

Modelling interlanguage intonation: the case of questions

Sophie Herment¹, Nicolas Ballier², Elisabeth Delais-Roussarie³, Anne Tortel⁴

¹ Aix-Marseille Université, UMR 7309 - Laboratoire Parole et Langage, France

² Université Paris-Diderot, CLILLAC – ARP (EA 3967), France

³ UMR 7110-Laboratoire de Linguistique Formelle, Université Paris-Diderot, France

⁴ Université Nice-Sophia Antipolis, UMR 7320 - Bases, Corpus, Langage, France

sophie.herment@univ-amu.fr, nicolas.ballier@univ-paris-diderot.fr,
elisabeth.roussarie@wanadoo.fr, atortel@unice.fr

Abstract

In this paper, we study the intonational patterns observed in learners' productions in order to evaluate what motivates the deviations observed: systemic differences between the learners' L1 and the L2, differences in phonetic implementation, etc. The analysis consists of a cross-comparison of the intonation of yes-no questions in French, English and English as an L2. It is based on five information-seeking yes-no questions that were extracted from the AixOx corpus, which contains a set of 40 texts that were read by 10 native French speakers, 10 Native English speakers and 20 French learners of English. The analysis of the data showed that the differences between native and non-native speakers do not affect the form of the nuclear contour. It mostly shows that French speakers of English have a tendency to assign a rising pitch movement at the end of any prosodic words, which leads to a clear difference in rhythm.

Index Terms: acquisition of prosody in L2, intonation, prosodic phrasing, rhythm, prosodic modelling, learner corpora, interphonologies.

1. Introduction

Research on interlanguage intonation has shown that the intonational patterns observed in learners' productions are often influenced by their L1s (see, among others, [1], [2] and [3]). As a consequence, the notion of L1 transfer is often invoked to account for the observed patterns. As pointed by [4], however, transfer may apply at the phonological as well as at the phonetic level. Transfers at the phonological level result from differences in the metrical structure or the tonal inventory. In a study on the intonation of tag questions in English, [5] have shown for instance that Spanish speakers of English use rises at the end of the question tag for confirmation request, whereas native English speakers will use falls, these patterns being thus analyzed as resulting from a phonological transfer. By contrast, transfers at the phonetic level occur when an identical phonological form differs in the way it is phonetically implemented in both languages. Differences in the temporal alignment of pitch accents may be for instance a case of phonetic transfer (see [1] for concrete examples). The distinction between different types of transfers or deviations is of great help to study interlanguage intonation, as pointed by [6].

In this paper, we will show however that classifying the observed deviations is not an easy task since the deviation type may change over time, and a mere cross-comparison of the surface forms may not always be sufficient. This will be done through the analysis of information-seeking *yes-no*

questions realized by French learners of English and extracted from the AixOx corpus ([7] and [8]).

The paper is organized as follows: in section 2, the prosodic characteristics of information-seeking *yes-no* questions in French and English are described; section 3 presents the data and methodology used for the prosodic analysis; the results obtained in the cross-comparison are given in section 4; section 5 discusses the results and offers perspectives and concluding remarks.

2. The intonation of *yes-no* questions

2.1. *Yes-no* questions in French

From a morpho-syntactic point of view, three distinct constructions may be used in French to build up a *yes-no question*: (i) declarative structures similar to the ones observed in assertive sentences (1); (ii) subject-object inversion, be the subject nominal or pronominal (2); and (iii) an interrogative particle *est-ce que* can be inserted in sentence initial position, the rest of the sentence having the same syntactic structure as in assertions (3).

- (1) *Vous avez appris des langues étrangères ?* ('Did you learn any foreign languages?')
- (2) *Pierre est-il venu ?* ('Did Pierre come?')
- (3) *Est-ce que c'est vrai ?* ('Is that true?')

As far as intonation is concerned, rising tones are seen as the canonical form associated with declarative questions (see, among others, [9] and [10]). By contrast, in *yes-no* questions in which the modality of the utterance is indicated by a morpho-syntactic or a lexical marker (subject-verb inversion or *est-ce que* particle respectively), non-rising tones may be used on a par with rising ones (see, amongst others, [9], [11] and [12]). Note, however, that the rising tune is by far the most frequently used in information-seeking *yes-no* questions, regardless of the construction (see, amongst others, [13]).

