Naïve listeners' perceptions of French prosody compared to the predictions of theoretical models

Caroline L. Smith

caroline@unm.edu

University of New Mexico

Abstract :

In this study, naïve French listeners' perceptions of prosody are compared to descriptions of prosodic structure in the literature, and to the results obtained in a similar experiment with American listeners (Cole et al. submitted). Untrained participants listened to recorded spontaneous speech, while following along on an unpunctuated transcript. The listeners were assigned to one of two groups: one group was asked to underline all the "important" (prominent) words, the other to mark locations where they perceived a break between two groups of words (phrasing). The French listeners demonstrated a strong tendency to mark as prominent those words immediately preceding locations perceived as phrase boundaries. This tendency agrees with descriptions of French accentual groups as ending with a prominence. However, the listeners marked boundaries on average approximately every ten words, implying phrasal groupings far longer than accentual groups. A possible explanation is that the listeners were actually marking Intonational Phrases.

1. Background

1.1. Prominences and boundaries in French prosody

French prosody has been analyzed by numerous researchers whose models, at least for the most part, share certain proposals in common. (A survey of several of these is given in Lacheret-Dujour and Beaugendre 1999.) At least two levels of prosodic structure are identified between the syllable and the utterance; here I will refer to the smaller of these as an 'accent group' and the larger as an 'Intonational Phrase'. (For a helpful diagram comparing terminology used by different authors see Di Cristo 2005:152.) Two potential locations for prominence are identified within the accent group. These are often referred to as initial and final accents. In some cases, these (particularly the initial accent) are further classified according to different roles that they may fulfill (Di Cristo & Hirst 1997, Di Cristo 1999).

The study reported here investigates the connection between two aspects of prosody, prominence and phrasing. Descriptions of French prosodic structure take for granted that prominence and phrasal boundaries are related, with a prominence falling immediately before a prosodic boundary. "L'accent final peut être considéré principalement comme un attracteur de frontière" (Di Cristo 2000:40) is a typical statement. An even stronger claim is that "final stress [= *accent final* in French] entails a right hand boundary of the intonation unit." (Mertens 2006:70) Writers seem to differ as to whether the occurrence of prominence derives from the boundary, or vice versa. Jun and Fougeron's (2002:147) description suggests that the prominence derives from the existence of a boundary: "the final full syllable of a word is realized with longer duration and higher intensity [that is, as more prominent] only if it is the last full syllable of a phrase." Di Cristo (2000:39), on the other hand, seems to imply that the final accent "generates" a boundary: "il existe une dissymétrie fonctionnelle potentielle entre l'accent initial et l'accent final, dans la mesure où le premier est, en règle générale, générateur

d'emphase et le second de frontière." In any case, the models predict, or assume, that the locations of prominent syllables and the locations of prosodic boundaries are linked. Note that this is in sharp contrast to English, in which there are no particular predictions about the locations of prominences relative to boundaries, except that in a neutral declarative utterance the word with nuclear accent is often the last content word. For example, in ToBI annotation of English prosody (Beckman and Ayers Elam 1997), the marking of pitch accents (prominences) and break indices (boundaries) are treated as two separate tasks.

"Prominence" is being used here without reference to any specific theory, to refer to any word or syllable that stands out from its neighbors by virtue of some combination of acoustic properties making it more salient. Referring to a syllable as prominent is perhaps a less clear-cut notion in French than it is in English. According to Vaissière (2002:151), "when asked which syllable is most prominent in an isolated French word, a naive Frenchman is likely to be puzzled." Models of accentuation in French predict prominence for individual syllables, but the specific syllable that is accented (prominent) will vary depending on structural, rhythmic and pragmatic factors (discussed by numerous authors, including Di Cristo 2000, Pasdeloup 1990, Post 2000). Thus a syllable can only be defined as prominent within a specific context. Above the syllable, the next larger unit relevant for accentuation is what I am referring to as the accent group. Apart from these prosodic influences, a lexical word may acquire prominence due to emphasis motivated by pragmatic or discourse factors.

Recent studies of French prosodic structure have generally been based on the analysis of a corpus of recordings. In most cases the detection of prominence or phrase boundaries is based on the perception of the researcher, and theoretical proposals are based on these individual perceptions. Researchers agree that speakers vary as to which potential accents they may realize in an individual production, but do not seem to have studied listeners in order to determine whether, and how, they vary as to which accents they perceive. Likewise, the notion that words can be grouped into functional units (prosodic phrases of some kind) seems to be easily accessible to untrained French speakers, but relatively few studies have investigated what groupings are perceived by listeners who lack specialized knowledge of posited prosodic structure. It is this gap that the present paper seeks to fill, by examining whether naive, untrained French listeners perceive prominences and phrasal boundaries as predicted by theoretical models, and in particular whether they perceive the correlation between the locations of prominences and boundaries that is a part of virtually every model.

1.2. Testing naive listeners

Previous studies have shown that it is possible to gain insight to the perceptions of prosody by language users who are not trained linguists. This expands the data on prosodic structure beyond prosodic transcriptions produced by a small number of experts. One study of Dutch (Streefkerk et al. 1997) tested the perceptions of listeners who had no special training. The speech material used in this study were phonetically rich sentences read aloud, which are likely to be less varied prosodically than spontaneous speech. Buhmann et al. (2002) also asked naive Dutch listeners to do prosodic annotation, but after 16 hours of training. That untrained listeners use linguistic knowledge in perceiving prosodic boundaries was shown by Mettouchi et al. (2007). They asked both native speakers and non-speakers of Kabyle and Hebrew to mark boundaries in samples of speech (the native speakers only worked with their own language). Native speakers listened to speech that had been filtered to render segmental information unintelligible, in order to ensure they responded purely on the basis of prosodic information. Their responses were closer to an expert transcription than were the non-speaker's transcriptions, which were presumably also based on gross prosodic patterning.

