratings based on accent status (nuclear, prenuclear, unaccented) and the ToBI accent label of a word. A novel finding from their study is that non-acoustic factors have a greater influence on prominence ratings for words that are unaccented in the ToBI annotation. These are very interesting findings that resonate with our view of prominence as a complex construct that varies in degree in relation to sound, structure, and information (see also Wagner & McAuliffe, in press).

6. Conclusion

Phrasal prominence is a core feature of current phonological theory. Prominence is realized on the word that is the structural head, or nucleus, of the prosodic phrase, and additional prominences are commonly reported in prenuclear position. Yet languages vary in the phonological patterning of prominence, in its phonetic expression, and in the association of prominence with the information structure properties of a word, such as contrastive or narrow focus. In light of such differences, this paper asks whether in languages with different prominence systems, native listeners perceive prominence similarly in relation to acoustic cues and non-acoustic contextual factors. We report on a study of prominence perception in English, French and Spanish—three languages whose prominence systems differ along phonological, phonetic, and pragmatic dimensions. Prominence ratings were obtained from 90 untrained listeners using the method of Rapid Prosody Transcription. Participants rated prominence in samples of conversational speech in their native language (English, French or Spanish) under two task instructions, which defined prominence in terms of acoustic properties or in relation to utterance meaning. Logistic regression models tested the role of task instruction, word-level acoustic cues, and non-acoustic contextual factors in predicting the binary prominence ratings of individual listeners, and examined differences among the three languages in the effects of these factors on prominence ratings.

The results provide clear evidence that acoustic measures related to timing, loudness and pitch play a similar role in cueing prominence in all three languages. Words with slower tempo (longer duration), increased intensity and higher peak F0 are more likely to be rated as prominent in each language. This finding points to a common element of signal-based processing that appears insensitive to differences among the languages in how the acoustic cues pattern in relation to prominence (cf., Zahner, Kutscheid, & Braun, in press, who argue that differences among languages may result in differences in how speakers interpret the relation between f0 peaks and prominence). There is also a common structural effect on prominence rating. In all three languages, a word is more likely to be rated as prominent if it precedes a prosodic boundary. This structural prominence supports claims from earlier work that the phrase-final nuclear prominence is perceptually stronger, and since the effect of boundary is independent of word-level acoustic factors, it provides evidence of top-down, expectation-driven prominence processing.

POS also has a common effect in all three languages, with prominence rating more likely for nouns and adjectives and less likely for verbs. This finding may reflect the greater tendency for nouns and predicate adjectives to occur in phrase-final (nuclear) position, but it suggests a possible role for prominence in the expression of the referential status of words, i.e., information status. Nouns and adjectives are typically the words used to introduce new referents into the discourse, and recent findings in German indicate that prominence encodes referential status through pitch accenting (Baumann & Riester, 2013). It is possible that the same association may be reflected in the increased likelihood for nominals to be rated as prominent in our data. This hypothesis can be explicitly tested in future work.

Differences among the languages arise in the effect of word frequency on prominence ratings. In all three languages, high-frequency words are significantly less likely to be rated as prominent than words with lower usage frequency. This effect is also independent of word-level acoustic factors, and is therefore further evidence for top-down, expectation-driven prominence processing. The word frequency effect is stronger for English; possibly due to the stronger pragmatic function of prominence in that language. Other differences among the languages emerge when we consider how prominence ratings are affected by the task instructions that define prominence in terms of acoustic vs. meaning criteria. In all three languages, listeners are more conservative in rating words as prominent under acoustic instructions, but the difference between acoustic and meaning-related prominence is much smaller for English than for French or Spanish. This difference was predicted, and can be traced to the finding that French and Spanish prominence ratings under meaning criteria are less successfully predicted by the factors we have examined as perceptual cues. We think it is highly likely that the addition of syntactic factors to our model, to capture word order variation that relates to information structure, will yield better predictions of meaning prominence in French and Spanish.

A final finding from this study is that the prominence ratings from untrained listeners in our study have a significant correspondence with ToBI pitch accent labels in each language (a finding also reported by Bishop et al., in press). Specifically, the contrast between unaccented, prenuclear and nuclear accents is reflected in corresponding differences in the empirical prominence ratings in each language. In addition, model predictions support a distinction between those pitch accents typically associated with contrastive or narrow focus and other pitch accents in English and Spanish, with contrastive or narrow focus-marking pitch accents more likely to be rated as prominent. This difference aligns with our expectation, as contrastive or narrow focus-marking pitch accents have greater acoustic prominence and are thus predicted to have greater perceptual salience. The status of a contrastive focus-marking pitch accent in French is controversial, and we failed to find an effect for the AP-initial pitch accent that is sometimes described as a focal accent. Our data provide no evidence of a role for prominence in conveying contrastive focus in French.

This study contributes novel evidence for common elements in signal-driven prominence processing across languages with phrasal prominence systems that differ in both form and function. It also points to a fundamental difference in how listeners perceive prominence in English, where phrasal prominence is strongly associated with utterance meaning, and French and Spanish, where phrasal prominence is perceived more narrowly on structural and acoustic grounds.

How does this study inform our understanding of prominence as a phonological and phonetic phenomenon? In metrical phonology, prominent words are categorically and structurally distinct from non-prominent words, and nuclear prominence is categorically distinct from prenuclear prominence (see also Wagner and McAuliffe (in press) for a more extensive discussion of the difference between structural prominence and focal prominence). On the other hand, acoustic prominence is not so obviously categorical. In the present study, participants rated prominence categorically, but under acoustic and meaning criteria that vary continuously. Are these prominence ratings phonological or phonetic? We see evidence for both in our statistical findings. We have an effect of categorical phonological structure on prominence ratings—a prosodic boundary increases the likelihood of prominence rating in each of the languages investigated here. On the other hand, the likelihood of prominence rating varies continuously in relation to continuous variation in acoustic cues—the greater the acoustic enhancement of a word, the more likely it is to be rated as prominent.

We suggest that these findings can be accommodated in a probabilistic model of prosody, i.e., a model where the categorical prosodic features are probabilistically related to the syntactic, semantic and pragmatic properties of a word. There is a probabilistic association between pitch accents and information structure conditions in speech production (Chodroff & Cole, 2018, 2019; see also Baumann & Grice, 2006; Cangemi, Krüger, & Grice, 2015 for German; Cangemi & Grice, 2016 for Neopolitan Italian; and Luchkina & Cole, 2017 for Russian), or between prosodic boundaries and constituent boundaries in syntactic or discourse representation (Calhoun, 2006; Peppé, Maxim, & Wells, 2000). Variability in the production of prosodic features means that the top-down influence of factors like information structure and syntactic context on listeners' (expectation-driven) perception of pitch accent and prosodic boundaries will be similarly variable. For example, a listener who infers a prosodic phrase boundary following a word that is also final in a major syntactic phrase may do so probabilistically, matching the likelihood of the prosodic boundary at that location in speech production. Many of the contextual factors found to predict prominence perception in this study, including the presence of a following prosodic boundary, are in fact probabilistically associated with prosody in speech production. In short, variation in the perception of a prosodic feature may arise due to top-down processing, as influenced by contextual features that are themselves probabilistically associated with prosody in speech production. A listener may perceive prominence on a word upon detecting the presence of a contextual factor associated with prominence in one instance, and may not perceive prominence in the presence of the same contextual factor in another instance. Similar differences may arise between listeners rating prominence on the same word, and such differences may lie behind the variable influence of contextual factors on prominence ratings in our study.

Our finding of a gradient effect of acoustic cues on the likelihood of prominence perception may suggest that prominence is itself a gradient feature, possibly reflecting gradations in metrical prominence among constituents in the same utterance (Wagner & McAuliffe, in press). Yet we must also recognize the possibility that the gradient effect of acoustic cues results from differences among listeners in how they process and/or interpret acoustic cues to prominence. Individual differences in prominence perception are reported for the English data analyzed here (Roy et al., 2017), and in a similar study of German (Baumann & Winter, 2018), each showing robust individual differences in the weighting or thresholding of cues in prominence rating. These studies, along with Bishop (2012), Bishop et al. (in press) and Wagner, Cwiek, and Samlowski (in press) also find evidence of individual differences in how non-acoustic, contextual factors influence prominence perception. At the population level, gradient variation in the likelihood of prominence perception could arise from individual differences in the selection of acoustic parameters as prominence cues, or in the cue value that marks a threshold between prominent and non-prominent words. At the level of the individual listener, it is possible that listeners perceive a categorical distinction in prominence and also perceive within-category variation, as has been shown for segmental contrasts (McMurray, Tanenhaus, & Aslin, 2002).

Ultimately, the evidence reported here is not conclusive on the status of prominence distinctions as categorical or gradient in phonological representation. Watson (2010) reviews a number of studies and argues that prominence is not categorical in production, but varies continuously with discourse structure. Kimball (2018) investigates the categorical status of the accented/unaccented distinction in English and offers evidence for a hybrid model of prosodic representation. In a series of experiments testing perceptual discrimination of the accented/unaccented distinction, Kimball shows that listeners perceive a categorical distinction between accented and unaccented words, while also perceiving within-category variation in pitch and duration as cues to accent status. We see a need for further research to examine the granularity of prominence distinctions in mental representation in relation to the patterns of variation in the production and perception of prominence, and the relationship between the two.

This study is the first to our knowledge to offer a direct comparison of the factors influencing prominence perception across languages, but we acknowledge that there is much more work to be done. For instance, we have not yet examined whether listeners would give consistent ratings of the same speech sample, with the same rating criteria, in rating tasks performed on different days. One promising line of research concerns prominence ratings with language learners, or with proficient L2 speakers of a language, for which the RPT method of prosodic annotation is already shown to be successful (Pintér, Mizuguchi, & Tateishi, 2014; Smith & Edmunds, 2013). It would also be interesting to compare prominence ratings of intelligible speech with ratings of a delexicalized, speech-like signal with intact acoustic-prosodic modulations from a source recording to determine the contribution of linguistic factors related to the lexical, syntactic and semantic content of an utterance.[12] We think these research questions, and others, can be fruitfully explored using the prominence rating method employed in the current study.

Acknowledgments

This work was supported by a grant to Jennifer Cole, José I. Hualde and Caroline Smith from the US National Science Foundation, USA [BCS-1251343 and BCS-1251134]. We thank Joseph Roy and Suyeon Im at the University of Illinois for help with data analysis, and David Escudero and Valentín Cardeñoso at the University of Valladolid for offering lab space and assistance with recruiting participants for our experiment in Valladolid, Spain. Caroline Smith thanks the Laboratoire Dynamique du Langage for assistance of various kinds during her working visits to Lyon. We thank the Guest Editors and anonymous reviewers for their insightful comments and helpful suggestions, and audiences at the Speech Prosody 8 and Laboratory Phonology 14 conferences, and at the University of Cologne and the University of Illinois for comments on earlier presentations of this research.

Appendix A. Methods

In the following sections, the word "schools" in "i do like columbus public schools but yknow i" from English stimulus utterance s10-1 will be used as an example.

1. TextGrids

The Buckeye corpus (Pitt et al., 2007) from which the English stimulus utterances were taken is already hand-aligned at the word and phone level in Praat (Boersma & Weenink, 2014). The syllable tier was added using a Python script written by one of the authors. The Glissando Spanish informal-dialogue sub-corpus (Garrido et al., 2013) from which the Spanish stimulus utterances were taken is force-aligned at the word, phone and syllable level, but the alignment is imper-

[12] We thank an anonymous reviewer for this suggestion.

fect. For this reason, the TextGrids were manually corrected. The Corpus de Français Parlé Parisien (Branca-Rosoff, Fleury, Lefeuvre, & Pires, 2012), henceforth CFPP, from which the French stimulus utterances were taken is orthographically transcribed but not annotated. TextGrids for the CFPP stimulus materials were created using the EasyAlign force aligner (Goldman, 2011) and checked and manually corrected by hand. A TextGrid for the example English utterance is given in Fig. A below.

2. Measurements

All measurements were taken at the word level. For each language, measurements were stored in a CSV file where each row corresponded to a word in the stimulus materials transcribers responded to, and each column represents a measurement of that word. So for English, the CSV has 931 rows, one for each word, and for French and Spanish there are 1079 and 897 rows respectively. Each of these three CSVs has 130 columns, which are described in the following subsections.

2.1. Item identification variables (3)

Three variables serve to identify items. The first is the word itself (i.e. “schools”). The second is the stimulus utterance in which the word occurs (in the case of “schools” the stimulus utterance was s10-1). The third is a unique item identifier, which was coded as a language-specific letter followed by a number corresponding to the order in which words were evaluated by the transcribers. So the word “schools” would be item e250 because it is the 250th word which the English transcribers responded to.

2.2. Transcriber response variables (120)

2.2.1. Prominence response

Each individual transcriber’s prominence marking of an individual word under a specific instruction was coded as 0 if the word was not marked, and 1 if the word was marked. So within each language, each item has two prominence responses from each transcriber, one under the acoustic instructions and one under the meaning instructions. These responses were stored as variables labeled with a language-specific letter, transcriber number, and instruction letter.

For example, variable “e11.pm” for item e250 is coded as 1 since the 11th English transcriber marked the word “schools” as prominent under the meaning instructions, while “e11.pa” is coded as 0 because the 11th English transcriber did not mark “schools” as prominent under the acoustic instructions.

