CROSS-LINGUISTIC STUDY OF FRENCH AND ENGLISH PROSODY
F_0 Slopes and Levels and Vowel Durations in Laboratory Data

Katarina Bartkova and Mathilde Dargnat
University of Lorraine, ATILF, France
e-mail: {katarina.bartkova, mathilde.dargnat}@atilf.fr

Abstract

Prosody conveys linguistic and extralinguistic information through prosodic features which are either language dependent or language independent. In addition, each speaker has unique physiological characteristics of speech production and speaking style, and thus speaker-specific characteristics are also reflected in prosody. Distinguishing the language-specific and speaker-specific aspects of prosody using acoustic parameters is a very complex task. Therefore, it is very challenging to extract and represent prosodic features which can differenciate one language from the other or one speaker from the other. The goal of our study is to investigate whether the prosody of isolated sentences in French and English is determined by their shared syntactic structures and whether the prosodic features used by the two languages are different or similar. In our cross-linguistic comparison of the prosodic parameters, two approaches are used. First, F0 slopes measured on target words in the sentences are analyzed by fitting mixed linear regression models (R package *lme4*). Secondly vowel duration and F0 values for each syllable are prosodically annotated using an automatic prosodic transcriber and the symbolic and numeric values are used in a more qualitative comparison of our data. It appears from the analyzed data that the observed F0 curves in our corpus do not always correspond to linguistic theory and that the output of the automatic prosodic transcriber provides relevent information for a cross-linguistic study of the prosody.

1 Introduction

Prosody is an important component of oral communication for transferring linguistic, pragmatic and extralinguistic information and gives the speech signal its expressiveness mainly through melody, intensity and sound duration. Variation of the prosodic parameters allows a listener to segment the sound continuum, and to detect emphasis on the speech signal (i.e., accent of words or expressions). The prosodic component of speech conveys the information used for structuring the speech message, such as emphasis on words and structuring the utterance into prosodic groups.

However the prosodic component of the speech signal is less easy to process than its segmental part as there are few constraints in the realization of its parameter values. Yet, prosodic information is difficult to add into the manual transcription of speech corpora, or other automatic speech processing. Hence, it is important to investigate automatic approaches for recovering such information from speech material.

Even if not perfect, the use of an automatic approach for prosodic annotation of the speech would be very useful especially as the agreement on manually annotated prosodic events (boundary levels, disfluences and hesitation, perceptual prominences) between expert annotators is quite low (68%). Even after training sessions, the agreement does not exceed 86% (Lacheret-Dujour et al., 2010) and the task can be considered even more difficult and complex when manual coding of pitch level is to be carried out. In fact, it is difficult for human annotators not to be influenced by the meaning of an utterance; annotators can be tempted to associate a prosodic boundary at the end of a syntactic boundary or at the end of a semantic group instead of focusing solely onto the prosodic events. Moreover, there can be a discrepancy between the parameter values and their perception by a human annotator. For instance, an acoustic final rise can be perceived as a fall depending on the preceding F0 curve (Hadding-Koch and Studdert-Kennedy, 1964). Moreover the same F0 contours can have non-standard occurrences (F0 rises can be found at the end of declarative sentences) and a human transcriber may be influenced by what he considers as being the norm, and standardize the transcription of prosodic phenomena, ignoring what he sees and what he hears.

A further advantage of an automatic processing is that, once the values of the parameters are normalized, they are then always compared to the same threshold values. This process is extremely difficult to follow when human (hence subjective) annotation is concerned.

The goal of the present study is to test an automatic approach for prosodic labeling in a cross-linguistic study of speech prosody in French and English. We use an automatic system, PROSOTRAN, in this study. This program is well adapted for annotation of languages, such as French, in which the syllable duration is one of the major parameters of stress. PROSOTRAN is able to annotate the prosody of sentences in French and English containing the same syntactic structures.

2 Prosodic annotation

Prosodic parameters are subject to a prosodic coherence governing parameter values across the prosodic group. It was observed in automatic speech synthesis (in diphone and data driven approaches) that a sudden unjustified change in f_0 or sound duration (beyond stressed syllables or prosodic junctures), is perceived either as a corruption of the speech signal or as an occurrence of a misplaced contrastive stress (Boidin, 2009). Most of the time transcribers focus on the transcription of parameter values of syllables considered as linguistically prominent, carrying pertinent linguistic information. The other syllables, linguistically non-prominent, remain

generally uncoded, although their prosody contributes to an overall perception of a correct pattern. Therefore, in order to keep a faithful prosodic transcription of the speech signal, all syllables should receive annotation of their different parameters. Moreover, some f_0 changes that can be perceptually crucial may not be transcribed in an appropriate way. Thus, a final f_0 rise generally indicates a question, an unfinished clause, or an exclamation, but it can also occur at the end of statements in spontaneous speech. A phonological transcription should avoid using one and the same symbol for these cases (for example, H%), as these types of rises, which may sometimes correspond to the same f_0 contours, are perceptually distinguished (Fónagy and Bérard, 1973).

Prosodic annotation is a complex and difficult task and linguists and scientists working in speech technology address this issue from various angles. A distinction can be made between phonological approaches (Silverman et al*.,* 1992; Hirst, 1998; Delais-Roussarie, 2005; etc.) and acoustic-phonetic prosodic analysis (Beaugendre et al., 1992; Mertens, 2004). Most of the prosodic transcription systems capture levels (*extra high, high, mid, low, extra low*) and movements of the f_0 values (*rising*, *falling*, or *level)*, or integrated F0 patterns (*Hat pattern*,…).

The prosodic transcription system, **ToBI** (Tone and Break Indices) (Silverman et al., 1992; Beckman et al., 2005), is often considered as a standard for prosodic annotation. However, ToBI appears to be a somewhat hybrid system. It is based on Pierrehumbert's abstract *phonological* description of English prosody (Pierrehumbert, 1980), but is often considered as a *phonetic* transcription, using the perception of the melody for its symbolic coding and the visual observation of the evolution of f_0 values.

INTSINT (an INternational Transcription System) is a production-oriented system. This system is a relatively language independent one; it has been used for the description of F0 curves in several languages (Hirst and Di Cristo, 1998). A limited number of symbols are used to transcribe relevant prosodic events. These include absolute (*Top, Mid, Bottom*) or relative (*Higher, Lower, Same, Upstepped, Downstepped*) designations. The limitations of the system stem from the use of the f_0 values alone.