A recent study undertaken in English (Cole et al. to appear a, b; Mo et al. 2008) included a larger number of listeners, 97 in four groups. This study served as the model for the experiment reported in this paper. Cole et al. asked untrained listeners, native speakers of American English, to mark prominence or boundaries while listening to a sample of spontaneous, conversational speech (extracted from the Buckeye Corpus, Pitt et al. 2005). As they listened, they followed along on a printed, orthographic but unpunctuated transcription. Half the listeners marked prominence first for one set of materials, then boundaries on a different set, while the other half performed the tasks in the reverse order. The listeners indicated their responses by underlining a word they perceived as prominent, or by marking a slash between two words where they perceived a boundary between two "chunks" of speech. Cole et al. obtained high rates of agreement among their listeners, higher for the marking of boundaries than for prominence.

Similar methodologies have been used in some previous studies of French prosody in which naive listeners labeled accents and/or boundaries in a recorded speech passage. Pagel et al. (1995) and Obin et al. (2008) do not report in detail the responses of their listeners, as their interest was in developing automated methods for prosodic labeling. Portes (2000) provides a detailed analysis of the responses of 12 naive listeners and 5 experts. Her naive listeners were first given a brief explanation of their task but no feedback or training. They were allowed to listen to the recording being analyzed as many times as they wished while labeling boundaries, accented syllables and emphasized words or expressions. A boundary, accent or emphasis was considered to be present at each location that was marked as such by a majority of the listeners. Portes found that the syllable preceding a boundary was marked as accented at 84% of the identified boundaries. She concludes that this supports the view that the end of an accent group is the best location for a boundary. Portes notes that it is impossible to claim that the boundary locations identified by listeners correspond to a specific linguistic unit: the labeled boundaries demarcate chunks that vary greatly in length and syntactic content. She notes this particularly for the non-terminal (comma) boundaries identified by the expert labelers, but the concern applies to those labeled by the naive listeners as well, and to a lesser extent, to the boundaries identified as terminal (in punctuation, corresponding to a period). Portes suggests that in many cases the unit demarcated by non-terminal boundaries is a clause, or a clause plus additional constituents. This issue will be raised again with respect to the results of the present study in section 3.5 below.

Portes's study raises many interesting issues, but is somewhat limited in that only one speech sample was analyzed and relatively few listeners participated. The study reported here uses a methodology similar to that of Portes and Cole et al., with more listeners than Portes. Most notably, by comparing two types of speech materials, it aims to uncover additional factors contributing to listeners' perceptions of the structure of spoken French.

2. Method

2.1. Speech Materials

Two types of speech materials were used. One set of ten extracts was prepared from recordings of a map task experiment that had been previously recorded at a Paris university (Smith 2007). The speakers are ten female undergraduates from the Paris region. They were recorded individually in a task which required them to give directions over the telephone as to how to use the Paris métro system to travel to various destinations around the city. They were speaking to an interlocutor who they could not see, but who posed questions and provoked discussion. These extracts thus consist of fairly informal, spontaneous task-directed speech.

The extracts were selected from portions of the conversations during which the one speaker had a relatively long conversational turn, and there was no overlap with the interlocutor. These extracts varied from 13 to 24 seconds in length, and are identified by speaker number (Extract1 – Extract10).

The second set of ten extracts was taken from a discussion/debate that was broadcast in December 2008 on a current affairs program on the France Info radio station. The subject is television advertising. These extracts also consist of single-speaker passages of spontaneous conversational speech, but the speakers are journalists and public figures. Their conversation was recorded for broadcast and thus illustrates a more formal register. The selected samples include two extracts from each of the five speakers who participated in the discussion. The duration of these extracts is from 26 to 53 seconds. These extracts are identified by a letter indicating the speaker (A – E, in the order that the listeners heard them) and a digit 1 or 2, indicating the first or second extract from that speaker that was presented to the listeners. For example, Extract D-2 was the second one presented from the fourth speaker.

Orthographic transcriptions of the extracts were prepared by the experimenter (a fluent non-native speaker of French), then edited by a phonetically-trained native speaker. These transcriptions were prepared for use in the listening test by removing punctuation and line breaks except as necessary to fit on the page, in order to avoid providing any hints as to the structure. Disfluencies such as repeated or partial words were included in the transcripts but filled pauses ("euh…") were not indicated. Three additional extracts (two from map task conversations, one from a radio program similar to the one used for testing) were also prepared to serve as practice samples.

2.2. Participants and testing procedure

Fifty-one listeners without advanced training in phonetics or prosody were recruited at higher education institutions in France. Most were undergraduate students in linguistics. In order to test listeners in groups for efficiency reasons, they were not screened for native language, and thus, a few were included who are non-native speakers of French. This was considered unlikely to be a problem because: (i) All participants are sufficiently fluent to participate in higher education (and informal conversation with the experimenter suggests all are very comfortable in French); and (ii) It is estimated that no more than four of the 51 listeners tested are non-native users of French, so their influence on the overall results will be minimal. The vast majority of listeners were female. Different listeners participated in the experiment in different settings at the Université Paris 3, Université Lyon 2 and the École Normale Supérieure – Lettres et Sciences Humaines in Lyon: some were tested in groups of 5-17 in a classroom, others individually or in groups of two or three in a sound-attenuated room. Each listener was presented with a packet containing instructions and the printed transcriptions of the practice and test extracts. They marked all their responses on these print-outs.

Each listener performed one of two tasks. 25 of the listeners were instructed to underline words that were important or highlighted ("mis en relief"); 26 others were asked to mark a vertical line between words at locations where they perceived a boundary between two phrasal units ("syntagmes", defined as groups of words that form a single unit for both meaning and function). All listeners heard the extracts in the same order, with brief pauses between extracts, controlled by the experimenter depending on the listener(s)' wishes. They practiced their task first on two map task extracts, then responded to ten map task extracts, then did a practice with an extract from a radio broadcast, then responded to the ten extracts from the radio broadcast. The ten extracts from the map task were presented in the (random) order in which the speakers had been recorded. The radio broadcast extracts were presented in random order, not in the order in which they occurred originally in the program. No two

extracts with the same speaker were presented consecutively.