2.2.2. Boundary response

Similarly, each transcriber’s boundary marking after an individual word under a specific instruction was coded as 0 if no boundary was marked after the word, and 1 if a boundary was marked after the word. So within each language, each item has two boundary responses from each transcriber. These were stored as variables in the same way as the prominence responses, except with “p” replaced with “b.”

For example, variable “e11.bm” for item e250 is coded as 1 because the 11th English transcriber marked a boundary after “schools” under the meaning instructions, while “e11.ba” is coded as 0 because the 11th English transcriber did not mark a boundary after “schools” under the acoustic instructions.

2.3. Meaning variables (2)

2.3.1. Word frequency

Raw word counts were taken from the Buckeye corpus for English, from the Glissando Spanish informal-dialogue subcorpus for Spanish, and from the Lexique 3.8 database (New et al., 2001) for French. For each item, the number of occurrences of that word in its corpus was divided by the total number of words in that corpus, and then the natural logarithm was taken.

For example, the word “schools” occurs 5 times in the Buckeye corpus and there are 35,009 words total. The value of word frequency for item e250 is thus ln(5/35009) = −8.9.

2.3.2. Part of speech

For English and Spanish, part of speech tagging was done by hand, considering the following labels: adjective, adverb, complementizer, conjunction, subject-verb contraction, negative contraction, polarity, determiner, discourse marker, disfluency, noun, preposition, pronoun, auxiliary verb, main verb. For French, part of speech tags were created using the Stuttgart TreeTagger (Schmid, 1994), hand-corrected, and then

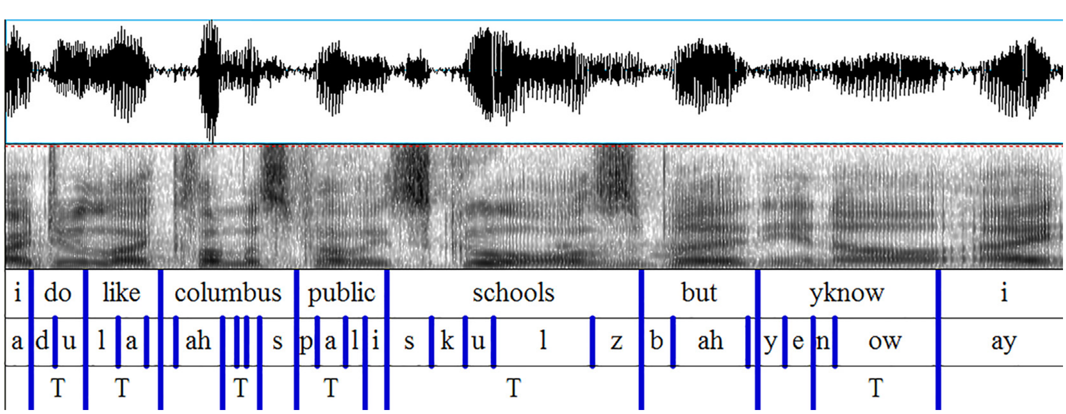

Fig. A. Example TextGrid aligned at the word, phone, and syllable level.

simplified to the system used for English and Spanish (for example, verb tense was marked by the Stuttgart TreeTagger, so we simplified both “verb:future” and “verb:present” to “main verb”). For example, item e250 “schools” is coded as a noun.

2.4. Acoustic variables (5)

2.4.1. Phonerate

First, a phonerate curve for each utterance was created using a Praat script written by one of the authors. The script used a fixed window length of 500 ms and a time-step of 10 ms. For each window that did not contain silences, the local phonerate was calculated using Pfitzinger (1998) RateLR formula.

Pfitzinger (1998) used this formula on syllablerate and phonerate and created a speechrate curve by combining these two, and used the inverse of the resulting function to normalize durations, thus removing the variance in duration caused by speechrate fluctuation. We instead used the mean and variance of these data points for each utterance to perform a normalization procedure on a different set of data computed at the word level. The phonerate was calculated for each word by dividing the number of phones in the word by the word’s duration. The resulting phonerate was z-scored using the mean and variance of the data points which comprise the phonerate curve for the entire utterance in which the word occurred. The results of this z-score thus indicate how much the word is lengthened or shortened relative to the utterance as a whole.

For example, there are 5 phones in the English word “schools” /skúlz/. The word is 0.536 seconds in duration, meaning the raw phonerate for the word is 5/0.536 = 9.3 phones/second. The word occurs in the English stimulus utterance s10-1. The phonerate curve created for this utterance with a 500-ms window length and 10-ms time-step has 1423 points (see Fig. B below). The mean of these 1423 phonerate values is 15.2 phones/second (solid line in Fig. B) and the standard deviation is 3.6 phones/second (1 standard deviation above and below the mean are dashed lines in Fig. B). The normalized phonerate value for item e250 is thus (9.3–15.2)/3.6 = −1.6.

2.4.2. Intensity

The mean intensity in dB of the stressed vowel in each word was extracted from the WAV files using a Praat script written by one of the authors. For Spanish words with no lexical stress, the first vowel was used (e.g. /e/ in the conjunction *pero* ‘but’). For French, the last non-schwa vowel in the word was analyzed (e.g., the /i/ in existe ‘exists’). The script created an Intensity object for each file using a minimum pitch of 100 Hz with a time-step of 1 ms and the mean pressure subtracted. The intensity value for each word was normalized using the mean and standard deviation of the stressed vowels of a 5-word window. For the first two and last two words in each stimulus utterance, the window was shortened (e.g. the first word is normalized using a 3-word window consisting of the first three words in the stimulus utterance). The result is a measure of how much a word does or does not stand out from its neighbors in terms of intensity.

For example, the mean intensity of the stressed vowel in “schools” is 67.9 dB. The mean intensity on the stressed vowels in “columbus public” (previous two words) and “but yknow” (following two words) are 63.9 dB, 64.6 dB, 64.3 dB, and 58.9 dB respectively. The mean of these 5 values is 63.9 dB and the standard deviation is 3.2 dB. The 5-word normalized value for item e250 is thus (67.9–63.9)/3.2 = 1.2. The intensity curve for this 5-word normalization is shown in Fig. C below.

2.4.3. Pitch

2.4.3.1. Pitch processing. First, each file was viewed in Praat with gender-specific pitch ranges of 70–250 for males and 100–300 for females (Vogel et al., 2009). If there was evidence that the gender-specific ceilings were clipping pitch accents, or evi-

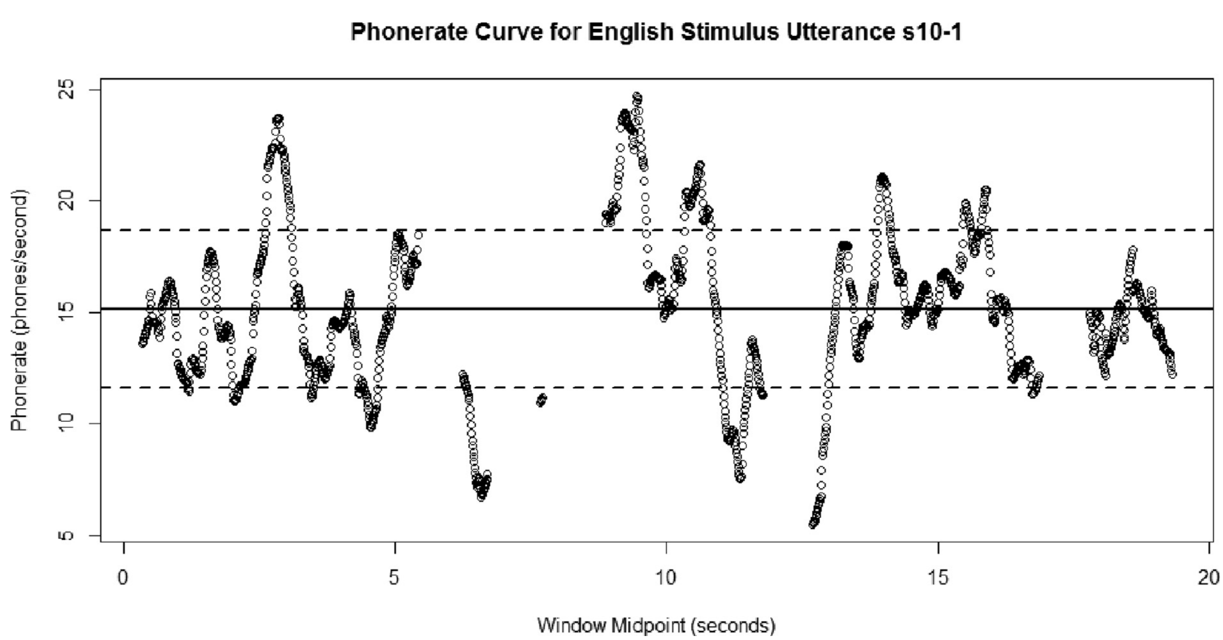

Fig. B. Phonerate curve for English stimulus utterance s10-1. The solid line represents the mean phonerate and the dashed lines represent one standard deviation above and below the mean.

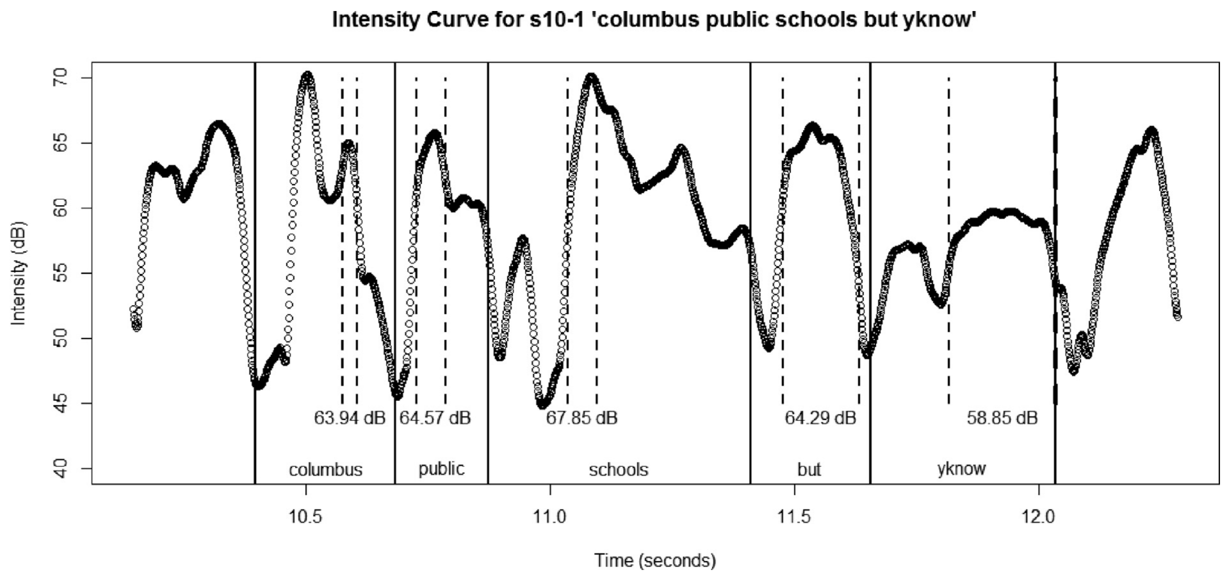

Fig. C. Example intensity calculation. Solid lines are word boundaries; dashed lines are stressed vowel boundaries. Intensity values are the mean for the stressed vowel.

dence that a speaker always spoke well above the gender-specific floor, the pitch ranges were manipulated to account for that speakers' idiosyncrasies. Using these speaker-specific pitch ranges, an auto-correlated pitch object was created for each file in Praat with a time-step of 1 ms and all other parameters left at default values. The individual pitch points from these pitch objects were written to CSV files in Hertz and processed in R (R Core Team, 2015). An R script written by one of the authors was then used check the pitch CSVs for halving and doubling. Halving and doubling were liberally defined as two consecutive pitch points less than 30 ms apart where the second point was less than or equal to 75% of the first value (halving) or greater than or equal to 150% of the first value (doubling). Instances marked as possible cases of halving and doubling were then examined in Praat and hand-corrected by doubling or halving the pitch contour respectively if a jump had occurred. After correcting for pitch jumps, each pitch file was median filtered with a filter width of 35. The reason the filter width is relatively high is due to the small time-step used in the creation of the files. It was chosen by comparing filters of varying widths and choosing the width that best dampened the effects of consonantal transitions without flattening pitch accents. The resulting filtered values were converted to semitones re 1 Hz (one of the options provided in Praat) by taking $12^{*}\log_2$(Hz-value).

2.4.3.2. Maximum F0. The maximum F0 of the stressed vowel in semitones was extracted from the pitch files for each word using an R script. For each word, the maximum F0 of the stressed vowel was normalized using the mean and standard deviation of the maximum F0 values of the stressed vowels in a 5-word window (with shorter windows for the first two and last two words as for the intensity normalization). The result is a measure of how much a word does or does not stand out from its neighbors in terms of the maximum F0 on the stressed vowel.

For example, the maximum F0 of the stressed vowel in "schools" is 91.9 semitones. The maximum F0 on the stressed vowels in "columbus public" (previous two words) and "but yknow" (following two words) are 88.2, 87.3, 86.6, and 83.7 semitones respectively. The mean of these 5 values is 87.6 and the standard deviation is 3.0. The 5-word normalized value for item e250 is thus (91.9–87.6)/3 = 1.5. The F0 curve for this 5-word normalization is shown in Fig. D below.

