The perception of phrasal prominence in English, Spanish and French conversational speech

José I. Hualde[1], *Jennifer Cole*[1], *Caroline L. Smith*[2], *Christopher D. Eager*[1], *Timothy Mahrt*[1,3] & *Ricardo Napoleão de Souza*[2]

[1] University of Illinois at Urbana-Champaign
[2] University of New Mexico [3]Aix-Marseille Université
[jihualde|jscole|eager2|tmahrt]@illinois.edu, caroline@unm.edu, ricardofns@gmail.com

Abstract

Since Bolinger's [1] discovery that pitch cues accentual prominence in English, a tension has arisen between two strategies: equating accent with pitch excursions and relying on perception for identifying accented words. This paper investigates the relation between prominence judgments from untrained listeners and accentual labels produced by trained transcribers. Naïve speakers of English, Spanish and French (30 per language) were asked to mark prominent words in excerpts of conversational speech from their native language (between 900-1100 words in each sample). Aggregated prominence scores (P-scores) were compared with experts' ToBI labels for each language. For all three languages, words ToBI-labelled as accented had substantially higher P-scores than unaccented words, and nuclear accents had higher P-scores than prenuclear ones. P-scores also discriminated among several accent types. Predictions from prior research on the relative prominence of accent labels were tested, and findings confirm that English L+H* accents are more likely to be judged as prominent than H* accents, and Spanish L+H* is more likely judged as prominent than L+>H*. However, for French, our prediction that Accentual Phrase-initial Hi is prominence-lending was not confirmed. The results establish the link between tonal accents and perceived prominence in three languages that differ in their use of contrastive prominence at the lexical and phrasal levels.

.

Index Terms: Phrasal prominence, prosodic labeling, intonation, English, Spanish, French

1. Introduction

Prominence relations among words in a phonological phrase, which reflect metrical structure, determine accent placement and affect the acoustic realization of words. Researchers investigating prosody can experience uncertainty in the identification of prominence due to variability in the acoustic cues to prominence, including F0. Uncertainty may also arise due to variability in the phonological context of the utterance (e.g., rhythmic factors), in the status of prominence as marking focus, or in factors affecting lexical accessibility [2], [3].

Dwight Bolinger famously stated that *"accent is predictable-- if you are a mind reader"* regarding phrasal accent in English. Whereas the placement of accents by speakers may be difficult to predict, an equally important issue is whether listeners can reliably identify those words that the speaker meant to make more prominent.

We are interested in examining the perception of prominence by untrained listeners across languages. Here we expand on previous work [3], [4] by considering together three languages that differ considerably in their prosodic properties: English, French and Spanish. Whereas both English and Spanish have lexically contrastive stress (e.g. English *p**e**rmit* vs *perm**i**t;* Spanish *pl**a**to* 'dish' vs *plat**ó*** 'stage'), French lacks this property. On the other hand, Spanish differs from English in having essentially predictable nuclear stress on the last word of the phrase, so that contrasts such as *the **phone** rings* vs *the phone **rings*** are usually conveyed by changes in word order, *suena el teléfono* vs *el teléfono suena* [5]. Other work has shown that Spanish speakers find contrast in accent placement in English and other Germanic languages very hard to learn [6], [7]. Given these differences, a question arises regarding how phrasal prominence is perceived by speakers of these three languages.

In particular, if we take consensus annotations by expert ToBI transcribers as the gold standard, we wish to answer the following questions: (1) How well does naïve listeners' perception of prominence align with metrical strength as determined by ToBI pitch accent placement? (2) Are nuclear accents perceived as more prominent than prenuclear accents? (3) Do different accent types in ToBI notation for each of the three languages examined differ in their prominence for naïve listeners?

2. Methodology

The methodology was the same for all three languages. For the American English part of the study, 30 native English speakers listened to excerpts of recorded conversations from the Buckeye corpus (15 excerpts produced by 15 different speakers, 864 words total). The experiment was run using a computer interface developed for this purpose (LMEDS [8]). Participants listened to each excerpt twice through earphones while they were simultaneously shown a transcript with no punctuation or capitalization on a computer screen. They were asked to mark a prominence where they heard a word stand out by 'being louder, longer, more extreme in pitch, or more crisply articulated.' Judgments of phrase boundaries, as well as prominence judgments using other instructions (focusing on information status) were also elicited, but are not reported here. This prominence judgment elicitation technique has been referred to as Rapid Prosodic Transcription or RPT in previous work by one of the authors [3], [4].