One listener in the boundary-marking group failed to follow directions, so that individual's responses were excluded from analysis, leaving a total of 50 listeners. All other responses to the experimental samples were retained, and coded in Excel spreadsheets. Listeners in the prominence-marking group were instructed to underline entire words, even if they only perceived a part of the word as prominent. In fact, many of them occasionally underlined parts of words. The difficulty was that the listeners had to be respond very rapidly (at the speed of the speech), and therefore listeners' markings were often rather imprecise. Because of this difficulty, it was judged infeasible to associate prominence with any unit smaller than a complete word. Thus, if a listener underlined any part of a word, the response was tallied as a prominence marking on the entire word. This ignores the fact (discussed in section 1.1) that French is described as having prominence associated with a syllable, rather than a word.

2.3. Statistical analyses

The data were tabulated in Microsoft Excel, and most of the results reported here are counts and proportions that were calculated using Excel. Agreement among listeners was assessed using a modified form of Cohen's Kappa. Kappa is a statistic that takes into account the amount of agreement that can be expected by chance. Kappa values can vary between 0 and 1. The particular form of Kappa used here is based on Brennan and Predinger (1981); it is suitable for tasks with multiple raters in which the raters are not constrained as to how many items they assign to each category ("free marginals"). Calculations were made using the Online Kappa Calculator (Randolph 2008). Kappa values were determined for each extract, pooling across all the listeners in each of the two groups.

A second type of analysis involved the calculation of a prominence score and a boundary score for each word. These were equal to the proportion of listeners who marked that word as having prominence, or as being followed by a boundary. A 'word' is defined orthographically, as any letter/number sequence separated by spaces from adjacent text. Those words marked by two-thirds or more of listeners (17 or more out of the 25 in each group) were considered to have "consensus" agreement. This criterion was arbitrary but indicates a substantial consensus.

The prominence and boundary scores were used to test the hypothesis of a correlation between the locations of prominences and boundaries. The large number of words receiving no marking (and hence scores of 0) would tend to inflate the correlation between these two, so all words that received two scores of 0 were excluded from calculation of the correlation, which was done using Analysis Tools in Microsoft Excel. The correlation values reported here are thus conservative estimates of any connection between the two variables.

3. Results

As a first step in evaluating the results of the experiment, the rates of agreement among the listeners were examined, in order to demonstrate the validity of the methodology.

3.1. Rates of agreement among listeners

Percentage rates of agreement are shown in Figure 1, as they provide a readily understandable representation of the distribution of listeners' responses. However, rate of agreement does not take chance agreement into account, and therefore the statistical analyses used the kappa statistic, as described above, rather than the raw rates of agreement.

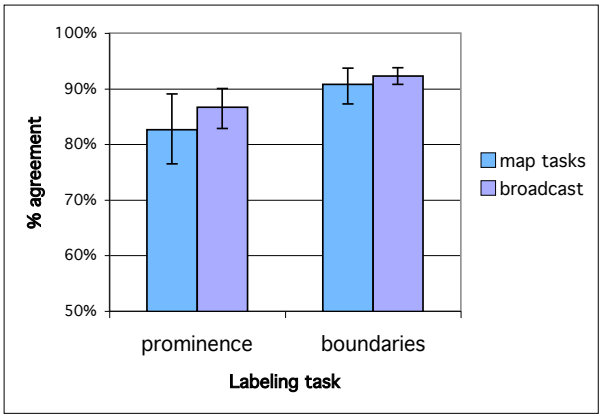

Figure 1. Percentage rates of agreement among listeners for the two sets of extracts and the two labeling tasks. The error bars show the range for the different extracts.

The values of the kappa statistic used to assess agreement ranged from .53 to .80 for marking of prominence, with a mean of .69 across the 20 extracts. For boundary marking, kappa ranged from .75 to .88 with a mean of .83. Randolph (2008) suggests that for this form of kappa, .7 or above is "adequate", which means that the agreement for prominence marking is borderline, but the rate for boundaries is well above this proposed cut-off. Because of the different calculation methods, kappa values reported here are not directly comparable to those in Cole et al. (to appear) for English or Buhmann et al. (2002) for Dutch. Lower rates of agreement might be expected in the present study, because these listeners were untrained and heard each passage only once, while Buhmann et al.'s listeners received training, and Cole et al.'s listeners heard the passages twice. Nonetheless, it is striking that in the present study, as in the earlier ones cited here, higher rates of agreement were obtained for marking of boundaries rather than for marking prominence. Again, small differences in experimental methodology could be a factor: Cole et al.'s listeners marked prominence in one set of extracts, then boundaries in another (or vice versa), so comparison of the two tasks is comparing the same participants. In contrast, the listeners in the present experiment were assigned to groups that marked either prominence or boundaries. Comparison of the two tasks is thus comparing the behavior of different listeners, who were, however, drawn from the same populations. The notably higher agreement for boundary-marking is especially surprising given the frequent assumption that in French, prominence derives from the occurrence of a boundary. This difference suggests that listeners' perceptions of the two aspects of prosody may not in fact be derived from the same information.

Perhaps the most striking examples of agreement among listeners are the thirteen locations where every listener in the boundary-marking group marked a boundary at the same place. Three of these locations occurred in three different map task extracts, and ten in four of the broadcast extracts. The extract with the largest number of boundaries marked by all listeners is given here. The four boundaries that were marked by all listeners are indicated by double vertical lines; three others marked by 67-99% of listeners have a single vertical line.