2.4.3.3. F0 range. First, the pitch points within the stressed vowel of the word which contained the minimum and the maximum F0 were identified. The minimum was subtracted from the maximum to obtain the F0 range in semitones. If the maximum occurred before the minimum, the range was multiplied by −1.

For example, the maximum F0 on the stressed vowel in "schools" is 91.9 semitones and this maximum occurs at a timestamp of 11.043 seconds. The minimum F0 on the stressed vowel in "schools" is 90.4 semitones and this minimum occurs at a timestamp of 11.097 seconds. Subtracting the minimum F0 from the maximum F0, we get a change of 1.5 semitones. Since the timestamps indicate that the maximum occurred before the minimum, meaning a falling pitch, the value was multiplied by −1 to obtain an F0 range value of −1.5 semitones for item e250. This range calculation is shown in Fig. E below.

2.4.3.4. F0 velocity. To obtain the velocity of the change in F0 over the stressed vowel, the F0 range described in Section 2.4.3.3 was divided by the absolute difference between the timestamps of the maximum and minimum F0.

For example, the stressed vowel in "schools" has an F0 range of −1.5 semitones and this fall occurs from 11.043 seconds to 11.097 seconds. The F0 velocity for item e250 is thus −1.5/abs(11.043–11.097) = −28.6 semitones/second. This is exemplified in Fig. E above.

3. Data cleaning

There were a total of 2907 items (931 English, 1079 French, 897 Spanish) prior to data cleaning.

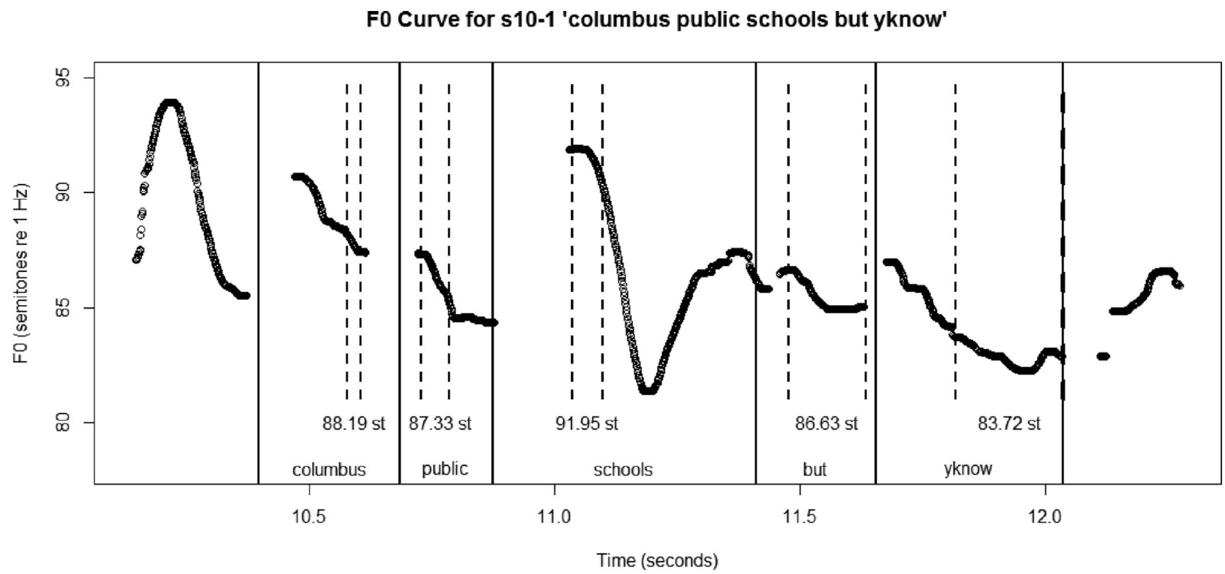

Fig. D. Example maximum F0 calculation. Solid lines are word boundaries; dashed lines are stressed vowel boundaries. F0 values are the maximum for the stressed vowel.

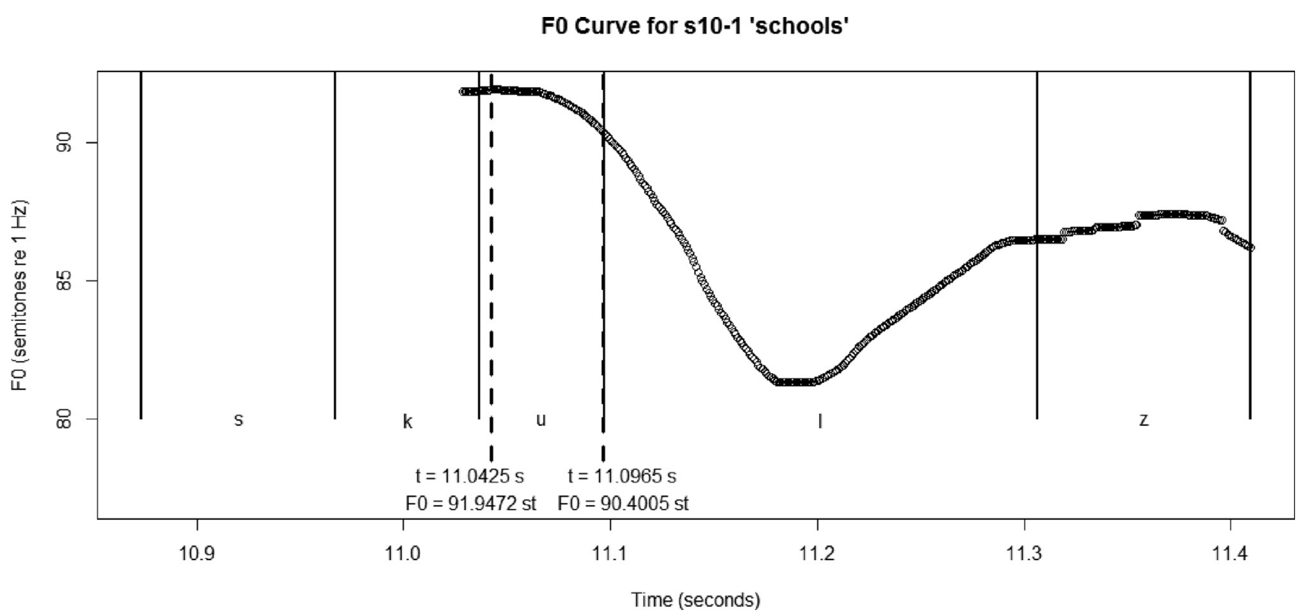

Fig. E. Example F0 range and velocity calculation. Solid lines are phone boundaries; dashed lines are the maximum and minimum F0 values for the stressed vowel.

3.1. Exclusion of English utterance s32-1

During pitch processing (Section 2.4.3.1), we found that the speaker in English stimulus utterance s32-1 had an abnormally extensive use of creaky voice, making his pitch measurements unreliable. For this reason, the entire utterance (67 items) was excluded from analysis, leaving 864 English items.

3.2. Exclusion of Spanish and French transcript mismatches

There were 10 items in the Spanish materials where there were mismatches between the audio and the transcript. For example, the discourse marker *o sea* has another variant lacking the *o*. The transcript several times had *o sea* but there was no evidence of the word *o* in the audio. Another example is *tenia* in the phrase *tenia que* with the verb conjugated in the transcript but in the audio it is actually *tener que*. Our pilot subjects did not report noticing anything odd about the transcripts, and the discrepancies we discovered were small and few, so we do not believe they affected the transcribers, but to be safe the items were excluded from analysis, leaving 887 of the 897 Spanish items. There were 5 mismatches of this type in the French transcripts which were also removed, leaving 1074 French items.

3.3. French contractions

Most contractions in the French speech samples (such as j'peux, 'I can') were transcribed as a single word, with no space between the contracted personal pronoun and the verb. However, there were 12 instances where these were printed as two separate words in the text presented to the transcribers in the LMEDS experiment. Transcribers thus had the ability to mark prominence on and boundaries after neither, either or both words, and there were 24 items in the spreadsheet where there should only be 12. We collapsed these items, taking the maximum prominence score of the pair for each transcriber, and the boundary score of the second item in the pair. So if a transcriber marked either *j'* or *peux* as prominent, then we coded that transcriber's prominence response to *j'peux* as 1. If the transcriber marked a boundary after *peux*, we coded that transcriber's boundary response as 1. The collapsing of these 24 items into 12 leaves 1062 French items.

4. Data exploration

4.1. Initial token counts

There are a total of 2813 items (864 English, 1062 French, and 887 Spanish).

4.2. Exclusion of function words from prominence analysis

First, 4 new variables were computed for each language by summing transcriber prominence and boundary responses within instruction for each item and dividing by 30 (the number of transcribers). The result is the proportion of transcribers who marked a prominence on a word or boundary after a word under a given instruction, ranging zero to one, termed "b-score" for boundary marking and "p-score" for prominence marking. So, for example, 7 English transcribers marked item e250 "schools" as prominent under the acoustic instructions, making the acoustic p-score for item e250 7/30 = 0.23, and 11 English transcribers marked a boundary after item e250 under the acoustic instructions, making the acoustic b-score for item e250 11/30 = 0.37. These variables (acoustic p-score, acoustic b-score, meaning p-score, and meaning b-score) are not the dependent variable used in regression analysis. Mixed effects logistic regression on each 0 or 1 response is used (described in Section 5). However, for exploratory data analysis, collapsing transcriber responses across items gives an approximation.

Defining nouns, main verbs, adjectives and adverbs as content words and all other parts of speech as function words, preliminary data analysis revealed that in all three languages, function words have a limited distribution, being highly unlikely to be marked as prominent under either the acoustic or meaning instructions. To exemplify this, Fig. F shows the probability density functions for p-scores separated out by language, instruction and status as content or function.

Because including the function words would more than triple the number of part of speech levels in regression (some of them with very few items), and could cause issues of separation with continuous variables (particularly word frequency), the 1667 function words (509 English, 630 French, 547 Spanish) were excluded from prominence analysis, leaving 1146 content words (355 English, 447 French, and 344 Spanish).

4.3. Exclusion of items with missing values

Of these content words, 5 English items, 15 French items, and 4 Spanish items had undefined F0 values. Examination of these items in Praat revealed that they were either too short for pitch analysis or were produced as devoiced or with creaky voice. They were excluded from further analysis, leaving 1122 items (350 English, 432 French, and 340 Spanish).

4.4. Checking assumptions of mixed effects logistic regression

Mixed effects logistic regression on prominence marking (each transcriber's 0 or 1) assumes a linear relationship between the log-odds of prominence-marking and the depen-

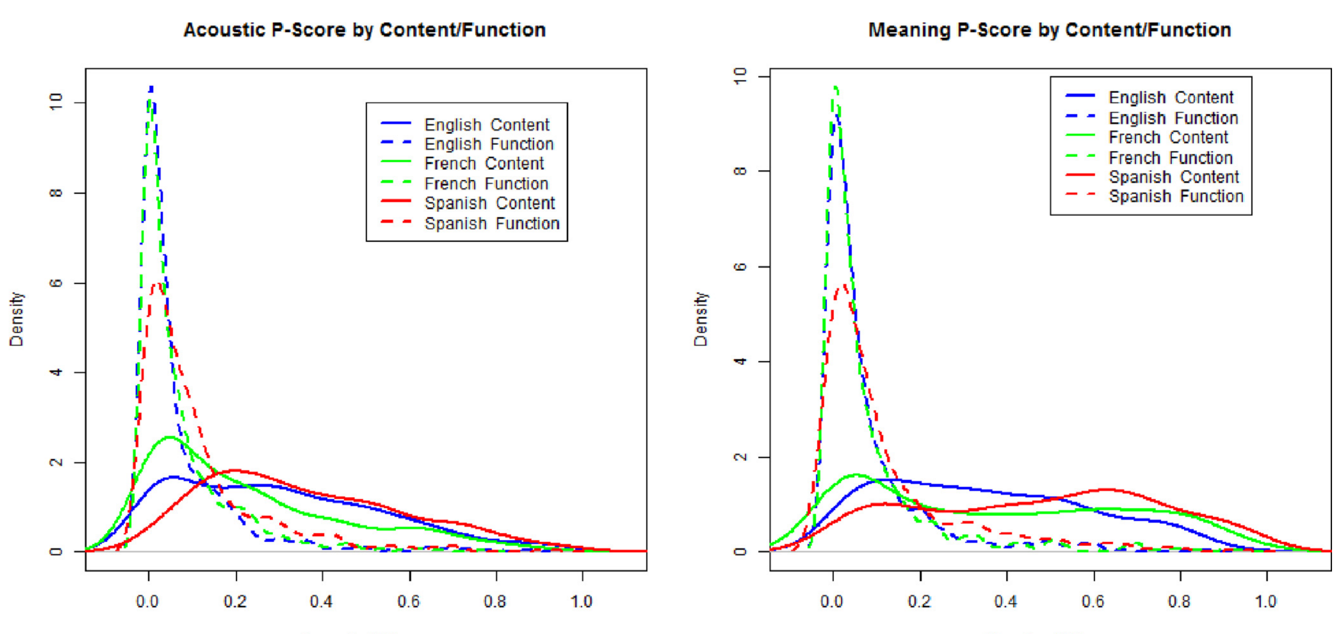

Fig. F. Probability density functions of content and function words' p-scores.

dent variables, and also assumes no gaps in the coverage of the continuous predictor variables. Gaps were visually checked from scatterplots as follows. Each item's p-score was converted to log-odds by ln(pscore/(1-pscore)) and each item's log-odds of prominence marking was plotted against the independent variables measured for the items. For a p-score of zero, the log-odds are negative infinity and for a p-score of one, the log-odds are positive infinity. Since the resolution of the p-score measure with 30 transcribers is 1/30, these p-scores of 0 and 1 were reassigned as 1/60 and 59/60 respectively for graphical purposes as well as for diagnosing model fit (see Section 6). Visual inspection of these scatterplots for the continuous predictor variables of **Max F0, Intensity, Word Phonerate**, and **Word Frequency** showed no gaps in the coverage of the continuous predictor variables over the range of p-scores.

