

The production and perception of L1 and L2 Dutch stress

Marie-Catherine Michaux¹, Sandrine Brognaux^{2,3}, George Christodoulides¹

¹Centre Valibel, Institute for Language and Communication, University of Louvain, Belgium

²Cental, University of Louvain, Belgium ³TCTS, University of Mons, Belgium

marie-catherine.michaux@uclouvain.be, sandrine.brognaux@uclouvain.be, george@mycontent.gr

Abstract

This study aims at exploring the production and perception of Dutch word stress by Francophone learners of (Belgian) Dutch. For this purpose a production experiment was first carried out. In line with other studies, it was hypothesized that participants would show a tendency to stress the final syllable. Even though this hypothesis was confirmed, there was also a substantial lack of agreement between the five labellers who perceptually annotated the data for stress position. To further investigate this matter, acoustic measures were extracted. The data suggest that both groups of speakers do not use acoustic correlates to signal prominence in the same way, the Dutch group using intensity, vocalic nucleus duration and pitch movement more, while the French group prefers duration and pitch movement. This study also led us to develop tools to phonetise, syllabify and facilitate the acoustic analysis of Dutch speech.

Index Terms: L2 prosody, Dutch as a Foreign Language, speech perception, speech production

1. Introduction

Due to discrepancies between the Dutch and French prosodic systems and to the lack of attention paid to pronunciation in Dutch didactics, Dutch word stress can be problematic for Belgian Francophone learners of Dutch as a Foreign Language (DFL). Dutch on the one hand is a variable-stress language where stress is a lexical property of words [1] that can be used contrastively (e.g., *voorkomen*, ‘to happen’, vs. *voorkomen*, ‘to prevent’). On the supra-lexical level, Dutch uses ‘accents’ to signal the informational status (linked to the concept of ‘focus’) of words. Stress is determined by the linguistic system, whereas accent depends on the communicational aims of a speaker [2: 41]. Dutch word stress (measured on words out of focus) is acoustically correlated (mainly) with duration, spectral tilt, and to a lesser extent overall intensity and timber. Dutch accent is mainly rendered by abrupt changes in f_0 , duration, spectral tilt and overall intensity [2].

French, on the other hand, does not have lexical contrastive stress: the standard ‘primary accent’ typically falls on the last syllable of ‘accentual’ word groups [e.g. 3, 4, 5]. Rather than being contrastive, this ‘primary’ accent has a demarcative function [6]. Besides this primary accent, French has several secondary accents falling on any syllable of the word group and covering rhythmic or emphatic functions [7]. The acoustic correlates of the primary accent are mainly duration, a change in f_0 and the potential use of pauses. An initial emphatic accent is rendered by a shorter duration but a change in f_0 , potentially preceded by a pause [7].

Although Dutch is taught in most primary and secondary schools in Francophone Belgium, pronunciation and prosody are often neglected in DFL courses according to our surveyed students and teachers [also 8]. This means that most learners may not be familiar with Dutch prosody.

The production of Dutch word stress by Francophones has been addressed in studies on Dutch as a Second Language (DSL) [9] and as a DFL [10, 11]. Based on the results of these studies it seems clear that the DFL population has to be analysed separately from the DSL one, as the latter group, probably as a result of receiving another type of input (viz. native spoken Dutch), has been found to be more proficient in producing correctly located stress in simple and complex words. As for the DFL group, it was concluded that learners tend to stick to their final L1 pattern, but can also evolve to a penultimate stress (yet not always being the required stress position) with time. Research on DFL stress in nominal compounds (with the first compounding part bearing stress as a main rule) [12] has also shown that, in addition to a preference for the final syllable, DFL speakers do not necessarily show a consistent stress pattern across words.

The current research focuses on Dutch stress production by Belgian Francophone learners. Our analysis first concentrates on the realised position of the stress by DFL speakers. This analysis relies on a perceptual annotation by multiple annotators. The considerable amount of inter-rater disagreement also led us to investigate the acoustic realisation of prominences by both DFL and native speakers. Discrepancies between the groups might explain annotation confusion and annotators might rely on different acoustic correlates.