(1) alors je je crois d'abord que le l'objectif de Nicolas Sarkozy n'est certainement pas de faire une bonne télé publique || mais une plutôt une télé aux ordres || le le fait qu'il nomme désormais le le directeur de de de de ces médias publiques | veut bien dire qu'il est dans une logique de soumission au pouvoir exécutif || ensuite les promesses de Nicolas Sarkozy n'engagent que ceux qui les croient || il peut promettre mondes et merveilles au secteur public | comme Xavier Darcos promet

à l'école de mieux fonctionner avec moins de profs et moins de moyens | ces formules magiques elles ne elles ne bernent personne [Extract E-1]

3.2. Connecting boundaries and prominence

3.2.1. Prominence of words adjacent to boundaries

As described in the methods section, "consensus" markings were identified as those locations where at least 67% of the participants had marked a boundary or prominence. The number of locations so identified is given in Table 1, together with the total number of words.

	Mean	Minimum	Maximum
Map task extracts			
Number of words in extract	55.9	23	92
Number consensus prominent words	2.2	1	3
Number consensus boundary locations	3.5	2	5
Broadcast extracts			
Number of words in extract	134.8	87	206
Number consensus prominent words	4.3	1	7
Number consensus boundary locations	8.5	5	12

Table 1. Number of words in each extract that were marked as prominent, or locations marked as boundaries, by two-thirds or more of listeners.

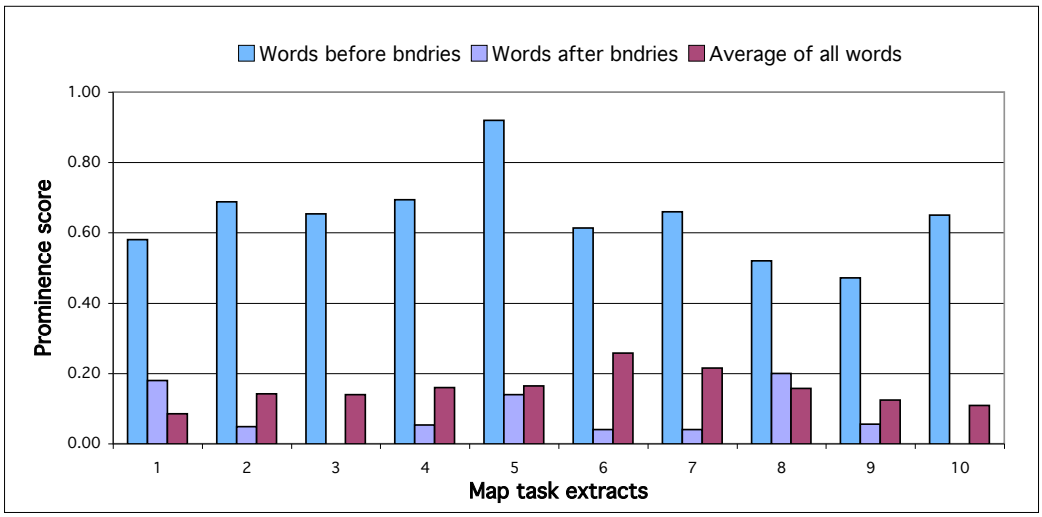

Figure 2a. Prominence scores for words before and after locations identified as boundaries by at least two-thirds of listeners, and average prominence score for all words in that extract.

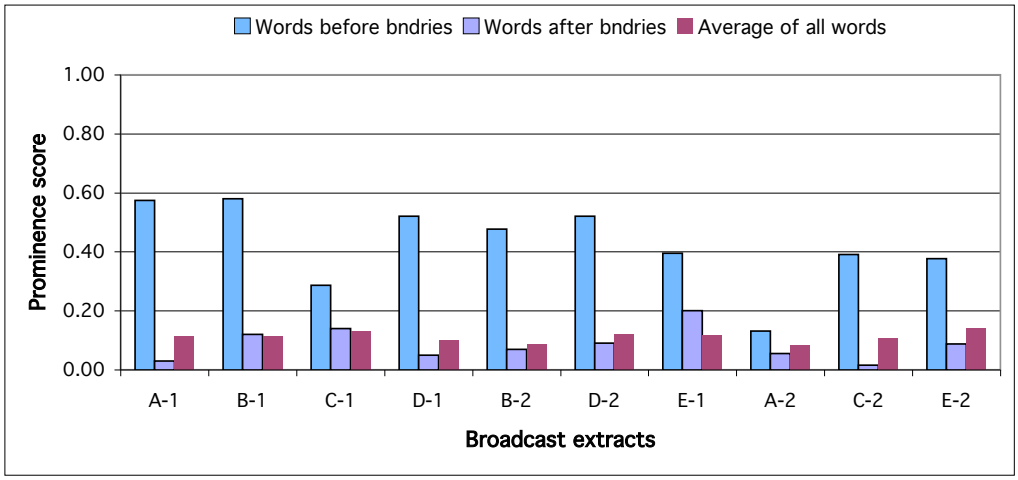

Figure 2b. Prominence scores for words before and after locations identified as boundaries by at least two-thirds of listeners, and average prominence score for all words in that extract.

8

Prominence scores of the words before and after the consensus boundaries were examined in order to determine whether the words in these positions have prominence scores that diverge from the average for the extract. Figures 2 shows that, as expected, words before boundaries received much higher prominence scores than average. Words after boundaries received much lower prominence scores, in fifteen cases lower than the average for all words in that particular extract. (The figure shows the extracts in the order the listeners heard them.)

3.2.2. Correlation between words marked as prominent and locations of boundaries

In order to further investigate the relation between the participants' marking of boundaries and of prominent words, the correlation was calculated between the prominence scores and boundary scores of the words in each extract. This analysis did not examine specific boundary locations; rather, it looked at the overall relation between locations marked as boundaries and the prominence of words before or after them. Because prominences were marked more frequently than boundaries, there cannot be a perfect correlation between them. In addition, the correlations will be reduced if the marked prominences are in some cases at the end of an accent group (where the prominence would precede the boundary) and in other cases at the beginning of an accent group (where the prominence would follow).