4.4.1. Gaps in coverage

Gaps in coverage were found for F0 range and F0 velocity in all three languages, and for phonerate in Spanish. Fig. G below shows English F0 velocity and Spanish phonerate as an example.

A total of 5 English Items, 4 French Items and 4 Spanish items were excluded from analysis, representing only 1.2% of the 1122 items, leaving a final count of 1109 items for regression analysis (345 English, 428 French, and 336 Spanish).

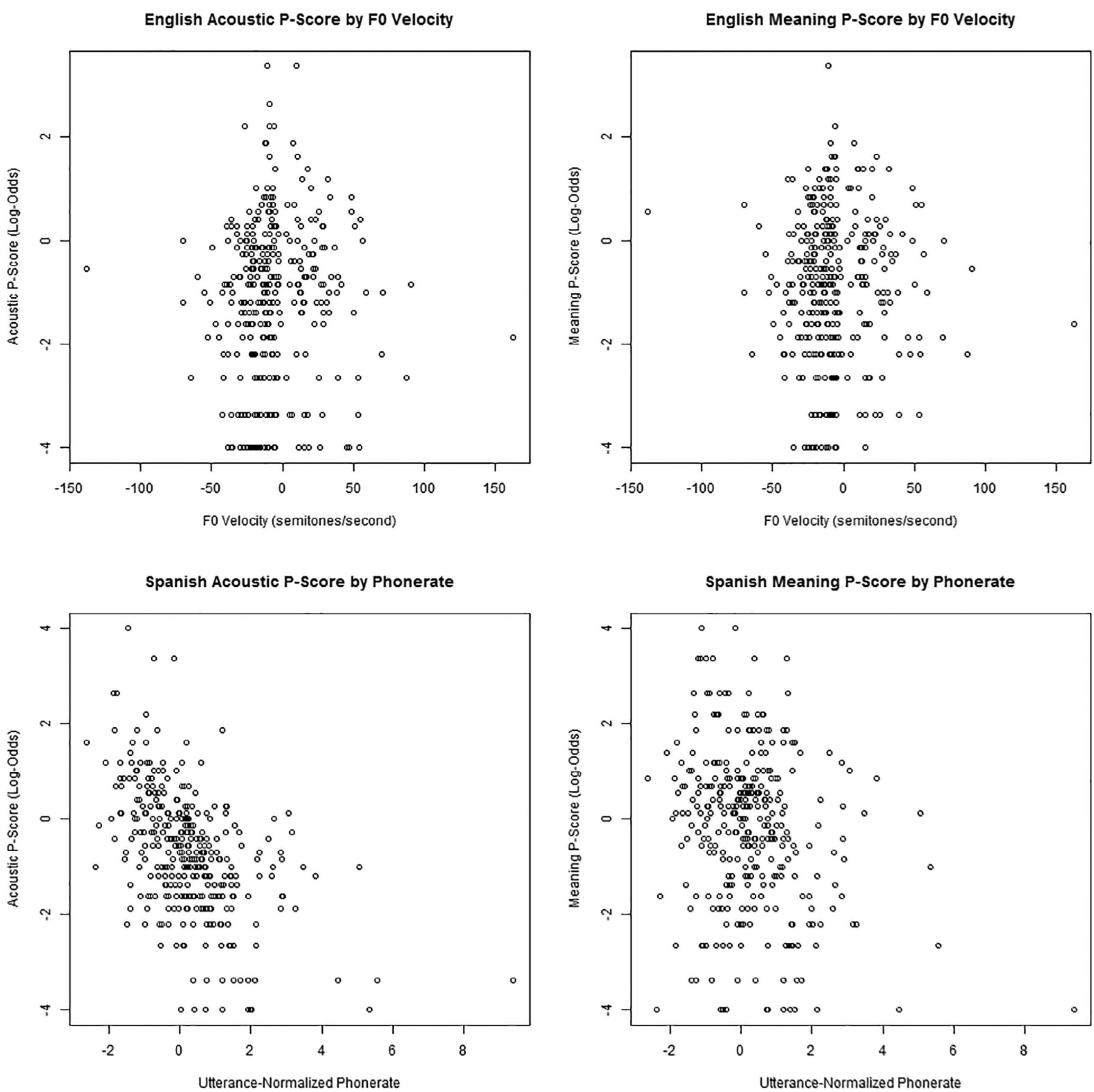

Fig. G. Examples of gaps in coverage and lack of F0 Range and F0 Velocity effect.

4.4.2. Exclusion of F0 range and F0 velocity

F0 range and F0 velocity showed no correlation with the log-odds of prominence marking under any language-instruction pair (exemplified with English F0 velocity in Fig. G). For this reason, they were not included in the regression analysis.

5. Statistical analysis

5.1. Data frame

The transcribers' prominence markings (0 or 1) for the 1109 content words were submitted to mixed effects logistic regression in R. An observation was defined as an individual transcriber's prominence and boundary markings of an item under a specific instruction. So for each item, 30 transcribers responded under 2 instructions, meaning 60 observations per item, and 66,540 observations total. Boundary marking was coded as each transcriber's 0 or 1 response to the item under the given instruction. Three additional factors were created to classify the observations: *language* with levels English, French and Spanish, *instruction* with levels Acoustic and Meaning, and *transcriber* with 90 levels, coded with a language-specific letter and unique transcriber number, resulting in a 66540 × 11 data frame, with the following columns: item, transcriber, language, instruction, p-mark (0 or 1), b-mark (0 or 1), log word frequency, part of speech (adjective, adverb, noun, or verb), utterance-normalized phonerate of the word, 5-word normalized mean intensity of the stressed vowel, and 5-word normalized maximum F0 of the stressed vowel in semitones. So, for example, the 11th English transcriber's responses to item e250 under the acoustic and meaning instructions would be coded as two observations:

obs	Item	Transcriber	Language	Instruction	P-Mark
6983	e250	e11	English	Acoustic	0
7328	e250	e11	English	Meaning	1

obs	B-Mark	Word Frequency	POS	Phonerate	Intensity	Max F0
6983	0	−8.9	Noun	−1.6	1.2	1.5
7328	1	−8.9	Noun	−1.6	1.2	1.5

The values for word frequency, part of speech, phonerate, intensity and max F0 are those which were exemplified in Section 2. The variables p-mark, b-mark, language, instruction, and transcriber are obtained from a decomposition of variables e11.pa, e11.pm, e11.ba and e11.bm as described in Section 2. There are 58 more observations for item e250, 2 for each of the other 29 English transcribers.

All factors were coded using sum contrasts, and all continuous variables were scaled in R prior to regression analysis, which allows straightforward interpretation of regression coefficients. Mixed effects logistic regression was run using glmer() in the lme4 package (Bates et al., 2014). The dependent variable is prominence marking (0 or 1). Fixed effects included in the model were the three way interaction between language, instruction, and each of the other six predictors (boundary marking, word frequency, part of speech, phonerate, intensity, and max F0), meaning 27 fixed effects in addition to the intercept, which represents the grand mean of the response variable, for a total of 54 fixed parameters. Random intercepts for transcriber and item were also included (the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013) was computationally infeasible given the size of the dataset and the number of random effects parameters each slope introduces). The likelihood ratio test was used to obtain *p*-values with mixed() function in the afex package (Singmann et al., 2015). Because there is a large number of levels of each grouping factor (90 transcribers and 1109 items) and the number of observations far outweighs the number of parameters, parametric bootstrapping is unnecessary (Barr et al., 2013) and would be too computationally demanding. Given the number of interactions included in the model, and the difficulty of interpreting these estimates directly, regression output was analyzed using the effects package (Fox, 2003).

6. Assessing model fit

Mixed effects logistic regression and logistic regression more generally do not have a straightforward way of measuring model fit such as R-squared and error plots like linear models do. Our goal in assessing model fit is to see how accurately we can predict the p-scores of the 1109 items, and to see if accuracy differs by language and instruction. To do this, the fixed effects were extracted from the model and used to predict the log-odds of each item being marked as prominent under each instruction, ignoring the random effects associated with transcriber and items, resulting in 2218 predictions. This tells us what the model predicts the probability of an item being marked as prominent is without being biased by the noise in our particular items and transcribers. Because boundary marking was included as a binary predictor coded with sum contrasts in the regression (b-mark of 0 coded as 1 and b-mark of 1 coded as −1), the average of the transcribers' responses was used. So, for example, if a boundary was marked after a word by 20 of the 30 transcribers, the regression coefficient for "no boundary marked" was multiplied by (1*10 + (−1)*20)/30 = −0.33.

To see if **the errors were normally distributed**, we subtracted the predicted log-odds of prominence marking from the observed log-odds of prominence marking for each item (coding actual zeros and ones as 1/60 and 59/60 respectively and converting to log-odds). To see **whether the fixed effects were more accurate for certain languages and instructions**, the log-odds predictions were converted to p-scores using the inverse of the logit transformation and subtracted from the true *p*-values and squared. A linear regression was run with squared error as the response and the interaction of language and instruction as predictors, and pairwise comparisons were made with Tukey-adjusted *p*-values. As **an estimate of the variance in the items explained by the fixed effects**, the square of the correlation coefficient between the actual p-scores and predicted p-scores was taken (it should be noted that this is not the same as R-squared for a linear model, but rather a rough estimate of the item variance explained).

7. Accuracy using only meaning or acoustic predictors

To see how accurate the p-score predictions would be using only meaning or acoustic predictors, two additional mixed effects logistic regressions were fit using different subsets of the predictors. In the acoustic model, phonerate, intensity, maximum F0, and acoustic instruction boundary marks were included along with their full interaction with language and instruction. In the meaning model, word frequency, part of speech, and meaning instruction boundary marks were included along with the full interaction with language and instruction. The squared correlation coefficient was then computed for each model, as described in Section 6.

Appendix B. Additional details on statistical results

1. Fixed effects (see Table B1).
2. Effects estimates for continuous-valued factors, by language and instruction (see Fig. B1).
3. Model fit (see Fig. B1 and Table B2).

The by-item errors in log-odds of prominence marking are normally distributed, confirmed by a Shapiro Wilk test ($p = 0.15$): (see Fig. B2).

Table B1
Regression estimates, standard error, *z*-value and VIF[*] for all fixed effects in the full model. The dependent variable is the binary prominence rating (0, 1) for each content word in the analysis dataset, as rated by each participant, and including all English, French and Spanish items. Factors: Language (Lang), Instruction (Instruct), Word frequency (Freq), Part of Speech (POS) Boundary, Phonerate (Rate), Intensity, Max F0. [*]*The variable inflation factor (VIF) is a measure of collinearity, and values under 5 are considered acceptable evidence of non-collinearity, i.e., that in independent variable is independent with respect to the other independent variables in the model.*