In this paper we also present some methodological aspects of our study. The tools we developed provide a detailed analysis of some aspects of DFL and native Dutch (DL1) prosody.

The paper is organized as follows. Section 2 presents the methodology and tools developed to phonetically align, annotate and analyse our corpus. The analysis of the prominences produced by DFL and native speakers is then presented and discussed in Section 3. Finally, Section 4 concludes the paper and discusses further works.

2. Method

2.1. Participants

20 DFL learners (age range 19-23, mean age 21.1, 14f, 6m) and 10 native speakers of Belgian Dutch (age range 20-51, mean age 28.6, 5f, 5m) took part in the experiment. French was the only mother tongue of the selected DFL speakers.

2.2. Materials

30 existing Dutch three-syllable words were used in the current study. They were selected and classified according to the stress rules for simplicia described in [13]. They were split into three canonical stress positions (SP): initial (*pagina*, ‘page’), medial (*collega*, ‘colleague’) and final (*anoniem*, ‘anonymous’). Each word X was randomly presented thrice in a carrier sentence (X heb ik gezegd ‘X I said’, Ik heb X gezegd

‘I X said’ and *Ik heb gezegd X* ‘I said X’), leading to a 90-sentence reading task. Each target was presented in bold, italics and was underlined, showing focus marking.

2.3. Procedure

Speakers were recorded individually in a quiet room. Prior to the recording they filled in a form containing questions about their learner profile (duration of Dutch learning, age at start of learning, etc.). The trial phase started after an instruction and training session similar to the trial. A Tascam-07 MKII recorder and a Sennheiser PC131 head-set microphone were used.

2.4. Perceptual analysis

The data were perceptually labelled independently by two DFL-speaking native-French speakers and three native Dutch speakers, all of whom were phonetically trained. After listening to the stimuli as often as required, the annotators indicated which syllable they perceived as prominent (1-2-3). Cases of doubt could be expressed as “1?3?”, etc. The annotators also gave a certainty score on a scale from 1 (very easy) to 5 (very difficult) representing the difficulty of making their decision as to which syllable was most prominent.

2.5. Linguistic analysis

An alignment between the speech signal and the phonetic transcription was necessary for the prosodic analysis of the syllables. Since manual alignment is a tedious and time-consuming task, several automatic alignment tools have been proposed in the literature [e.g. 14, 15]. They usually rely on pre-trained speaker-independent models to align new corpora. However, they cover a very limited number of languages and might not perform properly for different speaking styles. Most of these existing tools actually do not provide models for Dutch. To resolve this issue, we developed a new automatic phonetic alignment tool, *Train&Align* [16]. Its specificity is that it trains the models directly on the corpus to align, which makes it applicable to any language and speaking style. Previous experiments have shown that it provides results comparable to the other existing tools [17]. It also offers additional options like “bootstrap”, allowing for a manually-aligned part of the corpus to be used to improve the model quality. While a basic alignment of our corpus with *Train&Align* achieved rather poor alignment rates, the use of 40 seconds of bootstrap led to significant improvement, reaching alignment rates of about 82% with a 20 ms tolerance threshold. The aligned files provided in TextGrid format were easily imported in *Praat* for further prosodic analyses. This alignment was then manually checked.

The corpus was then syllabified in *Praaline* [18] (cf. 2.6). A basic rule-based syllabifier relies on the sonority sequencing principle (sonority should increase from the first phoneme of an onset to the nucleus), and the maximal onset principle, which states that a syllable’s onset should be extended at the expense of the preceding syllable’s coda. Such a simple rule-based approach has been shown to achieve an accuracy of 93-95% [19]. The syllabifier was adapted to Dutch by providing a list of valid onsets. The syllabification was manually checked.

2.6. Corpus processing

In order to further process the data, we used *Praaline* [18], a toolkit for corpus management, annotation, querying and

visualization. It interfaces with *Praat* and stores corpus data as a relational database, allowing the user to add external data sources. The annotator labels from the perceptual analysis were imported and linked to the corresponding corpus syllables. *Praaline* runs a cascade of scripts and/or external analysis tools, each of which may add features to an annotation level (e.g. syllables, words etc.). Using this interface, we applied *Prosogram* [20] for pitch stylisation on the entire corpus. *Prosogram*’s algorithm operates in two phases; for each syllable, vocalic nuclei are detected based on intensity and voicing. The f_0 curve on the nucleus is then stylised into a static or dynamic tone, based on a perceptual glissando approach. Several syllable features (duration, pitch, pitch movement etc.) were added to the database.