Between word prominence and a following boundary, the correlations averaged .68 with a standard deviation of .08 across the ten map task extracts. All of these are significant at $p<.001$. Across the ten broadcast extracts, correlations averaged .51 with a standard deviation of .17. These are also significant at $p<.001$, with one exception where $p=.001$. These results support the hypothesis that listeners tend to perceive prominence on words where they perceive a boundary following. Correlations between a prominent word and a preceding boundary were negative for all extracts, implying that a boundary tends **not** to precede a prominent word. The correlation averaged -0.28 for the ten map task extracts. For three of the ten, the correlation was significant at $p<.01$. The average correlation was -0.19 for the broadcast extracts. In six out of ten, the correlation was significant at $p<.01$. Both speech styles thus coincide in disfavoring a boundary preceding a prominent word, with a significant result occurring slightly more often for the broadcast extracts. This result is somewhat surprising, as it was expected that the broadcast speakers might make greater use of initial accents, which should result in more boundaries preceding prominent words. But this does not seem to occur in these data.

3.3. Listeners' patterns of marking prominence and boundaries

Listeners marked boundaries less often than prominences. The global median was one boundary marked every 9.7 words, and one prominence every 8.6 words. The tendency for more frequent marking of prominence than boundaries held true for 13 of the extracts. More frequent marking of prominence means that in some cases, a listener marked more than one word as prominent within the span delimited by two boundary-markings.

The overall rate of marking obtained here is somewhat lower than the means of one boundary per 7.4 words and one prominence per 7.5 words reported by Mo et al. (2008) for English. This difference may be due to the fact that the French listeners heard each extract only once, while Mo et al.'s American listeners heard their extracts twice, and thus could add additional markings as well as changing those they had marked on the first repetition.

3.3.1. Intervals between marked prominences

Combining all listener responses, the median interval between words marked as prominent was 8.6, and the mean 10.5. For the map tasks, the mean for each extract, averaging across all listeners, ranged from 4.3 to 18.4 words. For the broadcast extracts, the means varied from 9.2 to 16.0 words. Recall that for the map tasks, each extract is from a different speaker, and for the broadcast extracts, there are two extracts per speaker, so most of the variation among extracts could also be interpreted as variation among speakers.

Listeners varied in their behavior, also. Averaging across all extracts, the mean interval between marked prominences ranged for individual listeners from 3.1 to 27.3 for the map task extracts, and 3.6 to 23.5 for the broadcast extracts. As can be seen in Table 2, the great majority of listeners tended to mark prominences at intervals ranging from every four to twelve words. The other clear tendency is for prominences to be marked less frequently in the broadcast extracts. This seems slightly surprising, as the journalists participating in the broadcast discussion were arguing and emphasizing specific points, which might have led to more words being perceived as unusually prominent.

Number of words between marked prominences	Number of listeners marking prominences at different frequencies for extracts from	
	Map tasks	Broadcast
0-4	1	1
4-8	12	4
8-12	8	12
12-16	0	5
16-20	2	2
20-24	1	1
24-28	1	0

Table 2. Frequency count of number of listeners who marked prominences at different frequencies (intervals in numbers of words)

3.3.2. Intervals between marked boundaries

The intervals between locations marked as boundaries might be expected to delimit chunks of speech that correspond to a phrasal unit. Investigating the size of these chunks should shed light on the type of unit(s) that listeners are perceiving. The locations of boundary-marking will be discussed both in terms of the responses of individual listeners, and in terms of consensus marking, that is, those locations identified by at least two-thirds of listeners.

Looking at all listener responses, the median interval between boundary markings was 9.7 words, and the mean 11.5 words, combining all 20 extracts and 25 listeners who marked boundaries. The range of variation was greater among different listeners (averaging across all

extracts) than among different extracts (averaging across all listeners). For the map tasks, the mean for each extract (averaging across all listeners) ranged from 5.8 to 13.1 words. For the broadcast extracts, the mean ranged from 8.8 to 11.4 words. The range of variation among the different listeners (averaging across all extracts) was from 5.4 to 25.1 for the map tasks, and from 5.4 to 27.4 for the broadcast ones. Although the overall ranges were fairly similar for the two sets of extracts, as can be seen in Table 3, the two sets differed in that listeners tended to mark boundaries more often when listening to the map tasks than they did for the broadcast extracts. The mean interval at which boundaries were marked in the map task extracts was 10.7 words, for the broadcast extracts, 12.3 words. This difference suggests that listeners perceived slightly longer chunks in the speech of the journalists, which is plausible given the more formal speech and complex syntactic structures that they employed.

Number of words between marked boundaries	Number of listeners marking boundaries at different frequencies for extracts from	
	Map tasks	Broadcast
4-8	11	3
8-12	8	13
12-16	1	2
16-20	4	5
20-24	0	1
24-28	1	1

Table 3. Frequency count of number of listeners who marked boundaries at different frequencies (intervals in numbers of words)

The number of boundaries within an extract that were agreed on by at least two-thirds of the listeners ("consensus boundaries") varied from two to twelve. This closely correlates with the number of words in the extract ($r = 0.90$). The broadcast extract with the greatest density of consensus boundaries, indicated by a single vertical line, is shown here. These boundaries were agreed on by 72-96% of listeners.