Other approaches should be included to complete our short overview of prosodic annotations. The syntactic-pragmatic approach of French intonation integrates a morphological approach, where the intonation is built from sequences of prosodic morphemes, (Focus, Theme, Topic…) (Rossi, 1999). Another interesting approach to prosody is an abstract representation of relational "holistic gestalts", which integrated tonal and temporal whole word profiles, with pitch range variations. This type of system is well adapted to the representation of attitudinal patterns (Aubergé et al., 1997).

3 Cross linguistic study

The use of prosodic parameters is common in all the languages, but some of the uses are language independent. There are universal tendencies (Bolinger, 1978), but

also distinctions in intonational structure between languages ("semantic", "phonotactic", "pragmatic"…) (Ladd, 1996; Crystal, 1969). The comparison of the prosodic parameters among languages is very challenging precisely because of the universality and language specificity of prosody. This is especially true for Germanic (e.g., Dutch, English, German) and Romance languages (e.g., French, Italian, Spanish) (Hirst and Di Cristo 1998, Ladd 1996). Therefore, in order to conduct multi-language comparisons, several kinds of prosodic transcription should to be used: an acoustic-phonetic one (broad and narrow), a perceptual transcription for the perceptually relevant events in duration, intensity and melody, a phonological transcription, and a functional transcription.

3.1 French & English prosody

French uses a combination of segmental and tonal cues to signal prosodic phrases, and differs in this respect from a language like English, which relies almost exclusively on tonal boundaries (Gussenhoven, 1984). In French, lexical stress is mostly quantitative (Delattre, 1938), and the final syllable is the one which undergoes a potential lengthening. However, lengthening of the last syllable in a French word corresponds to final (pre-boundary) lengthening, which affects rhythm, and is not an accentual lengthening as in English (Campbell, 1992).

French is generally considered as a language with mostly 'rising' f_0 patterns accompanied by a lengthening of final syllables. According to Vaissière (2002), the French ear is trained to perceive rising *continuation* F0 patterns at the end of prosodic phrases: each prosodic phrase inside a sentence tends to end with a high rise (Delattre's *continuation majeure*), or a smaller rise (Delattre's *continuation mineure*). In Delattre's theory of French intonation, a categorical difference in intonation patterns is expected between minor and major continuation patterns, which are syntax-dependent. Furthermore, according to Delattre, major continuation patterns are only rising, whereas minor continuations can show rising or falling patterns. Prominence is not lexically driven in French (i.e., there is no lexical stress), but it is determined by prosodic phrasing (Delais-Roussarie, 2000).

3.1.1 F0 contours. French and English intonations are sometimes described by a set of contours. Delattre (1966) identified 10 basic contours that can describe the most frequent intonation patterns in French. Post (2000) also listed 10 contours although these contours differ from those proposed by Delattre. As far as English is concerned, 22 pertinent intonation contours are proposed by Pierrehumbert (1980) to describe English intonation.

It is common to use the term *assertion intonation* or *question intonation* to refer to falling or rising contours. Falling contours are associated with assertion or assertiveness (Bartels, 1999), whereas rising contours are associated with questions or aspects of questioning (uncertainty, ignorance, call for a response or feedback from the addressee, etc.). Although prototypical assertions are uttered with a falling contour and that prototypical confirmation or verifying questions are uttered with a rising contour, occurrences of assertions with a rising contour and occurrences of

confirmation or verifying questions with a falling contour are far from rare in everyday conversations (Beyssade et al, 2003).

In the following paragraphs, F0 contours in French and English sentences are measured and compared.Their difference was statistically evaluated.

3.2 Corpus

The corpus used in this study was recorded as a part of project Intonal, which focuses on intonation in French and English. The project was conducted by the University of Nancy2 and the LORIA research laboratory (2009-2012). The recorded corpus contains 40 short sentences belonging to 8 syntactic categories which were recorded by 20 French and 20 English native speakers. In a previous study, two prosodic parameters associated with f_0 slope were calculated for some target words in sentences. These words are bolded and underlined in the following sentences:

- (CAP). Continuative configuration at the end of the first clause in a two clause sentence, without any coordinating conjunction: "Il dort chez Maria, il va finir tard. / He'll sleep at Maria's, he'll finish late."
- (CAO). Continuative configuration at the end of the first clause in a two clause sentence, with a coordinating conjunction: "Il dort chez Maria car il finit tard. / He'll sleep at Maria's because it's too late."
- (CIS). Continuative configuration on a subject NP: "Les agneaux ont vu leur mère. / The lambs have seen their mother."
- (CIA). Continuative configuration on a NP subject in the first clause of a two clause sentence:
  "Nos amis aiment Nancy parce que c'est joli. / Our friends really like Nancy because it's pretty."
- (QAS). Question configuration at the end of a clause: "Il dort chez Maria? / Will he sleep at Maria's?"
- (QIS). Interrogative configuration on a simple subject NP: "Qui a appelé? Nos amis? / Who has phoned? Our friends?"
- (DIS). Short declarative sentence "Nos amis. / Our friends".
- (DAS). Longer declarative sentence: "Il dort chez Maria. / He'll sleep at Maria's".

Two kinds of non-conclusive f_0 slope configurations were studied here at two levels. First, on the syntactic level: the slope of the final segment of a subject NP in a declarative sentence, followed (CIA) or not (CIS) by another sentence. Second, on the discourse level: the slope of the final segment of A in a two clause utterance AB, where A and B are declarative clauses connected by a discourse relation, marked (CAO) or not (CAP) by a conjunction.

These sentences were used to investigate whether the intonation of the target words is realized in a similar manner in both English and French and whether:

- there is a significant difference between major continuation curves (expected in CAO and CAP sentences) and minor continuation curves (expected in CIA and CIS sentences).

- continuative rising slopes (expected in sentences CAO, CAP, CIA & CIS) are different from interrogative slopes (measured in QIS & QAS) sentences
- continuative falling slopes (measured in CIA and CIS types of sentences) are different from declarative slopes (measured on declarative sentences DIS & DAS).