Fixed Effect	Estimate (log-odds)	Standard error	Z value	VIF
Intercept (Grand Mean Prominence Marking)	−0.8012	0.0893	−8.9767	NA
Language English	−0.0837	0.1263	−0.6625	1.4098
Language French	−0.3415	0.1256	−2.7195	1.4147
Instruction Acoustic	−0.2689	0.0147	−18.3132	2.0896
Word Frequency	−0.7445	0.0519	−14.3454	2.1743
Part of Speech Adjective	0.5730	0.0842	6.8075	2.3464
Part of Speech Adverb	−0.2405	0.0751	−3.2018	2.4084
Part of Speech Noun	0.1863	0.0589	3.1661	1.9591
No Boundary Mark	−0.1819	0.0166	−10.9666	1.0694
Phonerate	−0.3253	0.0369	−8.8057	1.0842
Intensity	0.1336	0.0393	3.4043	1.2848
Max F0	0.1534	0.0396	3.8723	1.3131
Language English: Instruction Acoustic	0.1179	0.0213	5.5309	2.9259
Language French: Instruction Acoustic	−0.1409	0.0208	−6.7605	2.8664
Language English: Word Frequency	−0.2320	0.0781	−2.9697	1.6052
Lnaguage French: Word Frequency	0.1863	0.0600	3.1035	2.4603
Language English: Part of Speech Adjective	−0.0299	0.1146	−0.2607	2.7426
Language French: Part of Speech Adjective	0.2211	0.1166	1.8969	3.1483
Language English: Part of Speech Adverb	−0.0497	0.1036	−0.4792	2.8333
Language French: Part of Speech Adverb	−0.2119	0.0990	−2.1402	2.9238
Language English: Part of Speech Noun	0.0211	0.0827	0.2549	2.3486
Language French: Part of Speech Noun	−0.0678	0.0832	−0.8147	2.8381
Language English: No Boundary Mark	−0.0201	0.0249	−0.8083	1.5105
Language French: No Boundary Mark	0.0249	0.0225	1.1084	1.4793
Language English: Phonerate	0.0098	0.0533	0.1839	1.4287
Language French: Phonerate	−0.0137	0.0511	−0.2684	1.4407
Language English: Intensity	0.0707	0.0571	1.2380	1.6117
Lnaguage French: Intensity	−0.0954	0.0517	−1.8452	1.6243
Language English: Max F0	0.0617	0.0553	1.1140	1.5194
Language French: Max F0	0.0512	0.0529	0.9686	1.6225
Instruction Acoustic: Word Frequency	0.1315	0.0151	8.6850	2.0163
Instruction Acoustic: Part of Speech Adjective	−0.1592	0.0238	−6.6891	2.2458
Instruction Acoustic: Part of Speech Adverb	0.2310	0.0223	10.3367	2.2396
Instruction Acoustic: Part of Speech Noun	−0.0454	0.0167	−2.7216	1.8529
Instruction Acoustic: No Boundary Mark	−0.1132	0.0136	−8.3002	1.7448
Instruction Acoustic: Phonerate	−0.1337	0.0121	−11.0226	1.2172
Instruction Acoustic: Intensity	0.0141	0.0116	1.2133	1.2886
Instruction Acoustic: Max F0	0.1120	0.0116	9.6370	1.3186
Language English: Instruction Acoustic: Word Frequency	−0.0478	0.0230	−2.0800	1.6453
Language French: Instruction Acoustic: Word Frequency	−0.0431	0.0177	−2.4334	2.3191
Language English: Instruction Acoustic: Part of Speech Adjective	0.1014	0.0319	3.1750	2.7740
Language French: Instruction Acoustic: Part of Speech Adjective	−0.1242	0.0335	−3.7071	2.9911
Language English: Instruction Acoustic: Part of Speech Adverb	−0.1662	0.0311	−5.3490	2.9399
Language French: Instruction Acoustic: Part of Speech Adverb	0.1626	0.0304	5.3547	2.8565
Language English: Instruction Acoustic: Part of Speech Noun	−0.0001	0.0236	−0.0052	2.4158
Language French: Instruction Acoustic: Part of Speech Noun	0.0184	0.0238	0.7741	2.6973
Language English: Instruction Acoustic: No Boundary Mark	0.0454	0.0198	2.2889	2.4515
Language French: Instruction Acoustic: No Boundary Mark	−0.0534	0.0192	−2.7800	2.3380
Language English: Instruction Acoustic: Phonerate	0.0694	0.0173	4.0111	1.7471
Language French: Instruction Acoustic: Phonerate	−0.0403	0.0174	−2.3130	1.6356
Language English: Instruction Acoustic: Intensity	0.0122	0.0171	0.7163	1.7537
Language French: Instruction Acoustic: Intensity	0.0106	0.0156	0.6815	1.6801
Language English: Instruction Acoustic: Max F0	−0.0269	0.0161	−1.6695	1.5953
Language French: Instruction Acoustic: Max F0	0.0216	0.0160	1.3490	1.7020

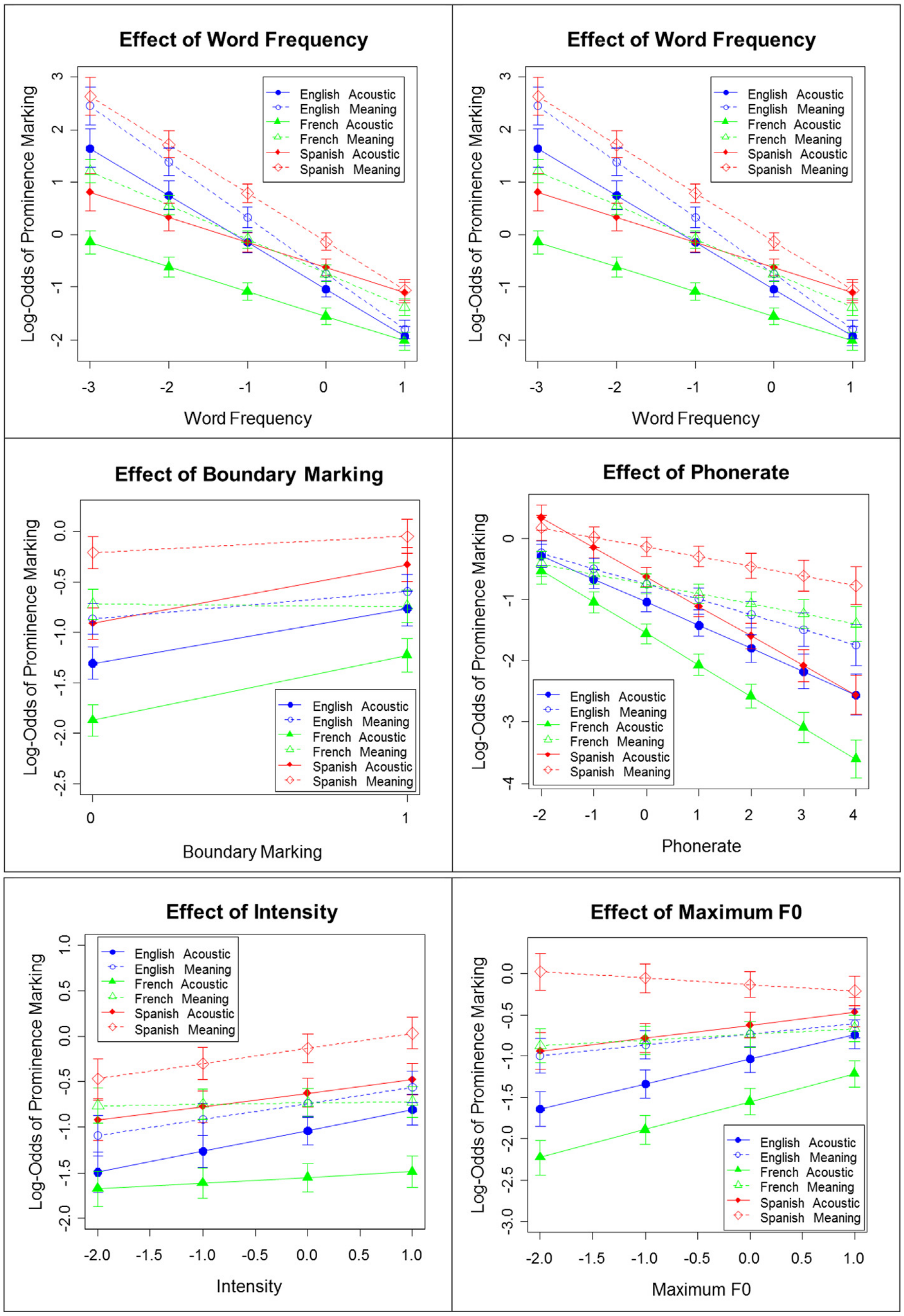

Fig. B1. Co-variation of predicted prominence and continuous word-level predictors (6 panels).

Appendix C. Rapid prosody transcription instructions

The prominence ratings reported in this paper were obtained from three prosody rating experiments conducted with speakers of English, French and Spanish. The experiments were administered through a web browser, using the Language Markup and Experiment Design Software (LMEDS; Mahrt, 2016a). After providing informed consent, participants were shown written instructions for the prominence rating task. Participants received instructions in their native language, which was also the language of the speech samples for which they rated prominence.

1. English RPT instructions

1.1. General task instructions

You will be annotating excerpts of recorded conversations. Each page contains the transcript for the excerpt and may also contain the audio recording. For each excerpt, you will complete two tasks. First you will break the text into smaller sequences of words. Second, you will mark certain words.

In the first task, based on some criteria (see next page) you will select a word and a vertical bar will be appear after that word:

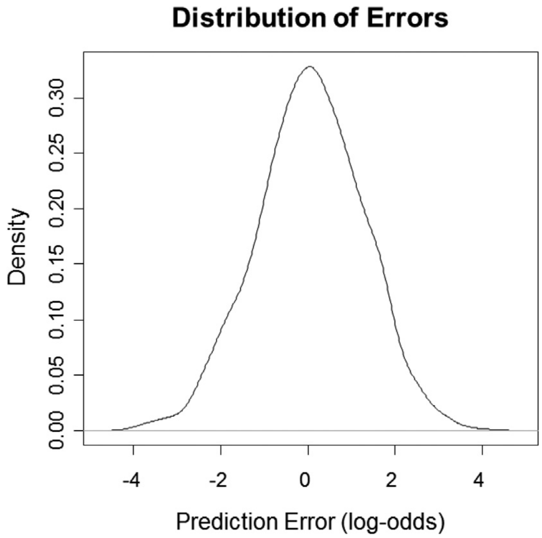

Fig. B2. Density plot of model errors (log-odds).

word word **word** | word word . . .

If you change your mind after selecting a word, just select it again and the bar will disappear. In each utterance, you will be able to select as many words that match the necessary criteria. When you have finished, click 'submit'. This will transition you to the next task.

In the second task you will again select words (based on different criteria—see next page), however, this time, once you select a word its color will change to red:

word word word word . . .

If you change your mind, just select the word again and it will revert back to black. During this phase, you will see where you marked the boundaries from the earlier task. When you are finished click submit to move on to the next excerpt.

1.2. Rating prominence and boundary based on acoustic criteria

The following instructions present the criteria for selecting words in this block.

As mentioned previously, for each audio file, you will conduct two tasks.

In the first task, you will mark the places where you hear a break, discontinuity or disconnection in the speech stream, strong or subtle. These break the fluid speech stream into different segments.

In the second task select the words that stand out in the speech stream by virtue of being louder, longer, more extreme in pitch, or more crisply articulated than other words in the same utterance.

There are sixteen audio excerpts in this block. Each audio file must be listened to twice in each task. Please use the second playback to review your work. You will not be able to listen to the audio more than twice.

1.3. Rating prominence and boundary based on meaning criteria

The following instructions present the criteria for selecting words in this block.

As instructed previously, for each audio file, you will conduct two tasks.

In the first task, select the words after which the audio file could be segmented with minimal disruption of the meaning of the speech.

In the second task select the words that convey the main points of information as you think the speaker intended.

There are sixteen audio excerpts in this block. Please use the second playback to review your work. You will not be able to listen to the audio more than twice.

2. French RPT instructions

2.1. General task instructions

Votre tâche sera d'annoter des extraits de conversations enregistrées. Sur chaque écran vous verrez une transcription de l'extrait et un bouton qui permet d'écouter l'enregistrement audio. Vous ferez deux tâches pour chaque extrait. D'abord vous diviserez le texte en groupes de mots. Deuxièmement vous sélectionnerez certains mots.

Pour la première tâche, vous sélectionnerez un mot (selon des critères qui seront expliqués prochainement) et une barre verticale apparaîtra après ce mot :

mot mot **mot** | mot mot . . .

Table B2
Results of a linear regression model predicting the squared errors of the model from the language-instruction condition of the item. P-scores showed that error magnitude can be separated into two significantly different groups: French Meaning, Spanish Meaning > English Acoustic, English Meaning, Spanish Acoustic, French Acoustic.

Contrast	Estimate (p-score square error)	SE	*t*	*p*
Acoustic, English – Meaning, English	0.003	0.005	$t_{(2212)} = 0.74$	0.976
Acoustic, English – Acoustic, French	0.003	0.004	$t_{(2212)} = 0.72$	0.979
Acoustic, English – Meaning, French	−0.021	0.004	$t_{(2212)} = -4.64$	<0.001
Acoustic, English – Acoustic, Spanish	0.006	0.005	$t_{(2212)} = 1.34$	0.760
Acoustic, English – Meaning, Spanish	−0.018	0.005	$t_{(2212)} = -3.75$	0.003
Meaning, English – Acoustic, French	0.000	0.004	$t_{(2212)} = -0.06$	> 0.999
Meaning, English – Meaning, French	−0.024	0.004	$t_{(2212)} = -5.42$	<0.001
Meaning, English – Acoustic, Spanish	0.003	0.005	$t_{(2212)} = 0.61$	0.991
Meaning, English – Meaning, Spanish	−0.021	0.005	$t_{(2212)} = -4.49$	<0.001
Acoustic, French – Meaning, French	−0.024	0.004	$t_{(2212)} = -5.67$	<0.001
Acoustic, French – Acoustic, Spanish	0.003	0.005	$t_{(2212)} = 0.70$	0.982
Acoustic, French – Meaning, Spanish	−0.021	0.005	$t_{(2212)} = -4.66$	<0.001
Meaning, French – Acoustic, Spanish	0.027	0.005	$t_{(2212)} = 6.01$	<0.001
Meaning, French – Meaning, Spanish	0.003	0.005	$t_{(2212)} = 0.66$	0.986
Acoustic, Spanish – Meaning, Spanish	−0.024	0.005	$t_{(2212)} = -5.06$	<0.001

Si vous changez d'avis après avoir sélectionné un mot, vous pouvez cliquer dessus encore une fois et la barre disparaîtra. Il n'y a pas de limite au nombre de mots que vous pouvez sélectionner dans un extrait. Cliquez sur tous qui vous semblent correspondre aux critères. Quand vous avez terminé un extrait, cliquez sur 'Suivant'. Vous continuerez ensuite à l'extrait suivant.