2.2. *Yes-no* questions in English

Contrary to French, in English information-seeking *yes-no* questions, the modality of the utterance is always indicated by morpho-syntactic means: either by subject/auxiliary inversion (4), or by the use of auxiliary *do* (5).

- (4) *Is Peter coming?* (vs. *Peter is coming.*)
- (5) *Does he live in Paris?* (vs. *He lives in Paris.*)

Declarative questions can be found in English, but they are usually echo-questions, which are not present in our corpus.

As opposed to *wh*-questions, for which the default tone is a fall in English, *yes-no* questions are uttered with a rising tone, even if a falling tone can also be heard, but less frequently (see amongst others [14], [15] and [16]).

3. Corpus and method

3.1. Corpus

3.1.1. The AixOx corpus

The AixOx corpus ([7], [8]) is a multilingual learner corpus which consists of recordings of 40 1-minute passages in English and French read by native speakers and L2 learners (all aged 20-35, see [8] for details about the informants). The passages were extracted from the Eurom 1 corpus ([17]). Non-native speakers were divided into two groups, B and C, according to their level of proficiency in the Common European Framework of Reference for Languages (CEFR). Learners of group B are independent users, B1/B2 in the CEFR and learners of group C are proficient users, C1/C2. Hence the corpus is composed of 6 groups of speakers, as shown in table 1 below. For each group, 10 speakers were recorded (5 females and 5 males), the corpus thus amounting to 60 speakers and about 30 hours of speech.

Table 1. *Speaker groups in AixOx.*

Language of recording	Native speakers	L2 learners B1/B2	L2 learners C1/C2
English	ENEN	FRENB	FRENC
French	FRFR	ENFRB	ENFRC

3.1.2. The extracted questions

Before extracting the questions on which the study is based, we had to take a few facts into account:

- The English and French corpora are not word-to-word translations, even if they are very close, since the pragmatic contexts in which the texts are uttered are similar; hence the English corpus contains 23 questions when the French one contains 22.
- *Yes-no* and *wh*-questions are present in the corpus.
- For the total questions, the French corpus includes different types of constructions, whereas *yes-no* questions in the English corpus all display subject/auxiliary inversion. In order to compare the data, only questions with an interrogative marker have been taken into account in the present study.
- Some *yes-no* questions in both languages are not neutral questions since they may be rhetorical questions addressed to oneself ('*What will 1992 really mean to the person in the street?*'), disguised orders ('*Can you give me a firm date now?*') or have the meaning of partial questions ('*Can you tell me what's on television tonight?*').

In order to make a cross-comparison between *yes-no* questions, we took into account the same types of questions, both on the syntactic and pragmatic levels. We therefore concentrated on 5 information seeking *yes-no* questions, in which the modality of the utterance is indicated by morpho-syntactic means. They are listed below (the English and French questions are not necessarily translations, since some of the French questions have been disregarded because of their declarative structure):

- French questions

F1 : Est-ce que vous pourriez me donner leur nouveau numéro de téléphone ?

F2 : Est-ce que c'est vrai ?

F3 : Est-ce que vous pourriez me donner la liste des restaurants de mon quartier ?

F4 : Est-ce que vous avez des tarifs spéciaux pour les collectivités ?

F5 : Est-ce qu'un organisme universitaire peut en bénéficier ?

- English questions

E1 Can you give me their new number please?

E2: Could you please tell me the best connections to Sheffield from East Greenstead?

E3: Do you take reservations by telephone?

E4: Can you give me a list of the restaurants in the neighbourhood?

E5: Do you have special corporate academic institutions?

We therefore studied 5 questions uttered by 30 speakers (10 in the three groups – natives and learners from groups B and C - for the 2 languages) in the two languages, the corpus thus amounting to 300 sentences.

3.2. Method for the prosodic analysis

In order to compare the intonation of the extracted questions uttered by the native speakers and by the B and C learners, we used two distinct approaches to encode the tones observed in the various utterances: a perceptual approach, and a semi-automatic approach.

3.2.1. The perceptual approach

Following [18] and the British tradition, the perceptual analysis relies for English on tonality (the division into intonation phrases), tonicity (the place of nuclear syllables) and tones (the distinctive pitch movements).

We consider, following [19], that there is only one level of boundary, associated with the intonation phrase (IP). In short questions, one IP will be realized but longer questions like *Could you please tell me the best connections to Sheffield from East Greenstead?* can be divided into 2 IPs by the natives, and even more by the learners. The IP boundaries are marked by a slash in the paper.