Subsequently, we constructed the datasets for further statistical analysis using a query editor. *Praaline* queries may include data from multiple levels of annotation, and the features of one level may be aggregated or normalised over another level. In this study, we correlated the perceptual annotations to the prosodic features of syllables. For each prosodic feature, we also calculated a z-score value normalised over each speaker. Queries may also include functions to calculate derived measures; we used this feature to obtain relative measures (cf. 3.4). The statistical analysis was performed using SPSS (v. 21) and R (v. 3.0.2).

3. Results

3.1. Perceived stress position

3.1.1. Inter-rater agreement

An inter-rater agreement analysis using the Fleiss’ Kappa statistic [21] was performed on the DFL and DL1 data to determine consistency among annotators. “Low certainty” comprises all the cases for which the majority of annotators ($n \geq 3$) expressed a low confidence ordeal about the decision they made as to which syllable was bearing stress (certainty score > 1 on a scale from 1 to 5, see 2.4.). “High certainty” refers to high confidence in the annotators’ decisions (score = 1, see 2.4.). As shown in Table 1, the κ -values are always lower for the DFL group than for the control. For all cases taken together the agreement is moderate ($\kappa = 0.570$) for the DFL group [22] and almost perfect for the DL1 group ($\kappa = 0.980$). Cases of high certainty comprise 75.88% of the cases for the DFL and 98.00% for the control group. While κ is substantial for the DFL group and almost perfect for the DL1 group for cases with high certainty, the κ -value drops to fair levels for low-certainty cases.

| | DFL | DL1 |
|----------------|----------------|---------------|
| General | 0.570 (n=1799) | 0.980 (n=900) |
| High certainty | 0.681 (n=1365) | 0.984 (n=882) |
| Low certainty | 0.239 (n=434) | 0.385 (n=18) |

Table 1: *Inter-rater agreement (Fleiss’ κ and counts) for overall annotations (“General”), and cases with high and low certainty per L1 group.*

3.1.2. Consensus

Based on the annotations of each labeller, a consensus variable was computed per word. Consensus is reached when *all* annotators marked the same syllable as prominent. Table 2 shows the consensus values per canonical stress position (SP)

for the DFL speakers. The shaded cells contain the cases where canonical and perceived stress concur, meaning that the stress was perceived and therefore probably produced in the expected position. The overall percentage of “correct” stress amounts to 26.7% (vs. 96.1% for the control group). Consensus over each syllable is not equally distributed over the three canonical SPs ($\chi^2(4) = 188.69, p < .001$). Canonical SP3 yields the best results (39.90% correct), followed by SP2 (25.8%) and SP1 (14.3%). On the whole DFL speakers tend to stress the 3rd syllable most often regardless of the canonical SP (25.80%), confirming our hypothesis. However this result is mainly due to the high percentage of 3rd-syllable stress in SP3.

| Canonical SP | No consensus | | Syll 1 | | Syll 2 | | Syll 3 | | Total | |
|--------------|--------------|-------|--------|-------|--------|-------|--------|-------|--------|--------|
| | % | (n) | % | (n) | % | (n) | % | (n) | % | (n) |
| 1 | 53.20 | (319) | 14.30 | (86) | 12.80 | (77) | 19.70 | (118) | 100.00 | (600) |
| 2 | 52.00 | (312) | 4.30 | (26) | 25.80 | (155) | 17.80 | (107) | 100.00 | (600) |
| 3 | 57.70 | (286) | 6.50 | (39) | 5.80 | (35) | 39.90 | (239) | 100.00 | (599) |
| | 51.00 | (917) | 8.40 | (171) | 14.80 | (267) | 25.80 | (464) | 100.00 | (1799) |

Table 2: Percentages (and counts) consensus between annotators for 1st, 2nd and 3rd canonical stress position broken down by perceived stress position.