Dans la deuxième tâche il s'agit toujours de sélectionner des mots, mais selon des critères différents. Cette fois, quand vous sélectionnez un mot il changera de couleur pour devenir rouge :

mot mot mot mot . . .

Si vous changez d'avis, vous pouvez cliquer sur le mot encore une fois et il redeviendra noir. Pendant cette tâche, vous pourrez voir les endroits où vous avez marqué les frontières dans la première tâche. Quand vous avez terminé un extrait, cliquez sur Suivant pour continuer à l'extrait suivant.

Attention ! Il faut sélectionner au moins un mot dans chaque tâche, sinon le test risque de planter. (Normalement vous sélectionneriez plus qu'un mot.)

Ne réfléchissez pas beaucoup avant de répondre ; ce sont vos premières réactions qui nous intéressent. De même, vous avez la possibilité de revoir vos réponses avant de continuer, mais nous préférons que vous passez à l'extrait suivant sans trop de délai. Le test prend environ une heure et demie au total si vous n'hésitez pas trop.

2.2. Rating prominence and boundary based on acoustic criteria

Ces instructions vous expliquent les critères pour sélectionner les mots.

Dans cette partie du test, vous écouterez les quatorze extraits audio, et vous aurez deux tâches à faire pour chaque extrait.

Dans la première tâche, sélectionnez les mots après lesquels vous entendez une rupture ou une discontinuité dans le flux de la parole, que ça soit majeure ou mineure. Ces endroits divisent le flux de parole en groupes de mots distincts.

Dans la deuxième tâche, sélectionnez les mots qui se détachent du flux de parole parce qu'ils sont plus forts, plus longs, articulés plus soigneusement que les autres mots dans le même énoncé, ou que la hauteur de la voix est particulièrement élevée ou baissée.

Vous écouterez chaque fichier audio deux fois pour chaque tâche.

Utilisez la deuxième écoute pour revoir vos réponses. Vous ne pouvez pas les écouter plus que deux fois.

2.3. Rating prominence and boundary based on meaning criteria

Ces instructions vous expliquent les critères pour sélectionner les mots.

Dans cette partie aussi, vous aurez deux tâches à faire pour chaque fichier audio.

Dans la première tâche, il s'agit de sélectionner les mots où une coupure du signal audio (après le mot) créerait la plus petite perturbation dans le sens de ce qui est dit.

Dans la deuxième tâche, vous sélectionnez les mots qui sont les plus importants pour transmettre le message tel que vous pensez le locuteur l'a conçu.

Il y a quatorze extraits audio. Vous écouterez chaque fichier audio deux fois pour chaque tâche. Utilisez la deuxième écoute pour revoir vos réponses. Vous ne pouvez pas les écouter plus que deux fois.

3. Spanish RPT instructions

3.1. General task instructions

Se le va a pedir que anote fragmentos de conversaciones. Cada página contiene la transcripcón del fragmento y puede incluir también una grabación en audio. Para cada fragmento, se le pedirá que complete dos actividades. Primero, deberá dividir el texto en secuencias más breves de palabras. Después deberá marcar ciertas palabras.

En la primera actividad, se le pedirá que seleccione ciertas palabras, siguiendo los criterios que se indicarán en la página siguiente. Cuando seleccione una palabra aparecerá una barra vertical inmediatamente después:

palabra palabra **palabra** | palabra palabra

Si cambia de opinión después de seleccionar una palabra, selecciónela otra vez y la barra vertical desaparecerá. En cada fragmento, puede elegir tantas palabra como le parezca adecuado según los criterios que se le indiquen. Cuando haya acabado, apriete “Enviar”. Con esto pasará a la siguiente actividad.

Para la segunda actividad deberá de nuevo elegir palabras, según otros criterios (véase la página siguiente). Ahora, cuando seleccione una palabra su color cambiará a rojo:

palabra palabra palabra palabra

Si cambia de opinión, simplemente seleccione la palabra de nuevo y se pondrá en negro otra vez. En esta fase podrá ver dónde puso las líneas divisorias antes. Cuando haya acabado apriete “Enviar” para pasar al siguiente fragmento.

3.2. Rating prominence and boundary based on acoustic criteria

Como se mencionó antes, para cada grabación, se le pedirá que lleve a cabo dos actividades.

En la primera actividad, se le pedirá que indique los lugares donde perciba un corte, discontinuidad o desconexión en el habla, ya sea fuerte o solo sutil. Estas discontinuidades dividen la cadena hablada en trozos o segmentos.

En la segunda actividad seleccione aquellas palabras que sobresalen en la cadena hablada por haber sido pronunciadas con más fuerza, por ser más largas, por tener un tono más extremo o por haber sido articuladas más cuidadosamente que otras palabras en la misma frase.

Hay 16 grabaciones en este bloque. Debe escuchar cada grabación dos veces en cada actividad. La segunda vez que escuche la grabación puede hacer cambios. Tenga en cuenta que no podrá escuchar cada grabación más de dos veces.

3.3. Rating prominence and boundary based on meaning criteria

Como se mencionó antes, para cada grabación, se le pedirá que lleve a cabo dos actividades.

En la primera actividad, se le pide que seleccione las palabras después de la cuales la grabación podría segmentarse con interrupción mí¬nima del significado.

En la segunda actividad seleccione aquellas palabras que en su opinión transmiten la información más importante que el hablante quiso transmitir.

Hay 16 grabaciones en este bloque. Debe escuchar cada grabación dos veces en cada actividad. La segunda vez que escuche la grabación puede hacer cambios. Tenga en cuenta que no podrá escuchar cada grabación más de dos veces.

Appendix D. Transcripts of corpus passages

1. English – Buckeye corpus

s02-1

i really don't know i think in today's world what they call the nineties that uh it's just like everything is changed like when i grew up we kids that grew up when we grew up and everything else we had standards and there were certain things we did or didn't do and then it came along that when our kids grew up then there was so many things that changed and

s03-1

there are those doomsayer from my point of view are saying that uh in in many years to come there won't be such a thing as a family uh it's just you you go out here you meet a an opposite sex you [procri-]procreate and you go on your way she goes on her way and she's just as satisfied about going her way as you are going your way

s04-1

she's attractive to you yknow but that's just something that's going through your mind and that i think that's like the devil getting into you because i would never sleep with a woman i would never do it even though i might be really would have been real drunk or yknow or not drunk and just

s10-1

they don't have the money to throw away and like catholic schools they don't have the money to put on those programs so some of those people kind of get pushed aside and so yknow as you can see i'm standing on my soapbox here i do like columbus public schools but yknow i i hate when people say yknow look at these test scores because it really doesn't reflect what's there because it's like saying all kids carry guns to school

s11-1

it's pretty much nobody's winning game mean you make stricter laws you're gonna get yknow even more more people yknow struggling to get a gun and shoot somebody's head off oh the school thing pretty much says itself metal detectors and police at the schools yknow it's come down to that nowadays

s14-1

i was born in dayton ohio i lived there almost thirty five years before moving to columbus in nineteen eighty one uh lived in uh north west columbus for eighteen years uh moved last january a little farther west about five miles further west near hilliard but still live in the city of columbus uh so i've been here about eighteen nineteen years i like columbus also

s16-1

i mean they're not teaching them to say thank you and please and just the normal yknow um thanks for having me when you go to someone's house just the normal i can't tell you how many kids i've picked up and driven home when i picked up my son because it was raining or something and they just jump out of the car and go bye slam the door

s17-1

well it could have been prevented but we didn't know it was gonna happen that our society was gonna change so intensely and we kind of hung back and thought things would stay the way they were and haven't and everybody's changing and especially the younger people

s21-1

i did some other ones yknow emailed and yknow if you have this program can tell me about it and and it was actually i think it was stanford it was on a tv show think it was like extra or something

s22-1

i was born on a homestead in northern montana and then my parents moved back to iowa which was their home area southern iowa and i grew up there attended high school there

s24-1

probably it was a blessing that moved to saint louis because probably could've been more uh problems legal problems yknow as far as going back and forth to court and putting her on the witness stand for stupid stuff

s25-1

i think the main thing that affects me now and since i've been home i've been reading the paper a lot more and yknow i'm more aware of what's going on and i really um gonna be some kind of activist for um i don't know a lot of the places that are coming up the gentleman's clubs and things like that i just i think that's gonna be really bad for the kids and and everybody around

s26-1

i think that their personalities are different too though so he was very independent but i'm not sure that that's just his personality or if it was just he was adapting to his environment and where he was put being put and being in daycare from a very very young age and

s32-1

i got to be the first graduating class of this brand new school that they built because the old one had overflowed beyondd all recognition and we called it the uh the taj mahal because they'd spilled bunches of information into it bunches of money it was pretty it was fun to be the first class in the school cause everything is brand new

s33-1

my thesis is on the sudeten germans those are the germans who lived in czechoslovakia uh before the second world war and specifically i'm looking at how they perceived their condition switching from being part of the austro-hungarian empire

s35-1

he was taking lessons from um gene walker that teaches jazz saxophone here at the university and he said that he thought charlie had the makings be a to make a living career as jazz saxophonist uh but there's also in the last four years or so a lot of uh mental health

2. French – CFPP corpus

CFPP2000 [11-01] Anita_Musso_F_46_11e

euh non elles marchent elles s' non non elles ont tendance à non non elles adorent marcher puisque moi voilà j'peux m'dire mais j'viens t'chercher tu finis tard tu seras voilà non non non laisse-moi marcher laisse-moi marcher donc euh non non pareil qu'elles marchent elles prennent pas l'métro pour faire quatre cinq stations elles prennent pas l'métro elles y vont à pied

CFPP2000 [Mo-01] Andre_Morange_H_58_Mo

t'as euh une fluidité entre les comportements des gens c'é- c'était naturel y avait pas ce cette réflexion qu'on vient d'ailleurs et qu'on est différent quoi j'crois qu'on s'enrichissait plutôt de cette différence plutôt que de dire de faire un blocage là-dessus quoi peut-être que c'est différent maintenant j'en sais rien mais et si vous voulez euh j'ai toujours ressenti euh la ville dans laquelle j'ai vécu où j'ai eu mes racines comme ça

CFPP2000 [SU-01] Dominique_Valin_F_37_SU

euh le supermarché je vais au Leclerc en allant chez mes parents le mercredi soir donc en fait les marchands d'Suresnes j'les connaissais pas énormément euh j'y étais allée une ou deux fois mais j'avais été un peu déçue non j'trouve qu'au niveau euh enfin à part peut-être qu'y a peut-être un un épicier qui est pas mal euh enfin primeur mais sinon l'reste j'ai l'impression que c'est quand même un peu moins bien que des s- enfin les produits sont quand même un peu moins bien qu'à Boulogne on voit quand même que c'est euh

CFPP2000 [13-01] Gabriel_Pujade_H_40_13e

bah les restaurants sont un peu populaires et enfin dans la rue les premiers sont un peu populaires pas trop chers donc ça amène pas de mal de monde d'étudiants tout ça et un peu plus loin dans la rue par contre y a des bistrots un peu français là qui sont un peu plus haut de gamme donc beaucoup plus chers mais très franchouillards

CFPP2000 [20-01] Gary_Collard_H_24_20e

alors moi c'que j'aime l'plus dans l'quartier c'est que à la base j'étais dans un petit passage euh passage Gambetta c'est un tout petit passage un peu éloigné de tout ce qui est commerces euh stations d'métro non pas très loin mais assez en marge avec beaucoup de verdure et tout et c'était un ça faisait un petit coin un peu euh on s- on se serait pas cru dans Paris en fait ça a rien à voir avec euh Châtelet ou République ou

CFPP2000 [Mo-02] Marie_Helene_Matera_F_67_Mo

sur la tête elle avait un les bérets vous savez c'était plutôt un béret d'homme mais bien enfoncé sur la tête avec ses lunettes ses mèches qui qui dépassaient de son de son bonnet elle allait à tout vent elle était grande et sèche je la véa- bien un petit peu un petit peu rouge sur les les joues parce qu'elle était toujours dehors certainement mais j'vois cette femme elle était elle était immuablement habillée pareil et pour moi elle a jamais vieilli elle était toujours pareille

CFPP2000 [IV-01] Jacqueline_Pelletier_F_65_Ivry

ma fille de deux ans mon fils de quatre ans donc après euh ben je suis pas enfin je je suis partie de Bondy euh on est revenus sur Paris euh et et après bon ben je suis rentrée chez Testut et c'est Testut qui m'a entre guillemets euh sauvé la mise hein j'étais collaboratrice d'un acheteur j'suis rentrée comme secrétaire au départ puis après collaboratrice d'un acheteur

CFPP2000 [07-02] Lucie_da_Silva_F_22_7e

oui alors euh on va dire que de mes depuis qu'je suis depuis mes 13–14 ans jusqu'à mes 18 ans c'était euh bah j'ai révisé mon bac au Champ-de-Mars j'ai révisé mon brevet au Champ-de-Mars euh j'ai c'était quand on avait rien à faire euh bah c'est sûr que j'pense que ça rassure un peu les parents quand même au on on était on traînait dans la rue mais au moins ils savaient où on était on était sur la pelouse du Champ-de-Mars