For Spanish 30 participants were recruited at the University of Valladolid, Spain. They were asked to perform the same tasks as described for English, using an identical computer interface. The audio stimuli were extracted from the Spanish semi-spontaneous speech part of the *Glissando* corpus [9] and contained 16 excerpts produced by 12 speakers, for a total of 887 words. The *Glissando* Spanish corpus was recorded in Valladolid and thus contains the same Spanish variety as that spoken by our participants.

Finally for French, 30 native French-speaking participants were recruited from introductory linguistics classes at the Université Lumière Lyon 2, France. They were asked to perform the same tasks as described for English, using an identical computer interface, and instructions translated directly from the English instructions. The audio stimuli were extracted from the Corpus du Français Parlé Parisien (CFPP, [10]), and contained 14 excerpts spoken by 14 speakers for a total of 1062 words. The CFPP consists of interviews in a style similar to those in the Buckeye Corpus. Dialect differences are minimal between Paris, where the corpus was recorded, and Lyon, where the listeners were recruited.

For all three languages, prominence labels from all transcribers were aggregated to produce a prosodic score or P-score for each word, representing the proportion of transcribers who judged that word as prominent. For each language, the discourse fragments that the naïve participants marked for prominence were also prosodically labeled by two of the authors using standard ToBI conventions for each of the three languages (MAE_ToBI [11], Sp_ToBI [12], [13], Fr_ToBI [14]). The same two authors labeled the English and Spanish excerpts and two other authors transcribed the French data. The last pitch accent in each phrase (preceding a prosodic boundary) was labeled as nuclear. Consensus ToBI labels thus obtained were compared with P-scores 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.

For each language a mixed effects logistic regression was run in R [15] using the lme4 package [16] 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, using the maximum random effects structure supported by the data [17], which was random intercepts for transcriber and for word token as item. Planned contrasts were then performed using the lsmeans package [18] 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.

3. Predictions

Unaccented words are predicted to result in lower P-scores than words bearing any pitch accent. This follows from the assumption that accents are related to the perception of prominence. This should be the case for all three languages examined here. For all three languages as well, accented words in nuclear position are expected to be judged as prominent more often than words bearing a prenuclear accent. This also follows from the standard assumption that nuclear accents are more salient than other prominences ([19], for French), or more meaningful [20]. Among the three languages considered here, a difference is that English has great flexibility in the location of nuclear pitch accents, whereas Spanish is less flexible in this respect but can use different word orders to locate words in phrase-final position, where they receive the nuclear accent [21], [22]. French, on the other hand, more often uses different syntactic constructions such as dislocation and clefts to highlight constituents. These differences, however, should not result in differences among the languages for the comparison that is being made here regarding the relative prominence of words receiving the nuclear accent in the phrase.

The existing literature on the topic allows us to make only a limited number of predictions regarding the relative prominence of specific accent shapes. For English, L+H* can be assumed to convey a higher degree of prominence than H* [23]. The choice among other pitch-accents has been discussed in terms of notions that are less directly relatable to differences in prominence. Thus Pierrehumbert & Hirschberg [24] claim that H* indicates that the item is salient and new in the discourse, whereas L* is used with items that are salient but are not to be added to the speaker's predication. Although no clear prediction regarding relative prominence emerges from these claims, we are also interested in discovering possible differences in this respect, in addition to the predicted contrast in prominence between L+H* and H*.

For Spanish, the claim in the literature is that the frequent L+>H* accent (a rising accent with a displaced peak) is mostly used in prenuclear position, whereas its counterpart without a displaced peak, L+H* mostly occurs in nuclear position in declaratives with a certain degree of emphasis [13]. This allows us to make a clear prediction regarding the relative perceived prominence of these two accents (to ensure this effect is evaluated independently from the expected contrast in prominence between prenuclear and nuclear accents, grand means were used as described in Section 2). Again, for other accent shapes no clear predictions emerge from the literature on Spanish intonation and our work here is exploratory.

For French, a high accent located on the initial syllable of the accentual phrase, notated Hi, has been said, in one of its uses, to mark the left edge of a prominent constituent, but has also been described as occurring in longer phrases where it may not be prominence-lending [25], [26]. A weak prediction can be made that this accent type is more likely to be perceived as prominent than other accents aligned with the last stressable syllable of the accentual phrase, such as H* and L*.

4. Results

We first present results regarding the marking of prominence of unaccented words and words with nuclear and prenuclear accent for the three languages together. Then, we discuss the results that have to do with specific ToBI labels for each language separately, in different subsections, since both the inventory of labels and the predictions are different for each of the three languages.