Strikingly, in 51.00% of the cases no consensus is reached. There are two possible reasons for this: either one or more annotators did not label the same syllable as being prominent due to a difference in perception or in acoustic correlates, or the speakers have produced multiple prominences within the same word, leading to ‘doubt’ cases (e.g. 1?3?). Table 3 shows all these cases where annotators labelled multiple syllables as prominent. “1?3?” doubt cases should be viewed separately from the other doubt cases as they probably signal double prominences within a word. The other cases might point to the use of ambiguous acoustic correlates.

| Doubt cases (n = 623) | DFL |
|-----------------------|--------------|
| 1?2? | 4.49% (28) |
| 1?3? | 80.90% (504) |
| 2?3? | 12.20% (76) |
| 1?2?3? | 2.41% (15) |

Table 3: Percentage (and counts) of perceived multiple prominences for the DFL group within all doubt cases.

As previously mentioned (see 3.1.), the annotators also gave a high-certainty score in 75.88% of the cases. This is interesting as it shows that sometimes the annotators showed great confidence in their annotations whereas they did not perceive the same syllable as being the most prominent one.

3.1.3. Analysis of the correct vs. incorrect results

For this analysis all cases where consensus and canonical SP concur were labelled as “correct” and all others as “incorrect”. A repeated measures ANOVA with Greenhouse-Geisser correction with canonical SP and word position in the sentence as within-subjects factors and L1 of the speakers as between-subjects factor was carried out on the percentage of (in)correct cases. As expected, an effect of L1 ($(F(1, 28)) = 28.92, p < .001$), but also of position of the word in sentence ($(F(1.761, 49.308)) = 3.917, p < .005$) was found. However, there is no effect of canonical SP for both speaker groups taken together ($(F(1.80, 50.34)) = 3.04, n.s.$). The analysis also reveals an interaction between canonical SP and L1 ($(F(1.80, 2.47)) =$

3.38, $p < .05$), and position in sentence and L1 ($(F(1.76, 68.96)) = 5.92, p < .001$).

The same analysis was carried out per L1 group as an interaction between SP and L1 and word position in sentence and L1 had been found. For the DFL group an effect of SP ($(F(1.70, 32.13)) = 10.33, p < .001$) and position in sentence was found ($(F(1.79, 33.95)) = 6.64, p < .05$). Pairwise comparisons show that the effect of SP is caused by the difference between SP1 (14.3% correct, see Table 2) and SP3 (39.90%) but not with SP2 (25.8%). The same appears to be true for word position in sentence: 21.3% of cases in sentence-initial, 27.7% in sentence-medial, 31.1% in sentence-final position are correct. Figure 1 shows the percentage of correct results per canonical SP and position in sentence. There seems to be a trend towards more correct results when words are sentence-final, especially on the 3rd syllable (cf. our hypothesis). However, this result does not reach significance as the interaction between SP and position in sentence is not significant ($(F(2.44, 45.39)) = 1.85, n.s.$).

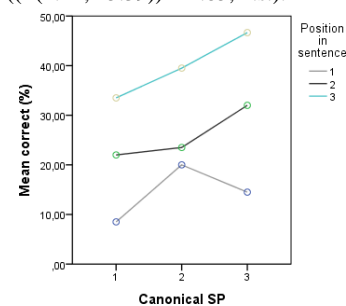


Figure 1: Percentage correct cases per canonical SP and word position in sentence for the DFL data.

For the control group, no effect of SP ($(F(1.70, 15.31))=1.39, n.s.$) or position in sentence ($(F(1.68, 15.01))=0.08, n.s.$) and no interaction between them ($(F(4, 36))=0.41, n.s.$) was found.

3.2. Acoustical realisation of prominent vs. non-prominent syllables

In order to study the prosodic correlates of perceived prominent syllables for the DL1 and DFL groups, we extracted several acoustic features of syllables. Inspired by the methodology presented in [23], four features were studied:

- Relative mean pitch: the difference of a syllable’s mean pitch relative to the mean pitch of the word (in semitones);
- Pitch movement: intra-syllabic upwards or downwards movement (in semitones);
- Relative vowel duration: the ratio of the vocalic nucleus of a syllable, relative to the duration of the vocalic nuclei of the word;
- Relative peak intensity: the difference of a syllable’s peak intensity in the vocalic nucleus relative to the mean intensity of the word (in dB).