The place of the tonic syllable, the nucleus, is looked at. The principle that there is only one nuclear syllable in an IP is adopted. It is the most prominent one, that bearing the tone (the distinctive pitch movement) of the IP (see for example [20] or [21]). The nuclear syllable is underlined in the given examples.

Finally, the tones are encoded. The tone is the distinctive pitch-movement, that bearing on or starting on the nucleus and extending on the post nuclear syllables (if any). The symbols used are F for a simple fall, R for a simple rise, HF for high fall and FR for fall-rise, again according to the British tradition but with a limited tone inventory, following [22] and [23].

The questions are therefore encoded as follows in (6), (7) and (8), respectively a native speaker, a learner from group B and a learner from group C:

(6) *Can you give me their new number, please F /*

(7) *Can you give me F / their new number, please FR /*

(8) *Can you give me their new number, please R /*

So as to allow cross-comparison, the French questions were encoded in the same way, as shown in examples (9), (10) and (11), respectively a native speaker and learners from groups B and C. The global contour of the IP was taken into consideration in our study, and not the contours of the accentual phrases.

(9) *Est-ce que vous pourriez me donner leur nouveau numéro de téléphone R /*

- (10) *Est-ce que vous pourriez me donner F / leur nouveau numéro de téléphone R /*
- (11) *Est-ce que vous pourriez me donner leur nouveau numéro de téléphone R /*

3.2.2. The semi-automatic approach

A semi-automatic approach was also performed. The questions studied were extracted under PRAAT [24], automatically aligned into words with SPPAS [25] and manually labelled into syllables. The SAMPA phonological representation of our syllabification is: k@n / ju / gIv / mi / @ / llst / @v / D@ / rEst / rQnts / In / D@ / neIb / @ / hUd. The syllable tier of the annotation was used for automatic extractions of acoustic parameters using ProsodyPro [26].

3.3. Revisiting the intonation of English yes-no questions

The first result we obtain concerns the intonation produced by native speakers. Table 2 below shows the results of the perceptual analysis: only the global contours (on the final IP), i.e. Falling (F and HF) or Rising (R and FR) of the questions studied are reflected in the table:

Table 2. Contours for ENEN and FRFR questions

Questions	English natives	French natives
E1/F 1	F 80%	R 100%
E2/F 2	F 100%	R 90%
E3/F 3	R 70%	F60%
E4/F 4	F 60%	R 90%
E5/F 5	F 80%	R 80%
Total	F 70%	R 80%

The table clearly shows, contrary to what the literature claims, that the rising tone is not the most frequent for information-seeking yes-no questions in English: 70% of the 50 English questions under scrutiny are uttered with a falling tone. For French, our data confirm that the rising tone is the default tone for yes-no questions.

3.4. The intonation of learners: global contour and phrasing

Table 3 gives the results for the French learners of English:

Table 3. Contours for FRENB and FRENC questions

Questions	FRENB	FRENC
E1	R 90%	R 60%
E2	R 80%	R 70%
E3	R 90%	R 60%
E4	R 80%	F 60%
E5	R 90%	R 80%
Total	R 86%	R 66%

The learners of both groups massively pronounce the questions on a rising tone, as in their mother tongue, but the tendency is less strong for the proficient group: 44% of the questions are uttered with a fall, which is quite a lot compared to the tendency for the native English speakers.

But the global contour is not the most relevant feature for proficiency. The phrasing is probably the most salient difference between the independent (group B) and the proficient (group C) learners. The independent learners tend to paste the French phrasing to the English sentence: they

produce a pitch movement at the end of word groups which would correspond to accentual phrases in French, as exemplified in (12) and (13):

- (12) *Can you give me a list R / of the restaurants R / in the neighbourhood R /*
- (13) *Could you please R / tell me R / the best connections R / to Sheffield R / from East Greenstead R /*

A clear evolution can be noted for the proficient learners, who no longer group words as in French, but, as a few native speakers do so too, divide E4 in 2 IPs for 8 speakers out of 10 (3 IPs for 1 speaker and 1 IP for 1 speaker) and E2 in 2 IPs for 100% of the speakers, as in the occurrences below:

- (14) *Can you give me a list of the restaurants F / in the neighbourhood F /*
- (15) *Could you please tell me the best connections to Sheffield F / from East Greenstead R /*