4.1. Prominence of accented and unaccented words

Our expert-labeling of the excerpts resulted in the distribution of accents in Table 1. For some additional accent types, less 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 our 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 our 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.

Table 1. Counts of ToBI accent labels by language

English		French		Spanish	
Accent	Count	Accent	Count	Accent	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				

As can be seen from Table 1, in all three languages, most words were judged not to bear an accent in our expert ToBI annotation. Among those words labeled with an accent, H* is by far the most common in all three corpora as well. There are also enough occurrences of all specific accents to perform our language-specific planned contrasts described in Section 3 (English L+H* vs. H*, Spanish L+H* vs. L+>H*, French Hi vs. L* and H*).

In Figure 1, we compare 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. From this comparison it appears 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, as predicted.

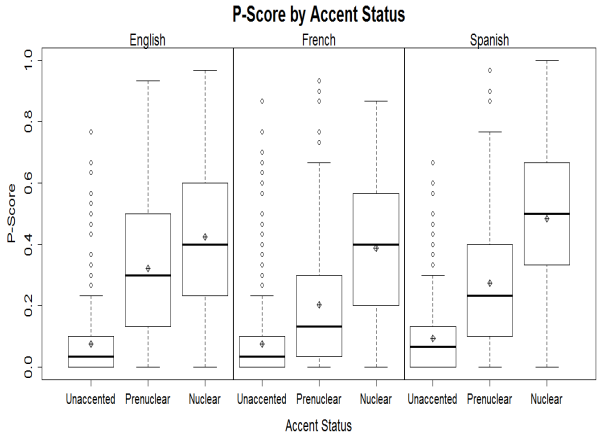

Figure 1. Distribution of P-scores of unaccented words, and words with prenuclear and nuclear accent by language. Diamonds represent means.

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 regressions 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 were significant, as shown in Table 2 below.

Table 2. 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	.18	14.7	<.001	.79	.33	2.4	.015
French	2.09	.14	15.4	<.001	.80	.28	2.9	.004
Spanish	2.31	.12	18.7	<.001	1.30	.21	6.3	<.001

As a main finding, we conclude that in all three languages *naïve prominence judgments are a good approximation of expert prosodic transcription* regarding the interpretation of words as prominent, and equating prominence with accenting. Within the group of accented words, naïve prominence judgments also approximate the distinction between nuclear and prenuclear accents.

We are also interested in determining the relative perceived prominence of different accent shapes as notated in ToBI. In sections 4.2-4.4, we test our planned ToBI label contrast for each language, and examine other accentual label differences descriptively.

4.2. English

Figure 2 shows the relation between P-scores (on y-axis) and different accentual labels for our English data.

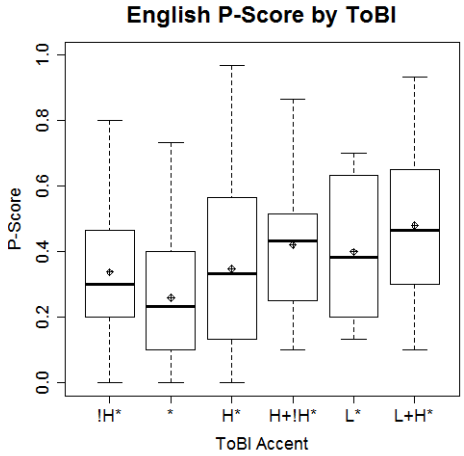

Figure 2. English P-scores for different accentual categories.

From Figure 2 it can be observed that words labeled with a L+H* accent were especially likely to be perceived as prominent, as expected from claims in the literature regarding the semantics of this accent as marking narrow or contrastive focus (L+H* vs. H*: log-odds difference 0.8, z = 3.2, p=.001). For other accents we were not able to derive clear predictions from previous work. Figure 2 includes words marked only with *, without a tonal label, which indicates some degree of prominence but not clearly marked with a visible pitch movement in the F0 display. Words marked with * are less often judged as prominent than other accents (mean log-odds=-1.54). A further observation is that words with a bitonal

accent, H+!H* and L+H*, have the highest median and mean P-scores among the accent types (mean log-odds for bitonal accents = -0.72; for monotonal = -0.94).

4.3. Spanish

P-scores of words with different accent shapes, using Spanish ToBI conventions are shown in Figure 3. Figure 3 shows that, as predicted, L+H* was more often perceived as prominent than L+>H* and the contrast obtained from the mixed effects model confirms that this difference is significant, independent of the contrast between prenuclear and nuclear accents (log-odds difference 0.6, z = 2.6, p=.010). Other pitch-accents receive intermediate scores.