3.3 Segmentation and annotation of the speech signal

In order to segment our speech data, knowing the orthographic transcriptions, a text-to-speech forced alignment was carried out using the CMU sphinx speech recognition toolkit (Mesbahi et al., 2011). This provided an automatic segmentation of the speech signal at the phoneme level. The automatic segmentation of each speech signal was then manually checked by an expert phonetician using signal editing software. Intonation slopes were computed as regression slopes (RslopeST) using f_0 values in semitones, which were estimated every 10 ms. Slopes were calculated on the last two syllables of the target segments (in underlined bold characters in 3.2) of every sentence.

3.1.1 Statistical analysis. f_0 slope data are analyzed by fitting mixed linear regression models (R package *lme4*). Using this approach, one can contrast the different configuration types and show the differences that are significant and those that are not (function *glht*, package *multcomp*).

The statistical analysis showed that in French, sentences where we expect minor f_0 patterns, continuation patterns (CIA-CIS sentence types) are mostly rising (95%). The major continuation sentence types (CAP-CAO) also have rising f_0 slopes (59%); but there is a significant difference between sentences with coordinating conjunctions (CAO), containing 73% of rising f_0 slopes, and paratactic (CAP) sentences containing only 46% of rising F0 slopes.

In the English data, the f_0 slopes measured in minor continuation (CIA-CIS) sentence types can rise (53%) and fall (47%) equally. In major continuation (CAP-CAO) sentence types, f_0 slopes are seldom rising (21%) and there is no marked difference between f_0 slopes in sentences with coordinating conjunctions (CAO, 18% of rising patterns) and f_0 slopes in paratactic sentences (CAP, 24 % of rising patterns).

In the French corpus, slopes measured on minor continuation (CIS-CIA) sentence types are not significantly different from juxtaposed sentence types where major continuation slopes (CAO) are expected, although they are significantly different from slopes measured on sentences with coordinating conjunctions (CAP) [see Figure 1 (left)]. Neither is there a significant difference between slopes measured on these two sentence types (CIA-CIS) (where minor continuation slopes are expected). However, the slopes of the latter are significantly higher than the slopes measured on short declarative sentences (DIS) and significantly lower than the slopes measured on simple subject NP questions (QIS). On the other hand, slopes measured on juxtaposed sentences (CAP) are significantly lower than those measured on sentences with a coordinating conjunction (CAO).
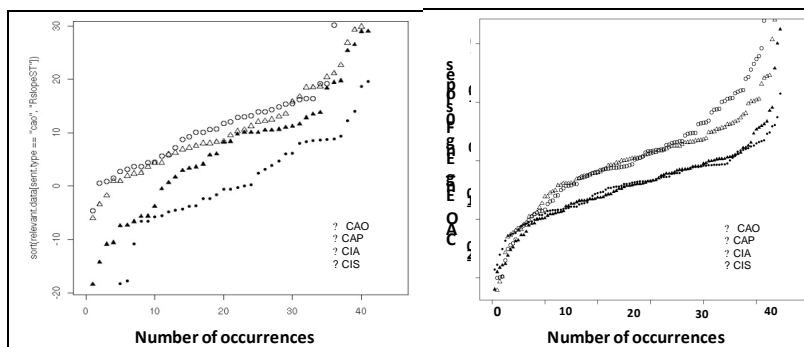
Figure 1. F_0 slope values for the French (left) and English (right) corpora in 4 sentence types. The *Y axis* corresponds to RslopeST value (RslopeST = slope of the regression line of the pitch data points in semitones) and the *X axis* to increasing ordering of observations (each point is an observation).

In the data recorded by English speakers, slopes of minor continuation sentence types (CIA-CIS) are significantly higher than slopes measured on major continuation sentence types (CAO-CAP) and are also significantly higher than slopes measured on short declarative sentences (DIS). However, no significant difference was found between minor continuation slopes (CI) and slopes measured on short questions (QIS). English speakers do not utter juxtaposed sentences (CAP) differently than sentences containing coordinating conjunctions (CAO) (see Figure 1 (right)). Furthermore, major continuation slopes (CAP-CAO) are not significantly different from slopes measured on longer declarative sentences (DAS) and interrogative (QAS) sentences (Bartkova et al., 2012).

3.4 Additional analyses using automatic annotations

As it appears from the previous analysis of the obtained results, the syntactic differences among the sentences studied are not necessarily marked, as expected by theory (Delattre, 1966) or by prosodic means, and there are not systematic and significant differences among the rising and falling f_0 slopes used. However, pertinent prosodic differences among these syntactic structures can be scattered all along the utterances and they are not necessarily concentrated on the final syllables of the target words alone. In order to compare the different syntactic structures and their prosody in a more precise way, and to conduct a deeper cross linguistic comparison of the prosody among French and English sentences, a subset of the data was annotated by our PROSOTRAN automatic annotation tool and the results of the obtained annotations were analyzed and discussed in the paragraph below. The corpus used was comprised of one sentence for each sentence type uttered by about 10 French speakers (as not all the speakers uttered all the sentences) and about 20 English speakers (all speakers uttered all sentences).

3.4.1 Speech data processing. The speech data processing used in this part of our study had 4 different stages. During the first stage, prosodic parameters are extracted from the speech signal. In the second stage, prosodic annotations are yielded by our

annotation tool PROSOTRAN using the extracted parameters and these parameters are hand checked by phoneme segmentation, as in our previous speech data processing (see 3.3). In order to check whether our annotation is faithful or not, the third processing stage recalculates the numerical f_0 values from the prosodic annotation and during stage four, the prosody of the speech signal is resynthesized using Praat (and the PSOLA technique). The resynthesis of the melody allows for checking whether or not the quality of the obtained signal was corrupted by the previous prosodic parameter manipulations.
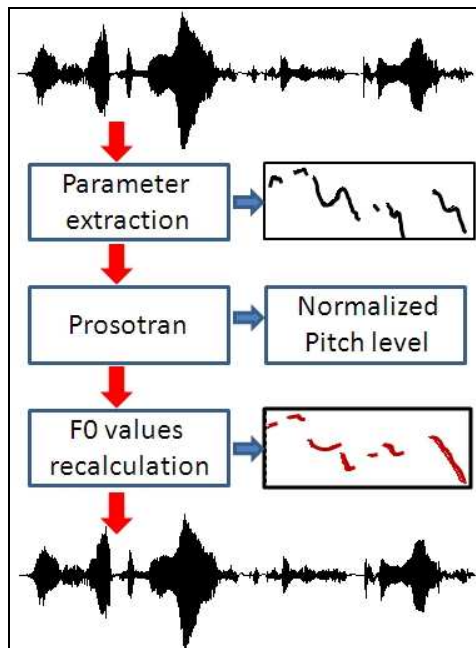
Figure 2. Illustration of the 4 stages of our prosodic processing: (1) parameter extraction, (2) prosodic labeling with PROSOTRAN, (3) f_0 value recalculation, and (4) resynthesis with the recalculated F0 values.