CFPP2000 [IV-02] Monique_Chaslon_F_53_frederic_chaslon_52_Ivry

je sais je sais même pas si on nous plaint mais en même temps voilà ce bon moi je sais que j'ai des collègues des vrais Parisiens avec voilà des gens qui sont installés enfin bon j'ai pas d' sou- enfin eux n'ont pas d'soucis avec moi mais par contre j'les inviterai pas non plus à à dîner ici parce que je sais que très bien qu'en arrivant ici pour eux ça serait le choc est serait trop rude

CFPP2000 [14-01] Nicole_Noroy_F_53_14e

la librairie elle existe depuis euh j'ai une cliente qui fait des recherches généalogiques un jour elle est retombée un petit peu comme ça par hasard sur des sur des registres de commerce et euh c'est elle qui m'a dit qu'apparemment euh la librairie avait été ouverte autour des années alors attendez euh ça doit être en vingt euh je crois que c'est vingt-six donc la librairie le lieu en tant que librairie a à peu près quatre-vingts ans quoi

CFPP2000 [07-03] Pauline_de_Bordes_F_67_7e

alors c'est donc euh c'est mon père qui est arrivé en et euh on a toujours habité vingt-huit boulevard Raspail hein et parce que lui est arrivé c'est un immeuble qui d'ailleurs est intéressant du point d'vue euh architectural on l'fait visiter maintenant par euh les étudiants parce que ça date de j'vous préciserai après mais mille neuf cent vingt-cinq je crois et c'est un immeuble qui euh est très intéressant du point d'vue architectural

CFPP2000 [07-04] Raphael_Lariviere_H_23_7e

moi j'suis quand même dans une dans une filière euh donc médicale donc c'est très spécifique donc on on voit t- souvent des gens euh on parle souvent avec d- ces gens que d'médecine et euh c'est un plaisir de pouvoir parler d'autres choses pouvoir respirer un peu avec des gens qui font du droit des lettres euh ou mm une prépa une école

CFPP2000 [07-05] Yvette_Audin_F_70_7e

c'qui est compliqué si vous voulez c'est que le monde que j'ai connu enfant dans le septième c'est-à-dire dans ce quartier de la rue Babylone Boulevard des Invalides rue Barbet de Jouy est extrêmement était extrêmement différent du monde actuel de ce quartier voilà ce quartier a énormément changé c'était un quartier de familles nombreuses euh pas forcément riches

CFPP2000 [Mo-03] Younes_Belkacem_H_59_Mo

c'était plus convivial y avait une solidarité euh nous quant à moi je je préfère quand j'habitais avec euh cinq ou six personnes que maintenant tout seul parce que euh tout l'monde sait qu'est-ce que vous avez si vous si vous manquez d'argent si vous avez pas d'travail maintenant personne ne vous connaît rien vous êtes tout seul isolé dans votre Assedic avec le R.M.I. des gens ils sont isolés dans les coins avant quand quelqu'un il il é- était embêté pa- j'parle par communautés par contre hein

3. Spanish – Glissando corpus

sp_f11r_fcd_01

pues que el en Valladolid son la ge- yo como no soy de Valladolid y puedo decirlo perfectamente son una gente muy cerrada muy cuesta mucho sa luego cuando son amigos son amigos pero es verdad que cuesta que se abran a la gente

sp_f11r_fcd_02

yo en cuanto vaya ahora para la radio ya le pregunto le digo que qué tal está la cosa ya sabes que no son buenos tiempos pero bueno a ti te conoce mejor algo conocido que lo bueno por conocer no

sp_f19s_fcd_01

pues nos ha nos hacían ir haciendo los patitos desde mi residencia hasta la alfonso octavo todo hasta mmm ciencias y luego allí cantar el clavelito eh mmm ciertas canciones de estas típicas de toda la vida

sp_f21s_fcd_01

ya no sé también se puede mirar sabes si preferimos ir a un sitio más de que tenga que hacer calor y tal pues ya se verá a lo mejor se puede retrasar un poco también hay que hablar

sp_f22s_fcd_01

y que te baja por todo porque que si de lo de las faltas si te sales del tema yo es que cuando empieza a decir que te baja por todo en el examen y nada te sube dije bueno esto va a ser más difícil aprobar

sp_f22s_fcd_02

pues no sé hombre en yo en verano me suelo ir a la pineda sabes dónde está lo de salou pues está muy bien tienes aqua park tienes discoteca la la pachá ésta famosa que hay por todas partes de españa está esa también y hombre lo único que está bastante lejos unas seis horas en tren cinco en alvia más o menos pero yo me lo paso muy bien cuando voy claro que también es semana santa el clima no es el mismo que en verano ese es el problema que tendremos

sp_f23s_fcd_01

y bueno Alex me ha comentado que tenía miedo porque dice jo es que lo tengo muy en sucio y al final la señora me va a decir que esto está mal y que no entiende nada vamos

sp_f24s_fcd_01

es que todavía no se puede mirar que ya lo miramos ayer todavía hasta el una semana antes o sea cuando pase en qué día vivo o sea este sábado ya podremos mirarlo para que te salga un poco en la semana todavía no te lo mira si no

sp_m09a_fcd_01

eso fue un ERE lo que hicieron fue renovar la plantilla por completo cogen a toda la gente la mandaron para fuera cogieron gente nueva porque les salía más barato indemnizar y despedir y coger gente nueva que no van a cobrarte ni los trienios ni nada que que mantener el personal que tenía un saneamiento de empresa

sp_m09a_fcd_02

habían despedido no sé si es un ERE habían despedido una serie de gente y habían cogido bueno convocaron unas plazas y de de técnico de imagen había trescientas plazas se supone que si eh no me sé las cifras pero pongamos que había de esa tacada ochocientos y luego tenía que coger a otros setecientos más a lo que pasa que pues imagino que todo habrá recorte presupuestario vamos yo no yo esa vía la doy

sp_m10a_fcd_01

oye he estado en Jerez el fin de semana que dicen que ha estado inundado y estaban los jerezanos cabreados porque resulta que dicen que han que se han quedado incomunicados y yo he estado el sábado por la mañana comiendo en una terraza en una plaza y la gente ahí tomándose el vermú con sol

sp_m12r_fcd_01

se pasa mal pero luego al final pues merece la pena no porque ini- inicias una nueva vida y la verdad es merece la pena eh merece la sí hay que tomar la decisión a veces cuesta mmm yo muchas personas conozco amigos también que lo están pasando mal y no se atreven pero hay que ser valiente

sp_m17s_fcd_01

claro porque clara la verdad es que todo el mundo le planteas una tercera guerra mundial y dicen sí sí suena de película o sea no nadie se lo espera pero es que yo creo que antes de las guerras mundiales tampoco se lo esperaban y ahora pff se está se está como dejando ahí mmm los organismos internacionales tampoco veo yo que que hagan nada o sea

sp_m18s_fcd_01

salían salían noticias que además me recordaba a todo lo que hemos estudiado de alemania mmm está en desacuerdo con el programa nuclear de irán no sé qué o sea me recuerda completamente a todo lo que sucedía antes de la segunda guerra mundial diciendo pues están en desacuerdo pero no hacemos nada y nos parece mal pero no hacemos nada

sp_m20s_fcd_01

ya pero no pero también es complicada yo me acuerdo vamos un amigo me en época exámenes bueno nosotros en época de exámenes que era el de derecho y me enseñaba en sus apuntes tochos bastante

grandes sobre todo con un un lenguaje muy técnico eran muchos dibujos pero nombres muy raros muchos conceptos nuevos es lo compicado de medicina luego por ejemplo creo que en quinto año tienen una asignatura que es sólo de medicamentos

sp_m20s_fcd_02

nosotros queríamos irnos a trabajar al extranjero este verano mmm na poca po yo que sé pues lo que tu dices sobre todo para dinero y para el idioma porque realmente el dinero que ganes allí lo vas a gastar allí Pero todo lo que ganas te aprendes el idioma todo eso cotiza mucho ahora mismo

Reference

Astésano, C. (2001). *Rythme et accentuation en français. Invariance et variabilité stylistique*. Paris: Editions L'Harmattan.

Astésano, C., Bertrand, R., Espesser, R., & Nguyen, N. (2012). Perception des frontières et des proéminences en français. In *Actes de la conférence conjointe JEP-TALN-RECITAL 2012, vol. 1: JEP* (pp. 353–360). Grenoble: ATAL & AFCP.

Astésano, C., Espesser, R., & Rossi-Gensane, N. (2008). *Quelques cas particuliers de détachement à gauche – ou la prosodie à l'aide de la syntaxe*. Avignon: Journées d'Etude sur la Parole.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. *Language and Speech, 47*(1), 31–56.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America, 119*(5), 3048–3058. https://doi.org/10.1121/1.2188331.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bartels, C., & Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. *The Journal of the Acoustical Society of America, 95*(5), 2973. 2973.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics, 38*(10), 1636–1657. https://doi.org/10.1016/j.pragma.2005.03.017.

Baumann, S., & Riester, A. (2013). Coreference, lexical givenness and prosody in German. *Lingua, 136*, 16–37.