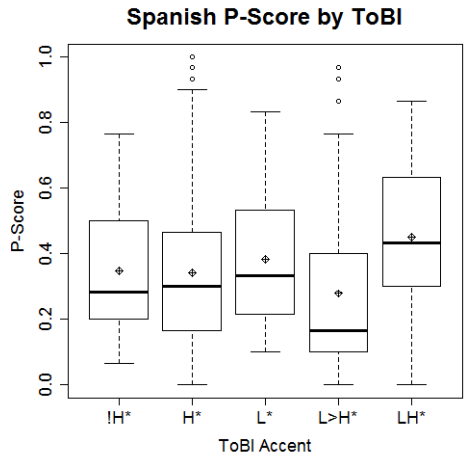

Figure 3. Spanish P-scores for different accentual categories.

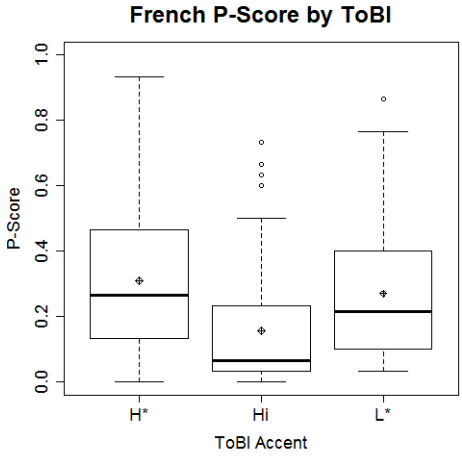

Figure 4. French P-scores obtained under acoustic-based notation of prominence for different accentual categories.

4.4. French

Regarding ToBI labels, our hypothesis was that words with initial prominence (Hi) would be perceived as especially salient. This prediction was not confirmed, see Figure 4. Rather, words with Hi received considerably lower P-scores than words with either a H* or a L* label. The contrast between prenuclear Hi and the grand mean of prenuclear H* and prenuclear L* confirms that Hi is perceived as significantly less prominent than H* and L* (log-odds difference -0.8, z = -2.9, p=.003), contrary to our predictions.

5. Discussion and conclusion

In spite of the notable phonological differences in prosodic properties across the three languages that we have compared in this study, the cumulative P-scores of our English, Spanish and French speaking participants all showed sensitivity to both the unaccented/accented contrast and the contrast between nuclear and prenuclear accent, as independently annotated by expert ToBI labelers. Both accentedness and nuclear accent thus appear to be perceptually salient phenomena in these three languages, at least when judgments of presence vs absence of prominence are elicited through RPT.

Our analysis of relative prominence among different accent types (as labeled according to the standard ToBI conventions for each the three languages) was partly confirmatory and partly exploratory. The two predictions made for English and Spanish were confirmed. In English H* accents received lower cumulative P-scores than L+H* accents. In addition, we noted that the other bitonal accent in MAE_ToBI, !H+H* also had a higher average P-score than all monotonal accents, suggesting a greater salience of more complex tonal excursions. For Spanish non-displaced L+H* accents were more often perceived as prominent than the displaced L+>H*, with a peak on the post-tonic syllable, even when controlling for the difference between nuclear and prenuclear position.

In French the initial Hi accent is often associated with focused constituents in at least one of its functions, and we expected that words with this accent label would receive higher P-scores than words with either a H* or a L* on the last stressable syllable. The results were just the opposite. The explanation may be that a high pitch excursion on the initial syllable of the accentual phrase most often has non-prominence-lending functions, at least in the type of discourse that we have examined.

This study is to our knowledge the first to demonstrate a relationship between prominence as perceived by naïve listeners and accent marks in a ToBI annotation system, regardless of the substantial differences among these languages in the grammatical use of prominence to mark focus, and in the status of phrasal stress. Further, we have shown a distinction in the perceptual salience of words that have nuclear vs. pre-nuclear prominence, a distinction which is, however, not directly encoded in the ToBI annotation. In future and on-going work we intend to examine the acoustic and non-acoustic factors of the stimuli that correlate with higher or lower P-scores for each language under varied instructions.

6. Acknowledgements

We thank Joe Roy for statistical advice. Thanks also to Suyeon Im and Sarah Little for help with various aspects of this research. The research reported here was made possible in part thanks to NSF grants BCS 12-51343 and BCS 12-51134

7. References

[1] Bolinger, D. Accent is predictable (if you're a mind reader). *Language 48*.3: 633-644, 1972.