3.4.2 Parameter extraction. Acoustic parameters, such as f_0 in semi-tones and log energy, are calculated from the speech signal every 10 ms with the Aurora front-end (Speech Processing, 2005). The forced alignment between the speech signal and its phonetic transcription provides phoneme durations, as well as the duration of the pauses. Synchronization between the phoneme units and their acoustic parameters (f_0 and log energy values) is carried out and prosodic parameters are calculated for every relevant phoneme.

3.4.3 PROSOTRAN. Our annotating tool, PROSOTRAN, is a system enabling automatic annotation of prosodic patterns. Since all linguistically relevant prosodic events are realized at the phonetic level by some sort of changes in the prosodic parameters, PROSOTRAN assigns a symbolic label to every syllabic nucleus for each prosodic parameter separately. The resulting annotation is multitiered, with

each tier being associated with a single parameter. PROSOTRAN encodes vowel duration, vowel energy, f_0 slope movement, f_0 level, delta f_0 values and some more information concerning the f_0 curve either symbolically or numerically. However, as for our cross linguistic study, only vowel duration and vowel f_0 levels are used, therefore only the calculation and coding of these parameters are explained in the following paragraphs (for more information about PROSOTRAN, see Bartkova et al., 2012).

3.4.3.1 Duration. Although the temporal axis of the speech signal is represented by all sound durations, PROSOTRAN uses only vowel durations in its prosodic annotation. This avoids the issue of syllabic structure variability, and vowel duration is considered to be more homogeneous and therefore more representative of speech rate variation than syllable duration (Di Cristo, 1985). Moreover, vowel nuclei constitute the salient part of the syllable and are hence the most important speech element used to convey the prosody (Segui, 1984).

In the French corpora processing, each vowel duration was compared to the mean duration and associated standard deviation of the vowels occurring in non-final positions (i.e. not at the end of a word nor before a pause) when measured on the speech data uttered by the same speaker. This way, stressed vowels whose duration is lengthened (vowel duration is one of the major prosodic parameter of French stressed vowel) are discarded from the calculation of the mean and standard deviation values. In the English corpora processing, the vowel durations are compared to the mean duration and standard deviation of all the vowels of all the speech material produced by the same speaker.

To represent sound durations, symbolic annotations are used, representing duration from extra short duration (Voweldur----) to extra long duration (Voweldur++++).

3.4.3.2 F0 range and levels. In order to represent the speech melody, a melodic range was calculated between the maximum and the minimum values of the f_0 in semi-tones. For each speaker, all speech material was used to build a histogram of the distribution of the f_0 values. To avoid extreme, often wrongly detected f_0 values, 6% of the extreme f_0 values (3% of the highest and 3% of the lowest ones) were discarded. The resulting range was then divided into several zones (9 in our case) and coded into levels (from 1 to 9). f_0 slopes were calculated for vowels and semi-vowels.

Results of the annotation are stored in text files and also in TextGrid files to make possible visualisation by Praat (see Figure 3 for annotation examples).
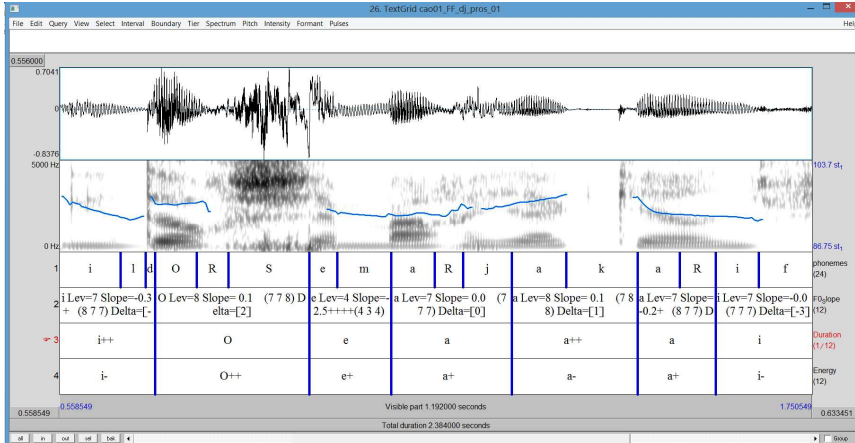
Figure 3. Example of the prosodic labeling provided by the PROSOTRAN tool

3.4.3.3 F0 level normalization. In order to compare the f_0 patterns of our French and English data, the f_0 level annotation produced by PROSOTRAN was used. However, to minimize the overall range differences among the speakers for a sentence type, f_0 level normalization of the different speakers was carried out. To obtain normalized f_0 level values, the f_0 pattern of one of the speakers was taken as a reference, and all other speaker f_0 patterns were adjusted in order to minimize the Euclidean distance between the individual speaker f_0 pattern and the reference pattern. Normalized f_0 levels were computed for each sentence and for each speaker.

Once the f_0 levels for all vowels were normalized by sentence type, a mean f_0 level value was calculated for each sentence type syllable to yield one representative f_0 level pattern of per sentence type (see Figure 4). Using this single representative f_0 level pattern per sentence enable us to compare the f_0 patterns of the French and the English sentence types and to carry on our cross linguistic study of the prosody.
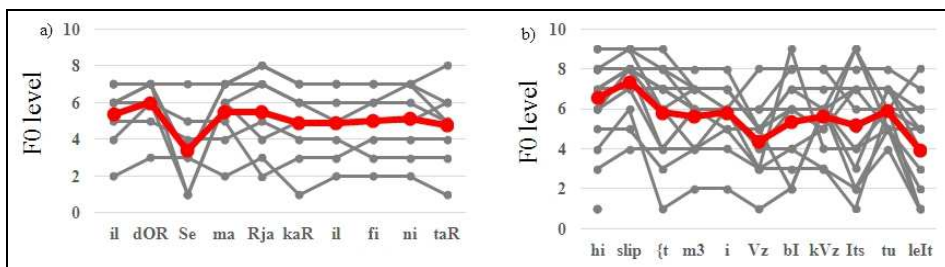

Figure 4. Calculation of a representative f_0 level pattern for a French (a) and an English (b) sentence.