(2) il y a une contradiction dans ce que vous dites | sur le la disparition de la publicité sur le service public | grosso modo la la la publicité disparaît | et la le service public perd en indépendance | présenter ju- juste je voudrais revenir là-dessus | parce que c'est un discours que notamment les patrons de presse | quand je dis les patrons de presse c'est les directeurs de rédaction | tiennent à toutes leurs équipes | je suis bien placé pour le savoir je pense que tu t'es un peu au courant aussi | c'est à dire que la garantie de l'indépendance d'un journal | c'est sa bonne santé économique | elle passe par la publicité aujourd'hui c'est à peu près cinquante cinquante entre les ventes d'un journal <et la publicité> [Extract B-2]

3.4. Acoustic properties of locations marked as boundaries

The consensus locations where at least two-thirds of listeners agreed on the presence of a boundary were examined to see if they shared any salient acoustic properties that might have stimulated the listeners to agree on the boundary. The first property investigated was whether boundaries were marked at the location of pauses. All pauses with duration greater than 150 ms were identified. These included silent pauses, filled pauses, breathing, or a combination of these. The duration of 150 ms was chosen as it is longer than a silence due to, for instance, the insertion of a glottal stop. Of the locations where pauses occurred, 59% in the map task extracts and 57% in the broadcast extracts were identified as consensus boundaries. The

average boundary score at the locations of pauses was 0.69 for the map tasks and 0.62 for the broadcast, compared to average boundary scores of 0.12 and 0.10, respectively, over all words. Thus, locations where pauses occur are definitely favored as locations for boundaries but are not a reliable indicator.

An attempt was then made to divide the extracts into Intonational Phrases (Nespor and Vogel 2007). This analysis is somewhat speculative, as definitions of Intonational Phrases (IPs) are far from explicit. D'Imperio et al. (2007:2) comment that they are defined in "a rather fuzzy way as a unit showing … 'melodic cohesion'." This notion of "melodic cohesion", coupled with what might be called "rhythmic cohesion", were the main criteria used for delimiting IPs in these speech samples. A clear break and re-start in the intonation or perceived rhythm of the speech was taken as the boundary of an IP. Since the speech examined here is spontaneous, a relatively small proportion of it consists of grammatically complete sentences. When there was a complete sentence, it was taken as coinciding with the end of an IP only if there was also some phonetic evidence of finality at the same location, such as a pause, lengthening, glottalization or an abrupt break in F0. In order to determined whether the end of the speech extract should be identified as an IP boundary, reference was made to the longer recordings from which these extracts were taken. In fact, all but one of the extracts did end at an IP boundary. The one extract was cut off at a point where another speaker interrupted, but the speaker in the extract continued talking simultaneously. Data on the relation between locations marked as IP boundaries and the locations that listeners perceived as pauses are given in Table 4.

	map tasks	Broadcast
total labeled as IP boundary by experimenter	82	175
total marked as boundary by at least $^2/_3$ of listeners	35	85
total labeled as IP boundary by experimenter and marked as boundary by at least $^2/_3$ of listeners	35	78
total labeled as IP boundary by experimenter but not marked as boundary by at least $^2/_3$ of listeners	47	97
total marked as boundary by at least $^2/_3$ of listeners but not labeled as IP boundary	0	7

Table 4. Number of locations marked as boundaries by more than two-thirds of listeners and locations identified as an IP boundary by the experimenter.

From the table it can be seen that far more IP boundaries were marked by the experimenter than there were consensus boundaries identified by the listeners. To a large extent (except for seven cases in the broadcast extracts), the consensus boundaries are a subset of the IP boundaries. Even though many IP boundaries were not marked by a consensus of the listeners, the average boundary score for IP boundaries (0.58) was far above the score for the rest of the extracts (0.03). A working hypothesis is that the locations marked as IP boundaries are potential sites for listeners to mark boundaries; those marked by a consensus of listeners are the most salient of these.

Looking just at the locations identified as IP boundaries, what acoustic patterns might have contributed to certain of these being perceived as more salient by the listeners? As discussed above, the presence of a pause globally tended to favor listeners marking a boundary. When the analysis is restricted to locations identified as IP boundaries, this becomes even clearer. (Recall that pauses were not a necessary or sufficient condition for marking an IP boundary; 11 locations where pauses occur were not marked as IP boundaries). Nonetheless, 92% (for the map tasks) and 79% (for the broadcast) of the IP boundaries that a consensus of listeners

marked as boundaries, did coincide with pauses. Since many IP boundaries that coincided with pauses were not marked as boundaries by the listeners, it seems that pauses are highly desirable but not sufficient cues for a boundary.

Another acoustic characteristic that was examined at IP boundaries is the presence of a salient high pitch during the last word immediately preceding the boundary. Pitch rises are commonly found at the end of an accent group that is non-final in the utterance, with the final high anchored to the last full syllable in the phrase (Welby 2006). In the present study, analysis was limited to the word level (not the syllable level), so the analysis tested for the presence of a high during the last word (which necessarily includes the last syllable of the phrase). Testing to see if the presence of an H favors the marking of a boundary has the major disadvantage of ignoring the fact that utterance-final boundaries would typically be produced with a pitch fall, and if the fall began before the last word, the analysis used here would not capture the presence of a H tone. However, the extracts being analyzed were all taken from a single conversational turn. Within a turn, speakers may end an utterance with a H tone to signal that they are holding the floor. Thus in these data, most phrases end with a H tone on the phrase-final word. In the map task extracts, there was a H tone during 73% of IP-final words that listeners identified as preceding a boundary, but H tones occurred during only 34% of IP-final words that were not marked as preceding a boundary. In the broadcast extracts, H tones were found during 65% of IP-final words identified as pre-boundary, but only 32% of those not marked as pre-boundary. Thus, like pauses, the presence of a H tone seems to favor boundary identification by listeners, but is in no way a necessary or sufficient condition.