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics, 70*, 20–38. https://doi.org/10.1016/j.wocn.2018.05.004.
Beckman, M. E., & Hirschberg, J. (1993). Guidelines for ToBI labelling. The Ohio State University Research Foundation X, (March), 1–43. Retrieved from https://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html.
Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBi System and the evolution of the ToBi framework. In S.-. A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 1–37). Oxford: OUP. https://doi.org/10.1093/acprof: oso/9780199249633.003.0002.
Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 255–309. https://doi.org/10.1017/S095267570000066X.
Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*(1), 92–111. https://doi.org/10.1016/j.jml.2008.06.003.
Beyssade, C., Hemforth, B., Marandin, J.-M., & Portes, C. (2009). Prosodic markings of information focus in French. In H.-Y. Yoo & E. Delais-Roussarie (Eds.), *Actes de la conférence Interface Discours et Prosodie 2009* (pp. 109–122).
Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (Eds.), *Prosody and meaning (Trends in linguistics)* (pp. 239–270). Berlin: Mouton de Gruyter.
Bishop, J. (2016). Individual differences in top-down and bottom-up prominence perception. *Proceedings of Speech Prosody, 2016*, 668–672.
Bishop, J., Kuo, G., & Kim, B. (in press). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*.
Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer. 7 http://www.fon.hum.uva.nl/praat/.Zugegriffen: 17.
Bolinger, D. (1954). English prosodic stress and Spanish sentence order. *Hispania, 37*, 152–156.
Bolinger, D. (1958). A theory of pitch accent in English. *Word, 14*, 109–149.
Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Palo Alto, CA: Stanford University Press.
Branca-Rosoff, S., Fleury, S., Lefeuvre, F., & Pires, M. (2012). *Discours sur la ville*. Corpus de Français Parlé Parisien des années 2000.
Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25*(7), 1044–1098. https://doi.org/10.1080/01690965.2010.504378.
Büring, D. (2016). *Intonation and meaning*. Oxford: OUP.
Buxó-Lugo, A., Toscano, J., & Watson, D. (2018). Effects of participant engagement on prosodic prominence. *Discourse Processes, 55*, 305–323. https://doi.org/10.1080/0163853X.2016.1240742.
Calhoun, S. (2006). *Information structure and the prosodicvstructure of English: A probabilistic relationship* (Unpublished doctoral dissertation). Edinburgh, United Kingdom: University of Edinburgh.
Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language, 86*(1), 1–42.
Cangemi, F., & Grice, M. (2016). The importance of a distributional approach to categoriality in autosegmental-metrical accounts of intonation pp. 1–20. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 7*(1), 9. https://doi.org/10.5334/labphon.28.
Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-specific perception of speaker-specific productions in intonation. *Individual Differences in Speech Production and Perception, 123–145*. https://doi.org/10.3726/978-3-653-05777-5.
Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 20–51). Amsterdam: John Benjamins.
Chodroff, E., & Cole, J. (2018). Information structure, affect, and prenuclear prominence in American English. *Proceedings of Interspeech 2018*. Hyderabad, India.
Chodroff, E., & Cole, J. (2019). The phonological and phonetic encoding of information structure in American English nuclear accents. *Proceedings of the International Congress on Phonetic Sciences, Melbourne*.
Cole, J., Mahrt, T., & Roy, J. (2017). Crowd-sourcing prosodic annotation. *Computer Speech & Language, 45*, 300–325.
Cole, J., Mo, Y., & Baek, S. (2010b). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes, 25*(7), 1141–1177.
Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010a). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology, 1*, 425–452.
Cole, J. (2015). Prosody in context: a review. *Language, Cognition and Neuroscience, 30*(1–2), 1–31. https://doi.org/10.1080/23273798.2014.963130.
Cruttenden, A. (1993). The de-accenting and re-accenting of repeated lexical items. In *Prosody-1993* (pp. 16–19).
Cruttenden, A. (2006). The de-accenting of old information: a cognitive universal. In G. Bernini & M. L. Schwartz (Eds.), *Pragmatic organization of discourse in the languages of Europe*. Berlin: Mouton de Gruyter.
D'Imperio, M., Bertrand, R., Di Cristo, A., & Portes, C. (2007). Investigating phrasing levels in French: Is there a difference between nuclear and prenuclear accents? In J. Camacho, V. Deprez, N. Flores, & L. Sanchez (Eds.), *Selected papers from the 36th linguistic symposium on romance languages* (pp. 97–110). New Brunswick: John Benjamins Publishing Company.
Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Jun, S-A., Martin, P., Meisenburg, T., Rialland, A., Sichel-Bazin, R., & Yoo, S-H. (2015). Intonational phonology of French: Developing a ToBI system for French. In S. Frota & P. Prieto (Eds.), *Intonation in romance* (pp. 63–100). Oxford: Oxford University Press.
Delais-Roussarie, E., & Rialland, A. (2007). Metrical structure, tonal association and focus in French. In S. Baauw (Ed.), *Romance languages and linguistic theory 2005* (pp. 73–98). Utrecht: Selected Papers from Going Romance. 8–10 December 2005.
Dell, F. (1984). L'accentuation dans les phrases en français. In F. Dell, D. Hirst, & J.-R. Vergnaud (Eds.), *La forme sonore du langage* (pp. 65–122). Paris: Herman.
Delais-Roussarie, E., Rialland, A., Doetjes, J., & Marandin, J.-M. (2002). The prosody of postfocus sequences in French. In *Speech Prosody 2002* (pp. 239–242). Aix-en-Provence.
Destruel, E., & Féry, C. (2015). Compression in post-verbal sequences in French (paper 179). In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the International Congress of Phonetic Sciences*. Glasgow. London: International Phonetic Association.
Di Cristo, A. (1998). Intonation in French. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems. A survey of twenty languages* (pp. 195–218). Cambridge University Press, Cambridge.
Di Cristo, A. (1999). Vers une modélisation de l'accentuation du français: première partie. *Journal of French Language Studies, 9*, 143–179.
Di Cristo, A., & Jankowski, L. (1999). Prosodic organisation and phrasing after focus in French. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. Bailey (Eds.), *Proceedings of the 14th international congress of phonetic sciences* (pp. 1565–1568). San Francisco.
Doetjes, J., Rebuschi, G., & Rialland, A. (2004). Cleft sentences. In F. Corblin & H. de Swart (Eds.), *Handbook of French Semantics* (pp. 531–552). CSLI Publications.
Escandell-Vidal, V., Marrero Aguiar, V., & Pérez Ocón, P. (2014). Prosody, information structure and evaluation. In G. Thompson & L. Alba-Juez (Eds.), *Evaluation in context* (pp. 153–177). Amsterdam: Benjamins.
Estebas-Vilaplana, E., & Prieto, P. (2008). La notación prosódica del español: Una revisión del Sp_ToBI. *Estudios de Fonética Experimental, 17*, 265–283.
Face, T. L. (2003). Intonation in Spanish Declaratives: Differences between Lab Speech and Spontaneous Speech. *Catalan Journal of Linguistics, 2*, 115–131.
Féry, C. (2013). Focus as prosodic alignment. *Natural Language and Linguistic Theory, 31*, 683–734.
Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software, 8*(15), 1–27.
Garrido, J. M., Escudero, D., Aguilar, L., Cardeñoso, V., Rodero, E., de-la Mota, C., González, C., Vivaracho, C., Rustullet, S., Larrea, O., Laplaza, Y., Vizcaíno, F., Estebas, E., Cabrera, M., & Bonafonte, A. (2013). Glissando: A corpus for multidisciplinary prosodic studies in Spanish and Catalan. *Language Resources and Evaluation, 47*(4), 945–971.
German, J., & D'Imperio, M. (2016). The status of the initial rise as a marker of focus in French. *Language and Speech, 59*, 165–195.
Goldman, J.-P. (2011). EasyAlign: an automatic phonetic alignment tool under Praat. In *Proceedings of Interspeech 2011, Florence* (pp. 3233–3236).
Goldman, J.-P., Auchlin, A., Roekhaut, S., Simon, A. C., & Avanzi, M. (2010). Prominence perception and accent detection in French. A corpus-based account. *Speech prosody*.
Gussenhoven, C. (1983). Testing the reality of focus domains. *Language and Speech, 26*, 61–80.
Gutiérrez González, Y., & Aguilar Cuevas, L. (2015). Nuevos datos empíricos sobre la entonación del español a partir del corpus de noticias Glissando. *Estudios de Fonética Experimental, 24*, 35–82.
Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge: CUP.
Hualde, J. I., Elordieta, G., & Elordieta, A. (1994). *The Basque dialect of Lekeitio*. Bilbao & Donostia: Universidad del País Vasco/Euskal Herriko Unibertsitatea & Gipuzkoako Foru Aldundia.
Hualde, J. I., & Prieto, P. (2015). Intonational variation in Spanish: European and American varieties. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 350–391). Oxford: Oxford University Press.
Im, S., Cole, J., & Baumann, S. (2018). The probabilistic relationship between pitch accents and information status in public speech. *Proceedings of Speech Prosody, 9*, 508–511.
Jun, S.-A. (2014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In S.-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 520–539). Oxford: Oxford University Press.
Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation. Analysis, modelling, and technology* (pp. 209–242). Dordrecht: Kluwer.
Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus, 14*, 147–172.
Kimball, A. (2018). *Toward a hybrid model of speech prosody* PhD dissertation. University of Illinois at Urbana-Champaign.
Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America, 118*, 1038–1054. https://doi.org/10.1121/1.1923349.
Krifka, M. (2007). Basic notions of information structure. In C. Féry & M. Krifka (Eds.), *Interdisciplinary Studies on Information Structure 6* (pp. 3–56). Potsdam: Universitätsverlag.
Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge and New York, NY: Cambridge University Press.
Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics, 31*, 81–112. https://doi.org/10.1016/S0095-4470(02)00073-6.
Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33.

Liberman, M. (1975). *The intonational system of English. Doctoral dissertation*. MIT.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8* (2), 249–336.

Luchkina, T., & Cole, J. (2017). Structural and referent-based effects on prosodic expression in Russian. *Phonetica, 73*(3–4). https://doi.org/10.1159/000449104.

Mahrt, T. (2016a). LMEDS: Language markup and experimental design software. https://github.com/timmahrt/LMEDS.

Mahrt, T. (2016b). PyAcoustics. https://github.com/timmahrt/pyAcoustics.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33–B42.

Nadeu, M. (2014). Stress- and speech rate-induced vowel quality variation in Catalan and Spanish. *Journal of Phonetics, 46*, 1–22.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris. Berlin: Mouton de Gruyter.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet. *LEXIQUE, L'Année Psychologique, 101*, 447–462. Version 3.80.

Oh, M. & Byrd, D. (in press). Syllable-internal corrective focus in Korean. *Journal of Phonetics*.

Olarrea, A. (2014). Word order and information structure. In J. I. Hualde, A. Olarrea, & E. O'Rourke (Eds.), *The handbook of Hispanic lingustics* (pp. 603–628). Malden, MA: Wiley Blackwell.

Ortega-Llebaria, M., Gu, H., & Fan, J. (2013). English speakers' perception of Spanish lexical stress: Context-driven L2 stress perception. *Journal of Phonetics, 41*, 186–197.

Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in central Catalan and Castilian Spanish. *Language and Speech, 54*(Pt 1), 73–97. https://doi.org/10.1177/0023830910388014.

Ortiz-Lira, H. (1994). *A contrastive analysis of English and Spanish sentence accentuation* Doctoral diss.. UK: University of Manchester.

Peppé, S., Maxim, J., & Wells, B. (2000). Prosodic variation in Southern British English. *Language and Speech, 43*(3), 309–334.

Pfitzinger, H. R. (1998). Local speech rate as a combination of syllable and phone rate. *Proceedings of ICLSP'98* (3, pp. 1087–1090). . Sydney.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

Pierrehumbert, J. (1980). The phonology and phonetics of English intonation (Doctoral dissertation). Cambridge, MA: Massachusetts Institute of Technology. Distributed by the Indiana University Linguistics Club (1988).

Pintér, G., Mizuguchi, S., & Tateishi, K. (2014). Perception of prosodic prominence and boundaries by L1 and L2 speakers of English. *Proceedings of Interspeech, 2014*, 544–547.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., et al. (2007). *Buckeye corpus of conversational speech* (2nd Release). Columbus, OH: Department of Psychology, Ohio State University. www.buckeyecorpus.osu.edu.

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication, 45*(1), 89–95. https://doi.org/10.1016/j.specom.2004.09.001 [www.buckeyecorpus.osu.edu].

Post, B. (1999). Restructured phonological phrases in French: Evidence from clash resolution. *Linguistics, 37*, 41–63.

Post, B. (2000). *Tonal and phrasal structures in French intonation*. The Hague: Thesus.

Post, B. (2011). The multi-facetted relation between phrasing and intonation contours in French. In C. Gabriel & C. Lleó (Eds.), *Intonational Phrasing in Romance and Germanic. Cross-linguistic and Bilingual Studies* (pp. 43–74). Amsterdam: John Benjamins.

R Core Team (2015). *R: A Language and Environment for Statistical Computing* (version 3.2.2). Vienna, Austria: R Foundation for Statistical Computing.

Rao, R. (2009). Deaccenting in Spontaneous Speech in Barcelona Spanish. *Studies in Hispanic & Lusophone Linguistics, 2*(1), 31–75.

Riester, A., & Baumann, S. (2017). The RefLex scheme – annotation guidelines. In SinSpeC: Working Papers of the Sonderforschungsbereich (SFB) 732 "Incremental Specification in Context". https://doi.org/10.18419/opus-9011.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics, 1*, 75–116.

Roy, J., Cole, J., & Mahrt, J. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology, 8*(1), 22. https://doi.org/10.5334/labphon.108.

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. Proceedings of international conference on new methods in language processing, 12 (pp. 44–49). Citeseer. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.1139&rep=rep1&type=pdf.

Schweitzer, A. (in press). Exemplar-theoretic integration of phonetics and phonology: Detecting prominence categories in phonetic space. *Journal of Phonetics*.

Selkirk, E. (1984). *Prosody and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, E. (1995). Sentence prosody: Intonation, stress and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550–569). Cambridge, MA: Blackwell.

Shattuck-Hufnagel, S., Ostendorf, M., & Ross, K. (1994). Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics, 22*, 357–388.

Shattuck-Hufnagel, S. (1995). The importance of phonological transcription in empirical approaches to "stress shift" versus "early accent": Comments on Grabe and Warren, and Vogel, Bunnell and Hoskins. In B. Connell & A. Arvaniti (Eds.), *Papers in laboratory phonology IV* (pp. 128–140). Cambridge, UK: Cambridge University Press.

Singmann, H., Bolker, B., & Westfall, J. (2015). afex: Analysis of factorial experiments. http://CRAN.R-project.org/package=afex.

Smith, C. L. (2012). Emphasis does not always coincide with phrasal boundaries in spontaneous spoken French. In *Actes de la conférence conjointe JEP-TALN-RECITAL 2012, vol. 1: JEP* (pp. 17–24). Grenoble: ATAL & AFCP.

Smith, C.L., & Edmunds, P. (2013). Native English listeners' perceptions of prosody in L1 and L2 reading. *Proceedings of Interspeech* 2013, Lyon, # IS130507, 235-238.

Smith, C.L., Erickson, E., & Savariaux, C. (in press). Articulatory and acoustic correlates of prominence in French: Comparing L1 and L2 speakers. *Journal of Phonetics*.

Smith, R., & Rathcke, T. (in press). Dialectal phonology constrains the phonetics of prominence. *Journal of Phonetics*.

Vaissière, J. (1997). Langues, prosodies et syntaxe. *Revue Traitement Automatique des Langues, 38*, 53–82.

Vallduví, E. (1991). The role of plasticity in the association of focus and prominence. *Proceedings of the Eastern States Conference on Linguistics (ESCOL), 7*, 295–306.

Vogel, I., Bunnell, T., & Hoskins, S. (1995). The phonology and phonetics of the rhythm rule. In B. Connell & A. Arvaniti (Eds.), *Papers in laboratory phonology IV* (pp. 111–127). Cambridge, UK: Cambridge University Press.

Vogel, A. P., Maruff, P., Snyder, P. J., & Mundt, J. C. (2009). Standardization of pitch-range settings in voice acoustic analysis. *Behavior Research Methods, 41*(2), 318–324.

Wagner, M. & McAuliffe, M. (in press). The effect of focus prominence on phrasing. *Journal of Phonetics*.

Wagner, P., Cwiek, A., & Samlowski, B. (in press). Exploiting the speech-gesture link to capture fine-grained prosodic prominence impressions and listening strategies. *Journal of Phonetics*.

Wagner, M., & Watson, D. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*(7–9), 905–945. https://doi.org/10.1080/01690961003589492.

Watson, D. G. (2010). The many roads to prominence: Understanding emphasis in conversation. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 163–183). Elsevier.

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition, 106*(3), 1548–1557. https://doi.org/10.1016/j.cognition.2007.06.009.

Watson, D., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science, 32*(7), 1232–1244.

Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics, 34*, 343–371.

Zahner, K., Kutscheid, S. & Braun, B. (in press). Alignment of f0 peak in different pitch accent types affects perception of metrical stress. *Journal of Phonetics*.

Zubizarreta, M. L. (1998). *Prosody, focus, and word order, linguistic inquiry monograph 33*. Cambridge, Mass.: MIT Press.