3.5. Multi-speaker modelling

With the semi-automatic approach, it is possible to visualize and compare the curves for the natives and the learners using R [27], as is shown in figure 1 below for question E4:

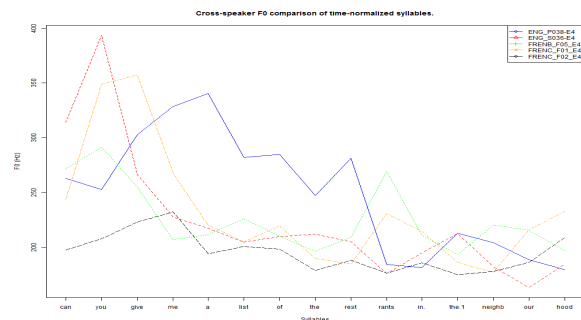


Figure 1. Comparison between two natives, two C learners and one B learner

For clarity's sake, the number of speakers represented has been limited and only female speakers have been taken into account in Figure 1. In addition, one should be made particularly aware that the curvature does not represent F0 variation *per se*, but the statistical software interpolations between the points representing the means estimated for each syllable. ProsodyPro offers a finer-grained representation, which is based on 10 successive measures of pitch over the same syllable. This corresponds to a kind of time normalization: the duration of each syllable might be different from one speaker to another, but each tenth of a syllable duration can be compared across speakers.

The nuclear syllables are easily identifiable on the curves. Figure 1 shows that for one native a pitch movement occurs on *RESTaurants* (the blue curve). For one FRENC learner, a peak is also clearly visible on *RESTaurants* (blackline), but the prenuclear part of the curves is quite different from those of the natives. As for the other C learner (yellow curve) and the B learner (green curve), a pitch movement also occurs on the word *restaurant*, but we see that the peaks have moved right, on the final syllable of *restauRANTS*. It is to be noted that the natives and the learners of group C pronounce *restaurants* with 2 syllables while many learners of group B pronounce 3 syllables as in French. The non compression of the median

syllable and the stress shift on the right are typical of a French learner's pronunciation. A closer look at the curves also confirms the perceptual impression that pitch movements take place on words preceding *restaurants*. The pitch contour in Figure 2 is representative of that tendency: peaks on *me* and *list* appear.

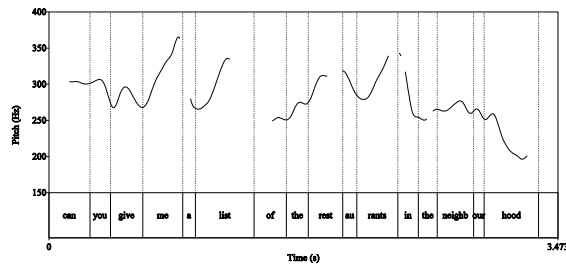


Figure 2: Pitch track of E4 uttered by a B learner

4. Discussion and conclusion

With the automatic procedure, cross-sectional comparisons are established on the basis of normalized duration, so that performances are equally considered on the basis of the prosodic target (here, pitch) for each syllable. In that sense, the semi-automatic procedure consists in discourse alignment, not in time-alignment, allowing prosodic realizations to be compared. This simplifies speech time variability and enables the design of “confidence intervals” for native prosodic realizations and potential non-native mismatches. A tentative representation of this modelling lies in the graphing and statistical representation, using R once more. Due to space limitation, we do not include the corresponding boxplot, but annotation at the word level is not so telling and syllable-based analysis is much more convincing. This semi-automatic approach paves the way for the characterization of learner profiles and their interaction with criterial features [28]. We give a rough outline of some of the interlanguage stages that can be detected with this kind of approach. Figures 3, 4 and 5 represent the boxplots synthesizing inter-speaker variation of mean fundamental frequency (computed for each syllable) for E4 realized by respectively the female natives, some of the female learners from group B and from group C. In the figures, the small rectangles show that the dispersion is limited and that there is a consensus on the prominence of the syllable. The large rectangles on the contrary point to a larger variation, *i.e.* some sort of non-consensus. The median (central line in bold) allows the visualization of the melody. If we first compare the median on the three figures, we clearly see that the natives (figure 3) divide the sentence into 2 IPs and favour a falling pattern on the first IP (on *RESTaurants*) and a rising one on the second (*NEIGHBourhood*); learners from group C (figure 4) also divide the sentence into 2 IPs, with similar nuclei as the natives, but realize a rise on both IPs; the less advanced B learners (figure 5) divide the sentence into what corresponds to the French phrasing with rising movements on *me*, *list*, *restauRANTS* and *neighbourHOOD*, followed by an F0 resetting after the first three words. If we now look at the rectangles, a large dispersion noticeably appears in figures 3 and 4 on tool words like *you*, *me*, *a* and *of*, which are far less important prosodically, as opposed to the stressed syllables of *restaurants* or *neighbourhood* for example, which display small rectangles (in figure 4 even more strikingly): the natives and the C learners therefore show strong agreement as to