These features are typically correlated with syllabic prominence. Syllables were included in the statistical analysis only when pitch could be detected and stylized by *Prosogram* in their corresponding word: in total 6891 syllables were analysed (2319 native, 4572 non-native; 1918 stressed, 4461 unstressed). The distributions of the four acoustic measures are

shown in Figure 2. Relative pitch and relative intensity follow a normal distribution. Relative vowel duration is positively skewed, while pitch movement follows a bimodal distribution, corresponding to falling and rising pitch.

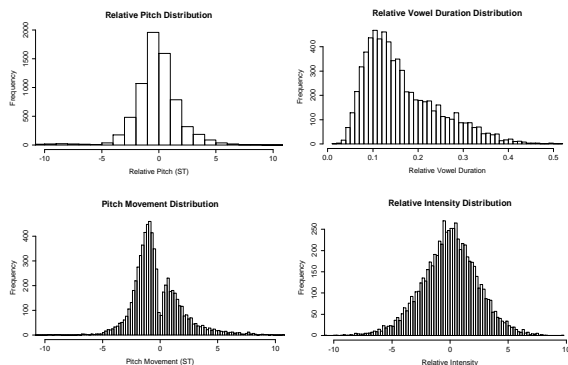


Figure 2: Distribution of the prosodic correlates of syllabic prominence under study (from left to right, top to bottom: relative pitch (ST), relative vowel duration (ratio), pitch movement (ST), relative intensity (dB)).

A multivariate ANOVA was carried out on every acoustic measure; the fixed factors were L1 and presence/absence of prominence. Table 4 shows that the language groups make different use of duration ($p < 0.001$) and maximum pitch ($p < 0.05$), but use similar overall pitch and intensity strategies. There is a significant difference for every acoustic measure between prominent and non-prominent syllables ($p < 0.001$) as well as an interaction between both factors ($p < 0.001$, $p < 0.05$ for falling pitch movement).

| Acoustic measure \ Factor | L1 | Prominence | L1x Prominence |
|----------------------------|------|------------|----------------|
| Relative syllable duration | ** | ** | ** |
| Relative nucleus duration | ** | ** | ** |
| Relative vowel duration | ** | ** | ** |
| Relative pitch minimum | n.s. | ** | ** |
| Relative pitch maximum | * | ** | ** |
| Relative mean pitch | n.s. | ** | ** |
| Pitch movement, rising | n.s. | ** | ** |
| Pitch movement, falling | n.s. | ** | * |
| Relative intensity | n.s. | ** | ** |

Table 4: Main effects for L1, absence/presence of prominence and interaction between them on all acoustic measures (* $p < 0.05$; ** $p < 0.001$)

In order to assess the relative importance of each prosodic correlate for prominent and non-prominent syllables for both groups, we applied a binomial logistic regression model. A syllable was considered prominent or not (binary dependent variable) as long as there was a consensus (per syllable) between 4 or all 5 annotators. The acoustic measures were the model's predictors. In the DL1 group, relative mean pitch was found non-significant; all other predictors were significant with $p < 0.001$. Table 5 summarises the standardised beta coefficients and z-scores for each predictor for the two models. The results suggest that DL1 speakers signal prominence mainly through relative intensity, duration and rising pitch movement (in decreasing order of importance). DFL speakers on the other hand, use duration, then rising pitch movement and relative mean pitch. It is noteworthy that these

were found to be the main prosodic correlates of syllabic prominence in French (along with succeeding pauses) [7].

| Acoustic measure | DL1 | | DFL | |
|-------------------------|---------|------|---------|------|
| | β | z | β | z |
| Relative vowel duration | 1.749 | 10.9 | 1.942 | 18.2 |
| Pitch movement, rising | 2.083 | 10.8 | 1.934 | 14.2 |
| Pitch movement, falling | 0.451 | 3.1 | 0.628 | 5.8 |
| Relative mean pitch | -0.221 | n.s. | -0.976 | -8.4 |
| Relative intensity | 3.715 | 18.4 | 0.325 | 3.4 |

Table 5: Standardised β coefficients and z-scores for the DL1 and DFL logistic regression models predicting prominence.