As mentioned before, the duration of each vowel was annotated symbolically. Using these symbolic annotations, a numeric coefficient was calculated expressing the degree of vowel lengthening produced by different speakers. Thus the coefficient value α indicates that the duration of a given vowel is on avarege equal to the mean duration value plus α times the standard deviation. A low value coefficient indicates

44

that the vowel was largely lengthened by only a few speakers or that the vowel was lengthened slightly by a large number of speakers.

3.5 Result analysis and discussion

The following figures contain the representative f_0 level patterns for the different sentence types. The circles indicate the prominent f_0 levels and the numbers show the vowel lengthening coefficient. Coefficients are indicated only for vowels whose duration was longer than the mean duration and greater than one times the standard deviation.
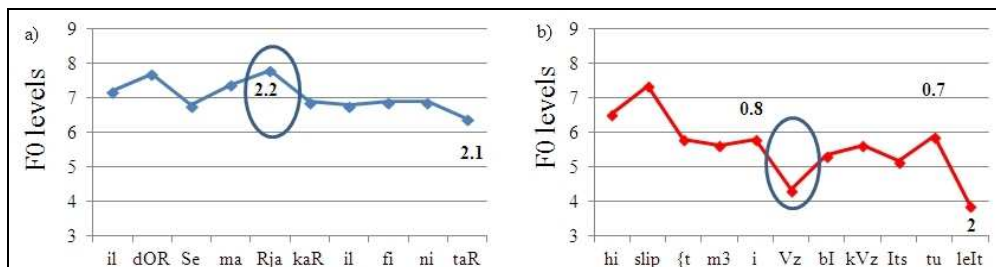
Figure 5. CAO - Continuative configuration at the end of the first clause in a two clause sentence, with a coordinating conjunction: (a) *Il dort chez **Maria** car il finit tard.* (b) *He'll sleep at **Maria's** because it's too late.*

For the continuative sentence types (Figure 5) French speakers marked the continuation with a rising f_0 while English speakers prosodically coded the same syntactic boundary with a lowering F0. In French, the general rising tendency of the f_0 was not very high but the prosodic boundary also was indicated with a lengthened vowel duration (high duration coefficient). On the other hand, the downwards movement of the f_0 in English was more important but there is no vowel lengthening in the final syllable. The sentence final f_0 movement was falling in the twoboth languages but the slope was steeper in English than in French.

In French paratactic sentences (Figure 6), the mean F0 level pattern contained a slight f_0 rise on the prosody boundary and the vowel duration was lengthened (even more than in the previous sentence) in the boundary final syllable. French speakers give preference to upward (though moderate) movement of the f_0 on the prosodic boundary, while the majority of the English speakers favor downward movement of the f_0 curve. In French, the inter-utterance prosodic boundary was marked by a lengthening of vowel duration, while in English the utterance final f_0 level was very low and the vowel duration was very clearly lengthened.

In two clause sentences with a continuative configuration (Figure 7), most French and English speakers realized a high level f_0 at the end of the noun phrase subject. But neither French nor English speakers used vowel duration to highlight the prosodic boundary. However, the second prosodic boundary of the sentence, although marked with a lower f_0 level, contained lengthened vowel durations. In English, the final boundary f_0 level was very low (level 3) and the vowel duration

was strongly lengthened. In French, the final prosodic boundary had a relatively high F0 level (level 7), but the final vowel lengthening was moderate.
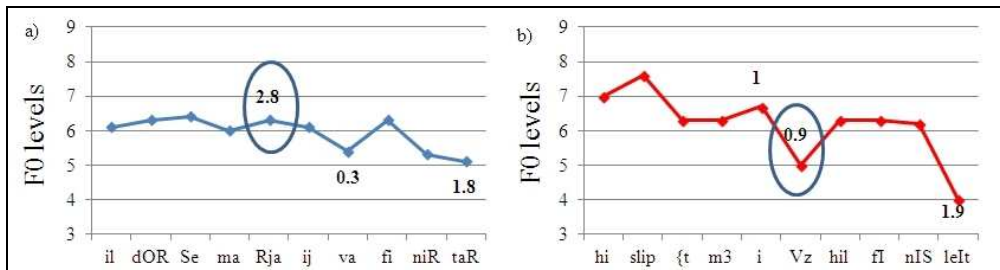

Figure 6. CAP - Continuative configuration at the end of the first clause in a two clause sentence, without any coordinating conjunction: (a) *Il dort chez **Maria**, il va finir tard.* (b) *He'll sleep at **Maria's**, he'll finish late.*
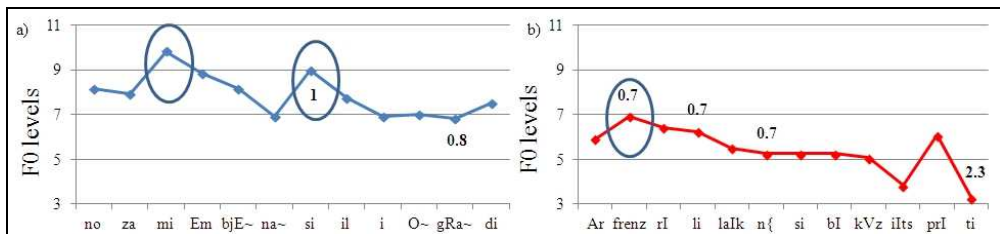

Figure 7. CIA - Continuative configuration on a NP subject in the first clause of a two clause sentence: (a) ***Nos amis*** *aiment Nancy ils y ont grandi.* (b) ***Our friends*** *really like Nancy because it's pretty.*

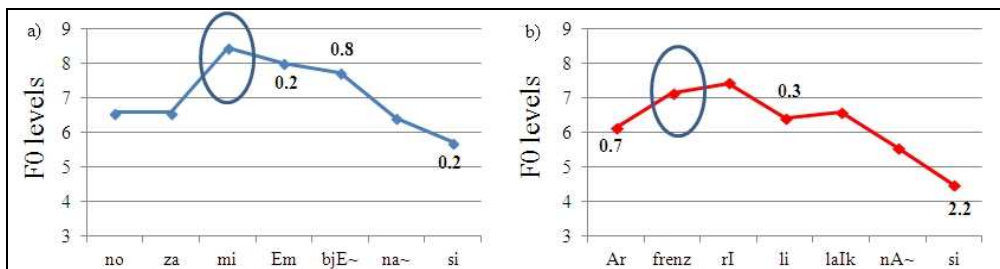*Figure 8.* CIS - Continuative configuration on a subject NP: (a) Nos **amis** aiment bien Nancy. (b) **Our friends** really like Nancy.