which syllable they should make prominent. The prenuclear contour, however, is still somewhat hesitant for the C learners (the dispersion is very high on *you* and *give*). Finally, the rectangles for B learners (figure 5) show much less difference in the dispersion, this reflecting isosyllabic realizations, typical of French rhythm.

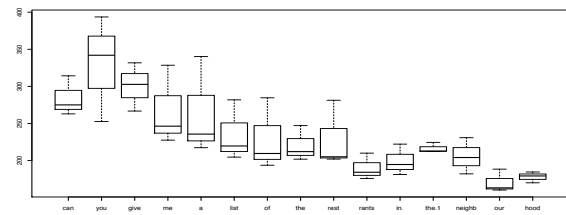


Figure 3. Native female speakers

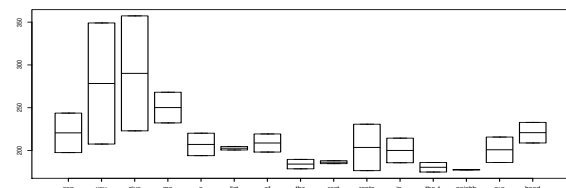


Figure 4. Female learners from group C

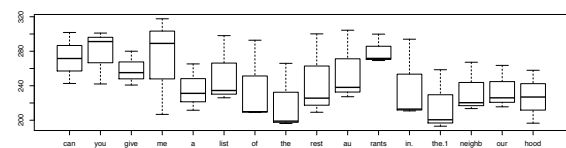


Figure 5. Female learners from group B

As explained in [29], there are limitations to this methodology (resyllabification issues, neutralization of rhythmic variation, pitch detection errors), but one consequence is for sure: modelling interlanguage intonation like this means that syllable division is high on the agenda for this kind of spoken learner corpus research.

The comparison of the productions of information-seeking *yes-no* questions by natives and learners showed that the form of the nuclear contour is not so much affected. The most important differences concern rhythm, and the prenuclear syllables. French speakers of English, in particular at B level, have a tendency to assign a rising pitch movement at the end of prosodic words, which leads to a clear difference in rhythm. The study of the intonation of the questions by English learners of French (which could not be developed in this paper for lack of space) also shows that the nuclear contour is somehow well realized too and it is to be noted that it seems easier for English learners of French to utter questions than for French learners of English, probably because the prenuclear syllables are better realized. Most studies on L2 intonation focus on the phonological nuclear form. The present paper encourages further study on the phonetic implementation of the prenuclear contour.