[2] Calhoun, S. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language,* vol. 86*:*1-42, 2010. 2010.
[3] Cole, J., Mo, Y., & Hasegawa-Johnson, M. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology,* vol. 1:425-452, 2010.
[4] Cole, J., Mahrt, T. & Hualde, J.I., Listening for sound, listening for meaning: Task effects on prosodic transcription. *Speech Prosody 7*: 859-863, 2014.
[5] Bolinger, D. English prosodic stress and Spanish sentence order. *Hispania,* vol. 37.2: 152-156, 1954.
[6] Zubizarreta, M. L. & Nava, E. Encoding discourse-based meaning: Prosody vs. syntax. Implications for second language acquisition. *Lingua,* vol. 121: 652-669, 2011.
[7] Maastricht, L. van, Krahmer, E. & Swerts, M. Prominence patterns in a second language: Intonational transfer from Dutch to Spanish and vice versa. *Language Learning*, 2015. doi: 10.1111/lang.12141.
[8] Mahrt, T. Language Markup and Experimental Design Software. http://prosody.beckman.illinois.edu/lmeds.html, 2013.
[9] Garrido, J. M., Escudero, D., Aguilar, L., Cardeñoso, V., Rodero, E., De-la-Mota, C., González, C., Rustullet, S., Larrea, O.m Laplaza, Y., Vizcaíno, F., Cabrera, M. & Bonafonte, A. Glissando: a corpus for multidisciplinary prosodic studies in Spanish and Catalan. *Language Resources and Evaluation,* vol. 47. 4: 945-971, 2013; DOI 10.1007/s10579-012-9213-0.
[10] Branca-Rosoff S., Fleury S., Lefeuvre F. & Pires, M. Discours sur la ville. Corpus de Français Parlé Parisien des années 2000. http://cfpp2000.univ-paris3.fr/CFPP2000.pdf, 2012.
[11] Beckman, M. & Hirschberg, J. The ToBI annotation conventions.http://www.ling.ohio-state.edu/~tobi/ame_tobi/ annotation_conventions.html
[12] Estebas Vilaplana, E. & Prieto, P. La notación prosódica en español. Una revisión del Sp_ToBI, *Estudios de Fonética Experimental XVIII*, 263-283, 2009.
[13] Hualde, J.I. & Prieto, P. Intonational variation in Spanish: European and American varieties. In Frota, S. & Prieto, P. (eds.), *Intonation in Romance*. Oxford: Oxford Univ. Press, 350-391, 2015.
[14] 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, H. Intonational Phonology of French: Developing a ToBI System for French. In Frota, S. & Prieto, P. (eds.), *Intonation in Romance*. Oxford: Oxford Univ. Press, 63-100, 2015.
[15] R Core Team. R: A language and environment for statistical computing. RFoundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org, 2013.
[16] Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,* vol. 67(1), 1-48, 2015.
[17] Barr, D., Levy, R., Scheepers, C. & Tily, H. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language,* vol. 68, 255–278, 2013.
[18] Lenth, R. & Hervao, M. Using the lsmeans package. 2013. http://CRAN.R-project.org/package=lsmeans.
[19] D'Imperio, M., Bertrand, R., Di Cristo, A. & Portes, C. Investigating Phrasing Levels In French: Is There A Difference Between Nuclear And Prenuclear Accents? In Camacho, J., Flores-Ferrán, N., Sánchez, L., Déprez, V. & Cabrera, M. J. (eds.), *Selected Papers from the 36th Linguistic Symposium on Romance Languages* (LSRL). New Brunswick: John Benjamins, 97-110, 2007.
[20] Brazil, D. *The communicative value of intonation in English*. Cambridge: Cambridge Univ. Press, 1997.
[21] Vallduví, E. *The informational component*. New York: Garland Publishers, 1992.
[22] Ladd, R. L. *Intonational Phonology,* 2nd ed. Cambridge: Cambridge Univ. Press, 2008.
[23] Steedman, M. The surface-compositional semantics of English intonation. *Language 90*.1: 2-57, 2014.
[24] Pierrehumbert, J. & Hirschberg, J. The Meaning of Intonational contours in the interpretation of discourse. In Cohen, P., Morgan, J. & Pollack, M. (eds.), *Intentions in Communication*, MIT Press, Cambridge MA. 271-311, 1990.
[25] Astésano, C., Bard, E. G., & Turk, A. Structural Influences on Initial Accent Placement in French. *Language and Speech,* vol. 50: 423-446, 2007.
[26] German, J. & D'Imperio, M. The Status of the Initial Rise as a Marker of Focus in French, *Language and Speech*, Online First. DOI: 10.1177/0023830915583082, 2015.