There were seven locations where a consensus of listeners marked a boundary, but the experimenter did not mark an IP boundary. These occurred at the end of a clause or major syntactic phrase, but with a continuous intonation contour and no pause, lengthening, glottalization or other interruption to the rhythm. In all but one of these, there was a substantial pitch rise just before the location marked as a boundary. Impressionistically, these occurred when the speaker was trying hard not to lose the floor. (All of these cases occurred in the broadcast extracts; in this discussion the speakers frequently interrupted each other.) Presumably listeners marked a boundary because they noticed the syntactic boundary, but no IP boundary was marked by the investigator because of the absence of prosodic indicators. These cases, along with other evidence summarized below, suggest that syntactic structure was a significant influence on listeners' boundary marking.

3.5. The role of syntax in listeners' response patterns

An account of the influence of syntactic structure on the listeners' responses is beyond the scope of this paper. Nonetheless, it clearly is an important factor. As suggested in the previous section, listeners sometimes marked boundaries at locations where there was no obvious prosodic cue to a boundary, but where there was a syntactic boundary. An example of syntax possibly dominating phonetic cues in listeners' responses occurs in the following sample from one of the broadcast extracts. The values in parentheses are the boundary scores for those locations. The square brackets mark locations identified as IP boundaries.

(3) je veux bien le croire (0.24)][je veux bien en être un représentant (0.84) mais (0.28)] [vous êtes certain (0.56) que les sujets sont (0.32)][sont fermés (0.72)][et …

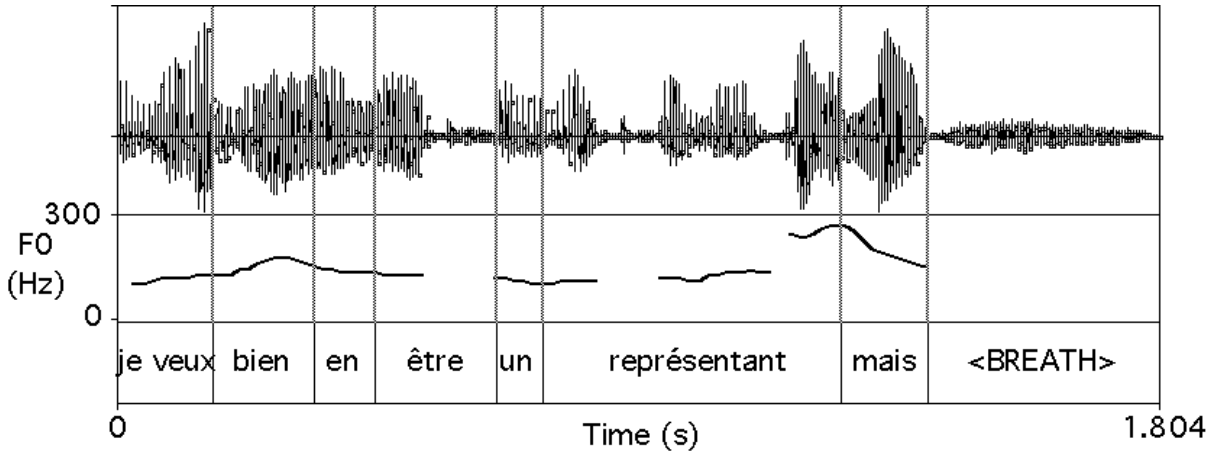

Figure 3. Acoustic waveform and F0 trace for part of the utterance shown in (3). The figure shows the breath but not the filled pause that followed "mais".

Note that far more listeners marked a boundary after "représentant", which ends a major clause, than after "mais", which was followed by a breath and filled pause that in total lasted 1.403 seconds (entire duration not shown in the figure). This distribution of responses suggests that listeners were marking structural (in this case, syntactic) boundaries, rather than just attending to interruptions in the flow of speech. No IP boundary was marked after "représentant" because the absence of lengthening and lack of any pause suggests that F0 rise is marking only an accent group boundary. The listeners who marked a boundary after "mais" were presumably attending to the pause. Overall, it seems likely that both syntactic structure and phonetic factors influenced listeners' responses in this experiment, as would be expected, since prosodic structure is considered to derive from both syntactic structure and properties of the actual production (pauses, intonation contour, speech rate, lengthening, etc.), as well as other factors such as pragmatics (Shattuck-Hufnagel and Turk 1996).

4. Discussion

The results for French obtained in this experiment resemble those for English obtained by Cole et al. (to appear) more closely than was expected. The most striking similarity is the higher rates of agreement for the marking of boundaries than for the marking of prominences that were obtained not only in the current experiment, and in Cole et al.'s work, but also by Buhmann et al. (2002) for Dutch. The rates of marking boundaries and prominences were very similar to each other in Cole et al.'s study, whereas the listeners in the present experiment marked boundaries less often. One important consideration is that in Cole et al.'s study, each listener marked both boundaries and prominences, but for different samples of speech drawn from the same corpus. In the present experiment, different groups of listeners marked prominences and boundaries, so conceivably the different rates of agreement and different frequencies of labeling could reflect individual differences among participants. Although possible, this explanation seems somewhat unlikely because the listeners were drawn from the same population, and assigned to the two different groups randomly.

There is good support in the results presented here for the argument that the prominences marked by the listeners must correspond to final accents, rather than initial ones. The high prominence scores for words before boundaries, and the high correlations between prominent words and following boundaries, agree in supporting theoretical accounts in which the final position before a boundary is a favored location for prominence in French. Cole et al. (2008)

did not investigate whether there was a correlation in their data between the locations of boundaries and of prominences, so we cannot directly compare the results for French with their results for English. But the fact that this question did not seem worth investigating suggests that there was no obvious relation between the two, as expected for English.

A more detailed analysis of the syntax of these extracts could help to determine the importance of syntax in conditioning listener responses. Spontaneous speech includes many syntactically incomplete fragments, so it may be difficult to determine how listeners interpret its structure. Listeners' perceptions most likely reflect a combination of syntactic and phonetic factors: the acoustic analyses suggest that pauses and strong F0 rises contributed to listeners perceiving boundaries. Further research will be necessary to disentangle the different factors that contribute to listener perceptions.