5. References

- [1] Mennen, I., "Bi-directional interference in the intonation of Dutch speakers of Greek", *Journal of Phonetics* 32, 543–563, 2004.
- [2] Jilka, M., *The Contribution of Intonation to the Perception of Foreign Accent*, Doctoral dissertation, University of Stuttgart, 2000.
- [3] Rasier, L., & Hilgsmann, P., "Prosodic transfer from L1 to L2. Theoretical and methodological issues", *Nouveaux cahiers de linguistique française* 28, 41-66, 2007.
- [4] Mennen, I., "Phonological and phonetic influences in non-native intonation", in J. Trouvain & U. Gut [eds.], *Non-native Prosody: Phonetic Descriptions and Teaching Practice*, pp. 53–76, Mouton De Gruyter, 2007.
- [5] Ramírez, D. & J. Romero, "The pragmatic function of intonation in L2 discourse: English tag questions used by Spanish speakers", *Intercultural Pragmatics* 2 (2), 151-168, 2005.
- [6] Mennen, I., "Beyond segments: towards an L2 intonation learning theory (LILT)", in Delais-Roussarie, E., Avanzi, M. & S. Herment [eds.], *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations*, Springer Verlag, accepted, to appear.
- [7] Herment, S., Loukina, A. & A. Tortel, "The AixOx corpus", *SLDR*, 2012. <http://sldr.org/sldr000784/fr>
- [8] Herment, S., Tortel, A., Bigi, B., Hirst, D. & A. Loukina, "AixOx, a multi-layered learners' corpus: automatic annotation", in Diaz Pérez, J. and A. Diaz Negrillo, [eds.], *Specialisation and variation in language corpora*, Bern: Peter Lang, to appear.
- [9] Delattre, P., "Les Dix Intonations de base du français", *The French Review* 40 (1), 1-14, 1966.
- [10] Di Cristo, D., "Intonation in French", in Hirst, D. & A. Di Cristo [eds.], *Intonation systems: A survey of twenty languages*, 195-218, Cambridge: Cambridge University Press, 1998.
- [11] Martin, P., "Une grammaire de l'intonation de la phrase française 2", *Rapport d'Activité de l'institut de phonétique* 9/2, pp. 77-96, Institut de Phonétique de l'Université Libre de Bruxelles, 1975.
- [12] Martin, P., *Intonation du français*, Paris: Armand Colin, 2009.
- [13] Santiago-Vargas F. & Delais-Roussarie, E., "Acquiring phrasing and intonation in French as a second Language: the case of Yes-No questions produced by Mexican Spanish Learners", *Proceedings of Speech Prosody*, Shanghai, China, 2012.
- [14] Cruttenden, A., *Intonation*, Cambridge: Cambridge University Press, 2nd ed. 1997 (1st ed. 1986).
- [15] Wells, J.C., *English Intonation, an Introduction*, Cambridge: Cambridge University Press, 2006.
- [16] Roach, P., *English Phonetics and Phonology, A practical course*, Cambridge: Cambridge University Press, 4th ed. 2009 (1st ed. 1983).
- [17] Chan, D., Fourcin, A., Gibbon, D., Grandström, B., Huckvale, M., Kokkinakis, G., Kvale, K., Lamel, L., Linderg, B., Moreno, A., Mouropoulos, J., Senia, F., Transcoso, I., Velt, C. & J. Zeiliger, "EUROM – A Spoken Language Resource for the EU", *Proceedings of Eurospeech '95*, (Madrid) 1, 867-880, 1995.
- [18] Halliday, M.A.K., *Intonation and Grammar in British English*, The Hague-Paris: Mouton, 1967.
- [19] Grabe, E., Post, B. & F. Nolan, "Modelling intonational variation in English. The IViE system", in Puppel, S. & G. Demenko [eds.], *Proceedings of Prosody 2000*, Adam Mickiewicz University, Poznan, Poland, 2001.
- [20] Tench, P., *The Intonation Systems of English*, London: Cassell, 1996.
- [21] Halliday, M.A.K. & W.S. Greaves, *Intonation in the Grammar of English*, London-Oakville: Equinox, 2008.
- [22] Hirst, D., "Form and function in the representation of speech prosody", in Hirose, K., Hirst, D. & Y. Sagisaka [eds.] *Quantitative prosody modeling for natural speech description and generation (Speech Communication* 46 (3-4)), 334-347, 2005.
- [23] Gussenhoven, C., *On the grammar and semantics of sentence accent*, Dordrecht: Foris, 1984.
- [24] Boersma, P. & D. Weenink, "Praat, a system for doing phonetics by computer", *Glott International*, 5:9/10:341-345, 2001.
- [25] Bigi, B., "SPPAS: a tool for the phonetic segmentation of speech", *Proceedings of the Language Resource and Evaluation Conference*, Istanbul, Turkey, 1748-1755, 2012.
- [26] Xu, Y., "ProsodyPro, A Tool for Large-scale Systematic Prosody Analysis", *Proceedings of the TRASP conference*, Aix-en-Provence, France, 2013.
- [27] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org>
- [28] Cauvin, E. "Intonational phrasing as a potential indicator for establishing prosodic learner profiles", In Granger, S., Gilquin, G. & F. Meunier [eds.], *Twenty Years of Learner Corpus Research: Looking back, Moving ahead. Proceedings of Corpora and Language in Use*, Louvain-la-Neuve: Presses universitaires de Louvain, 75-88, 2013.
- [29] Ballier, N. & P. Martin, "Developing corpus interoperability for phonetic investigation of learner corpora" in Diaz-Negrillo, A., Ballier, N. & P. Thompson [eds.], *Automatic Treatment and Analysis of Learner Corpus Data*, Amsterdam: Benjamins, 33–64, 2013.