In sentences with a continuative configuration on a subject NP (Figure 8), the same phenomena was observed as in the CIA sentences (Figure 7): both speaker groups favored a high f_0 level (corresponding to a rising F0 curve). This level was again higher in French than in English and no vowel lengthening was used to strengthen the prosodic boundary. The final f_0 level was low in both languages (although lower in English than in French) and the final vowel was significantly lengthened in English, while moderately lengthened in French.
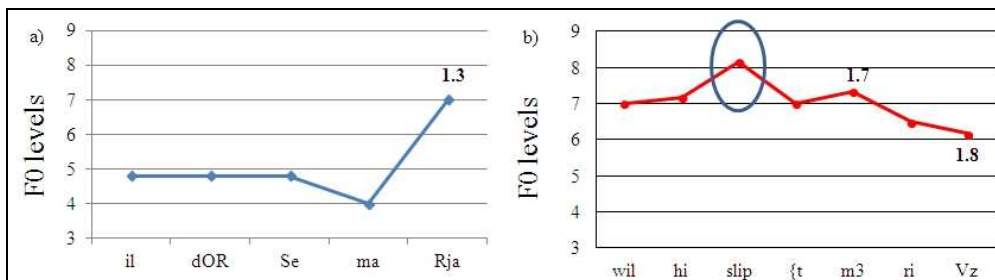
Figure 9. QAS - Question configuration at the end of a clause: (a) *Il dort chez **Maria**?* (b) *Will he sleep at **Maria's**?*

In French, the yes/no question configuration (Figure 9) of f_0 levels is similar to the configuration found in QIS type sentences (figure 10): a huge level rise preceded by a rather flat f_0 level. The pattern in English sentences contained a lowering of the f_0 level at the end of the sentence as the interrogative character was expressed here by syntactic means (subject-verb inversion); therefore there was no need for prosodic marking.
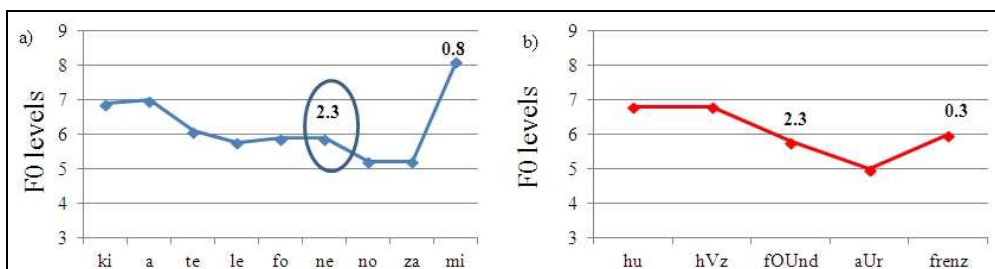

Figure 10. QIS - Interrogative configuration on a simple subject NP: (a) *Qui a téléphoné? **Nos amis**?* (b) *Who has phoned? **Our friends**?*
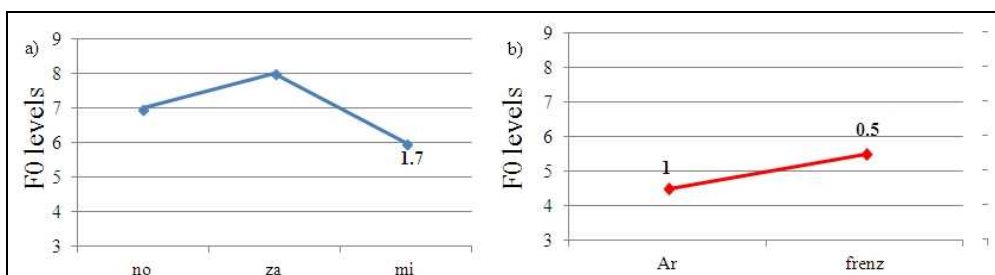

Figure 11. DIS - Short declarative sentence (a) ***Nos amis***. (b) ***Our friends***.

The French and English versions of the previous sentences contained final F0 rise (high f_0 level), however the level was much higher in French sentences than in English. The first part of the sentence contained a clause containing an interrogative pronoun and its occurrence explained the falling pattern of the f_0 levels. The vowel

47

duration was used in both sentences to mark the prosody boundary in the first part of the sentence.

The short declarative sentence had a falling f_0 (low f_0 levels) in French pronunciations. However, in the English realization of the sentence, the pattern was slightly rising. In both sentences (French and English), the final vowel duration was also lengthened and used as a boundary marker.
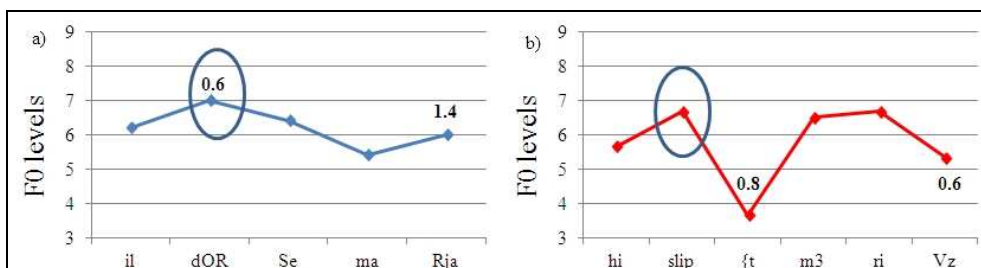

Figure 12. DAS - Longer declarative sentence: (a) Il dort chez **Maria**. (b) He'll sleep at **Maria's**.