Although simple, this experiment demonstrates that untrained listeners can make their perceptions of prosody explicit in performing a meta-linguistic task with sufficient consistency to provide a useful confirmation of the connection between the two aspects of prosody, prominence and phrasing.

Acknowledgments

Merci à tous les auditeurs qui ont participé (sans récompense !) à cette expérience, et surtout, merci beaucoup à Frédérique Bénard pour son aide avec la transcription des extraits.

References

Beckman, M. & G. Ayers Elam (1997). Guidelines for ToBI labelling. Ms, Ohio State University. http://www.ling.ohio-state.edu/~tobi/ame_tobi/.

Brennan, R. & D. Prediger (1981). Coefficient Kappa: Some uses, misuses and alternatives. *Educational and Psychological Measurement* 41, pp. 687-699.

Buhmann, J., J. Caspers, V. van Heuven, H. Hoekstra, J.-P. Martens & M. Swerts (2002). Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus. *Proceedings of LREC 2002* (Las Palmas), pp. 779-785.

Cole, J., Y. Mo & M. Hasegawa-Johnson (to appear a). Signal-based and expectation-based factors in the perception of prosodic prominence. To appear, *Laboratory Phonology 1*.

Cole, J., Y. Mo & S. Baek (to appear b). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. To appear, *Language and Cognitive Processes*.

Di Cristo, A. (1999). Vers une modélisation de l'accentuation en français : première partie. *Journal of French Language Studies* 9, pp. 143-179.

Di Cristo, A. (2000). Vers une modélisation de l'accentuation en français (seconde partie). *Journal of French Language Studies* 10, pp. 27-44.

Di Cristo, A. (2005). Éléments de prosodie. Nguyen, N., S. Wauquier-Gravelines & J. Durand (eds.), *Phonologie et phonétique : forme et substance*. Hermès Science Publications. Lavoisier, Paris, pp. 117-157.

Di Cristo, A. & D. Hirst (1997). L'accentuation non-emphatique en français : stratégies et paramètres. Perrot, J. (ed.), *Polyphonie pour Ivan Fónagy*. L'Harmattan, Paris, pp. 71-101.

D'Imperio, M., R. Bertrand, A. Di Cristo & C. Portes (2007). Investigating phrasing levels in French: Is there a difference between nuclear and prenuclear accents? Camacho, J., V. Deprez, N. Flores & L. Sanchez, *Selected Papers from the 36th Linguistic Symposium on Romance Languages* (LSRL). John Benjamins, New Brunswick, p. 97-110.

Jun, S.-A. & C. Fougeron (2002). Realizations of accentual phrase in French intonation. *Probus* 14, pp. 147-172.

Lacheret-Dujour, A. & F. Beaugendre (1999). *La prosodie du français*. CNRS Editions, Paris.

Mertens, P. (2006). A predictive approach to the analysis of intonation in discourse in French. Kawaguchi, Y., I. Fonagy & T. Moriguchi (eds.), *Prosody and Syntax*. Usage-Based Linguistic Informatics 3. John Benjamins, Amsterdam, pp. 64-101.

Mettouchi, A., A. Lacheret-Dujour, V. Silber-Varod & S. Izre'el (2007). Only prosody? Perception of speech segmentation in Kabyle and Hebrew. *Nouveaux cahiers de linguistique française* 28, pp. 207-218.

Mo, Y., J. Cole & E. Lee (2008). Naive listeners' prominence and boundary perception. *Proceedings of Speech Prosody 2008* (Campinas). http://prosody.beckman.illinois.edu/publications.html.

Nespor, M. & I. Vogel (2007). *Prosodic phonology: with a new foreword*. Walter de Gruyter, Berlin.
Obin, N., X. Rodet & A. Lacheret-Dujour (2008). French prominence: a probabilistic framework. *Proceedings of ICASSP 2008* (Las Vegas), pp. 3993-3996.
Pagel, V., N. Carbonell, Y. Laprie & J. Vaissière (1995). Spotting prosodic boundaries in continuous speech in French. Elenius, K. & P. Branderud (eds.), *Proceedings of the XIIIth ICPhS, Stockholm*, Vol. 4, pp. 308-311.
Pasdeloup, V. (1990). *Modèle de règles rythmiques du français appliqué à la synthèse de la parole*. Diss, Université de Provence.
Pitt, M., K. Johnson, E. Hume, S. Kiesling & W. Raymond (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication* 45, pp. 89-95.
Portes, C. (2000). *Approche du rôle de la prosodie dans la structuration du discours oral en français*. DEA thesis, Université de Provence.
Post, B. (2000). *Tonal and phrasal structures in French intonation*. Diss, Katholieke Universiteit Nijmegen.
Randolph, J.J. (2008). Online Kappa Calculator. http://justus.randolph.name/kappa.
Shattuck-Hufnagel, S. & A. Turk (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25, pp. 193-247.
Smith, C. (2007). Prosodic Accommodation by French speakers to a non-native interlocutor. Trouvain, J. & W.J. Barry (eds.), *Proceedings of the 16th ICPhS*, Saarbrücken, Germany, pp. 1081-1084.
Streefkerk, B., L. Pols & L. ten Bosch (1997). Prominence in read aloud sentences, as marked by listeners and classified automatically. *IFA Proceedings* 21, Institute of Phonetic Sciences, University of Amsterdam, pp. 101-116.
Vaissière, J. (2002). Cross-linguistic prosodic transcription: French versus English. Volslkaya, N.B., N.D. Svetozarova & P.A. Skrelin (eds.), *Problems and methods in Experimental Phonetics*. In honour of the 70th anniversary of Prof. L.V. Bondarko. St. Petersburg State University, St. Petersburg, pp. 147-164.
Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics* 34, pp. 343-371.