In the longer declarative sentence, the f_0 level of the last vowel was low in English (falling movement) and slightly rising in French. In both cases, the vowel duration was lengthened and marked the prosodic boundary, while the first prosodic boundary was marked by slightly higher f_0 level.

3.6 General discussion

In French, the f_0 level was high at a major prosodic boundary. In fact, the level was higher than in English, especially in yes/no questions. English speakersused falling f_0 patterns to mark major continuation prosodic boundaries and strongly falling patterns to mark the end of declarative sentences. The duration of the last vowel was often lengthened in English and was used to mark the prosodic boundary.

In French declarative sentences, the f_0 range was narrower (1.8 levels on average) than in English (3.5 levels on average). In interrogative sentences, the mean f_0 pattern values was 3 in French and 2 in English. In French, the f_0 was more strongly rising on prosodic boundaries than in English. The final f_0 movement in assertive sentences was more moderate in French (falls through 1.2 levels) than in English (falls through 2.1 levels).

The declarative sentences in French were uttered at a higher f_0 level (mean level value 7) than English sentences (mean level value 5.4). The level range used in English sentences was larger (the f_0 on average evolves through 3 levels) than in French sentences, where the mean level range used is 2.

Interrogative sentences in French were uttered at relatively lower range (5 and 6.2) compared to assertive sentences. English speakers used a relatively higher range level for interrogative sentences than assertive sentences (6.9 and 6.2 levels).

The general tendency for French intonation in the phrases studied here is as follows: in French, speakers gave preference to a more flat f_0 (narrower range of f_0

levels used), with mainly upward movement on prosodic boundaries. In English, the range of f_0 levels was broader with mainly downward f_0 movement.

Vowel duration wass used in both languages to indicate prosodic boundaries. In French, a slight f_0 movement on a prosodic boundary was completed by lengthened vowel duration, which indicated the boundary location and its depth. In English, vowel lengthening typically took place at boundaries where the f_0 movement was important. The lengthened vowel duration was used in both languages, however vowel durations were longer on non-final prosodic boundaries in French (mean coefficient value of vowel lengthening 1.8) than in English (mean coefficient value of vowel lengthening 0.8). Moreover, vowel duration was slightly more lengthened in English in sentence final syllables (followed by a pause) than in French. Indeed, in English, the mean vowel lengthening coefficient value was 1.4, while in French its value was 1.2.

3.7 Speech synthesis

In order to verify whether our approach to prosody representation and coding is correct, the f_0 pattern represented as a range of 9 levels was transformed to semitones values and these values were used to synthesize the melody of the sentences in our corpus. According to our preliminary perception tests, made by only 2 expert phoneticians (a French and an English native), all of the resynthesized sentences sounded very natural and there was very little difference between the modified and unmodified sentences. The listening tests were carried out by MOS (*Mean Opinion Score*) tests and the re-synthezided and natural sentences were judged on a 5 point scale (0-very bad, 5-excellent). According to this very preliminary test, the appreciation of naturalness in non-modified sentences was 4,4 out of 5 and the f_0 resynthesized sentences obtained a score of 4,2. Naturally, this very preliminary test will be completed in the future using more listeners in order to verify the validity of our preliminary tests.
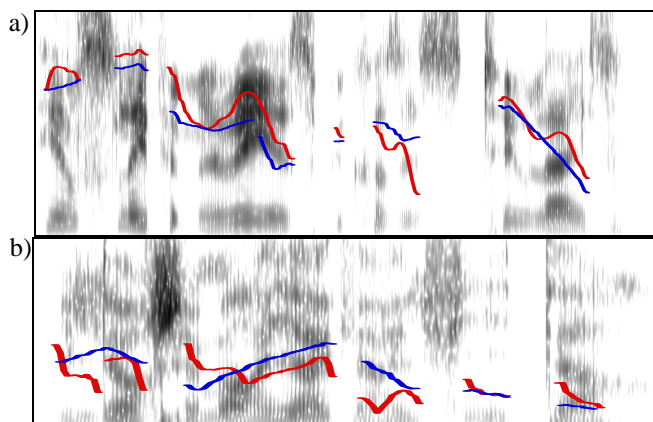

Figure 13. Examples of resynthesis of the melody (a) of an English and (b) of a French sentence. Natural melody curve in red and the synthesized melody curve in blue.

4 Conclusion

The goal of our study is to use an appropriate coding schema for prosody representation in a cross linguistic study of French and English prosody. The data used are laboratory data produced by a group of French and English native speakers and they contain sentences sharing the same syntactic structures in both languages. This syntactic specificity of the data base is well adapted to cross-linguistic study as it allows for comparison of prosodic phenomena relatively easily. However, a methodological problem remains: how to represent prosodic parameters in such a way that comparison would be pertinent.

Two approaches are tested in this study; the first is a general statistical analysis, which compares f_0 slopes measured on the last syllable of some of the words considered as pertinent from a prosodic point of view. This analysis showed that the prosody used in different syntactic structures is not necessarily supportive ofprevious prosodic theory (Delattre, 1966).

The second part of the study was dedicated to a more qualitative comparison of French and English prosody. Two prosodic parameters, vowel duration and F0 values were coded by an automatic prosodic transcriber (PROSOTRAN), which provided symbolic and numeric annotations for use in our cross-linguistic study. The cross-linguistic comparison of these two parameters highlighted the same basic general differences or similarities on the use of prosody in these two languages. An attempt was also made here to verify how faithful the prosodic coding was by transforming the symbolic values of F0 levels back to physical parameter values and then reconstructing the prosody of the sentences with F0 synthesis. The preliminary results are very encouraging but further study is needed in order to get reliable perception test results.

References

Aubergé, V., T. Grepillat, A. Rilliard 1997, Can we perceive attitudes before the end of sentences? The gating paradigm for prosodic contours. In *EuroSpeech'97*. Rhodes, Grèce, pp. 871-877.

Bartkova, K., A. Bonneau, V. Colotte, M. Dargnat 2012. Productions of "continuation contours" by French speakers in L1 (French) and L2 (English). In *Proceedings of Speech Prosody*. Shangaï, China, 22-25 May 2012, pp. 426-429.

Bartkova, K., E. Delais-Roussarie, F. Santiago-Vargas 2012. PROSOTRAN: a tool to annotate prosodically non-standard data, In *Proceedings of Speech Prosody*. Shangaï, China, 22-25 mai 2012, pp. 55-58.

Bartels, C. 1999. *The Intonation of English Statements and Questions.* New-York: Garland Publishing.

Beaugendre, F., Ch. d'Alessandro, A. Lacheret-Dujour and J. Terken 1992. A perceptual study of French intonation. In *ICSLP 92 Proceedings: 1992 International Conference on Spoken Language Processing*. Edmonton, Canada: Priority Printing, pp. 739-742

Beckman, ME., J. Hirschberg and S. Shattuck-Hufnagel 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*. Oxford: University Press, Chapter 2, pp. 9–54.

Beyssade, C., J.-M. Marandin, and A. Rialland 2003. Ground/Focus: a perspective from French. In R. Nunez-Cedeno et al. (eds): *A Romance perspective on language knowledge and use: selected papers of LSRL 2001*. Amsterdam/Philadelphia: Benjamins, pp. 83-98.

Boidin, C. 2009. *Modélisation statistique de l'intonation de la parole expressive*. PhD thesis, Rennes, published by University of Rennes 1.

Bolinger, D. 1978. Intonation across languages, Intonation across languages. In: *Universals of human language 2*. Stanford, Stanford UP, pp. 471-524.

Campbell, WN. 1992. Syllable-based segmental duration. In Bailly and Benoît (eds): *Talking machines: theories, models and design*. Amsterdam: Elsevier, pp. 211-224.

Crystal, D. 1969. *Prosodic systems and intonation in English*. Cambridge, Cambridge UP.

Delattre, P. (1938), "A comparative study of declarative intonation in American English and Spanish", Hispania XLV/2, pp. 233-241.

Delattre, P. 1938. L'accent final en français: accent d'intensité, accent de hauteur, accent de durée. *The French Review* 12(2), pp. 141-145.

Delattre, P. 1966. Les dix intonations de base du français. *The French Review* 40(1), pp. 1-14.

Delais-Roussarie, E. 2005. Interface Phonologie/Syntaxe: des domaines phonologiques à l'organisation de la Grammaire. In Durand, J., N. Nguyen, V. Rey and S. Wauquier-Gravelines (eds): *Phonologie et phonétique: approches actuelles*. Paris: Editions Hermès, pp. 159-183.

Delais-Roussarie, E. 2000. Vers une nouvelle approche de la structure prosodique. In *Langue Française* 126: Paris: Larousse, pp. 92-112.

Di Cristo, A. 1985. *De la microprosodie à l'intonosyntase*. Thesis. Université de Provence.

Di Cristo, A. 1998. Intonation in French. In A. Di Cristo and D. Hirst (eds): *Intonation systems: a survey of twenty languages*. Cambridge, Cambridge UP.

Di Cristo, A. 2010. *A propos des intonations de base du français*. Unpublished MS.

Fónagy, I., E. Bérard 1973. Questions totales simples et implicatives en français parisien, Interrogation et Intonation. In A. Grundstrom and Léon P. (eds): *Studia Phonetica* 8, Ed. Paris: Didier. pp. 53-98.

Fónagy, I. 1980. L'accent français, accent probabilitaire: dynamique d'un changement prosodique. In I. Fónagy and Léon (eds): *L'accent en français contemporain. Studia Phonetica* 15. pp. 123-233.

Gussenhoven, C. 1984. *On the grammar and semantcis of sentence accents*. Dordrecht: Foris.

Hadding-Koch H., M. Studdert-Kennedy 1964. An experimental study of some intonation contours. *Phonetica* 11, pp. 175-185.

Hirst, D. 1998. Intonation of British English. In D. Hirst and A. Di Cristo (eds): *Intonation Systems: A survey of twenty languages*. Cambridge: Cambridge University Press. pp. 56-77.

Hirst, D. and A. Di Cristo 1998. *Intonation systems: A survey of twenty languages*. Cambridge, Cambridge University Press

Ladd, R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Lacheret-Dujour, A., N. Obin, M. Avanzi 2010. Design and evaluation of shared prosodic annotation for French spontaneous speech: from expert's lnowledge to non-experts annotations. In *Proceedings of the 4th Linguistic Annotation Workshop*. Uppsala, Sweden, pp. 265-274.

Mertens, P. 2004. The Prosogram: semi-automatic transcription of prosody based on a tonal perception model. In *Proceedings of Speech Prosody 2004*. Nara, Japan, pp. 549-552.

Mesbahi, L., D. Jouvet, A. Bonneau, D. Fohr, I. Illina, Y. Laprie 2011. Reliability of non-native speech automatic segmentation for prosodic feedback. In *Proceedings Workshop on Speech and Language Technology in Education*. Venice, Italy, pp. 41-44.

Pierrehumbert, J. 1980. *The phonology and phonetics of English intonation*. PhD thesis, MIT. Distributed 1988, Indiana University Linguistics Club.

Post, B. 2000. *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.

Rossi, M. 1999. *L'intonation, le système français: description et modélisation*. Paris:

Editions Ophrys.

Segui, J. 1984. The syllable: A basic perceptual unit in speech processing? In H. Boumam and DG Bouwhuis (eds). *Attention and performance Vol.10: Control of language processes*. Hillsdale: Erlbaum. pp. 165-181.

Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert and J. Hirschberg 1992. ToBI: a standard for labeling English prosody. *Proceedings of the Second Int. Conf. on Spoken Languages.* No 2, pp. 867-70.

Speech Processing, Transmission and Quality Aspects (STQ) 2005. Distributed speech recognition; extended advanced front-end feature extraction algorithm; compression Algorithms. *European Telecommunications Standards Institute, European Standards (ETSI ES).* pp. 202-212.

Vaissière, J. 2002. Cross-linguistic prosodic transcription: French vs. English. In NB. Volskaya, ND. Svetozarova, PA. Skrelin (eds.): *Problems and methods of experimental phonetics. In honour of the 70th anniversary of Pr. LV. Bondarko.* St Petersburg: St Petersburg State University Press. pp. 147-164.