4. Conclusion and perspectives

This paper investigated the realisation of Dutch stress by Belgian Francophone learners of Dutch. Our study showed low scores of correct stress position for the DFL group, pointing at their poor grasp of Dutch word stress position. On the whole, the DFL speakers relied on their final L1 pattern but mainly in canonical SP 3. This globally supports our hypothesis. There also seems to be a trend towards more correct 3rd-syllable stress in sentence-final position, but this result does not reach significance.

The manual annotation of perceived prominence in the corpus sometimes reached low agreement rates. In an attempt to explain this phenomenon, acoustical analyses were carried out. The analyses of variance seem to signal a different use of duration and maximum pitch by the language groups. The binomial logistic regression model points out that the DL1 group uses relative intensity, duration and rising pitch movement to signal prominence. The DFL group uses L1 accentuation strategies (duration, rising pitch movement and relative mean pitch). If annotators are more sensitive to different sets of acoustic correlates, this would explain the overall low consensus.

Further studies will focus on the comparison of the acoustic realisation of prominent syllables with high vs. low-certainty score cases. Speaker variability will also be investigated as it might also account for the lack of annotation agreement. Furthermore the analysis of the “doubt” cases (Table 3) should help us find evidence for multiple prominences within words.

While our study relied on a binary prominence label, it should be noted that research [e.g. 23, 24, 25] suggests that syllabic prominence is perceived as a gradual rather than a binary phenomenon. A manual annotation of relative prominence levels might give more insight into DFL stress production. Finally, it should be highlighted that the studied acoustic correlates are actually those of focus accent. While lexical stress should be studied in words out of focus [2], the main goal of our study was rather to compare stress position between DFL and DL1 speakers, stress and accent falling on the same syllable in Dutch. In an attempt to avoid total lack of prominence on the DFL stimuli, all words were put in focus. The acoustic correlates are used here as an attempt to explain perceptual differences between annotators and should not be considered as acoustic correlates of lexical stress.

5. Acknowledgments

The first two authors are supported by F.R.S.-FNRS grants. We would like to thank Dr. Thomas François for his advice on κ -coefficients.

6. References

- [1] Rietveld, A. C. M., & Heuven, V. J., van. *Algemene Fonetiek*. Bussum: Coutinho, 2009.
- [2] Sluijter, A. *Phonetic Correlates of Stress and Accent*. Leiden University, Den Haag: HILL, 1995.
- [3] Lacheret-Dujour, A., & Beaugendre, F. *La prosodie du français*. Paris: CNRS Editions, 1999.
- [4] Di Cristo, A. “Vers une modélisation de l’accentuation du français (seconde partie)”, *French Language Studies*, 10, 27-44, 2000.
- [5] Rasier, L. *Prosodie en vreemdetaal-verwerving. Accentdistributie in het Frans en Nederlands als vreemde taal*. Diss. Université catholique de Louvain, 2006.
- [6] Vaissière, J. “Cross-linguistic prosodic transcription: French versus English”, *Problems and methods in experimental phonetics*, In honour of the 70th anniversary of Prof. L.V. Bondarko. St.-Petersburg, St.-Petersburg State Univ, 147-164, 2002.
- [7] Simon, A.C. *Phonologie et Prosodie*. Université catholique de Louvain, 2009.
- [8] Hiligsmann, Ph. “Uitspraakvaardigheid van Franstalige leerders van het Nederlands: een ondergeschoven kind?”, *Handelingen*, LII, 157-167, 1999.
- [9] Caspers, J., & Santen, A., van. “Nederlands uit Franse en Chinese mond. Invloed van T1 op de plaatsing van klemtoon in Nederlands als tweede taal?”, *Nederlandse Taalkunde*, 11 (4), 289-318, 2006.
- [10] Heiderscheidt, S., & Hiligsmann, Ph. “De accentuering in de tussentaal van Franstalige leerders van het Nederlands”, *Leuvense Bijdragen*, 89 (1/2), 17-131, 2000.
- [11] Michaux, M.-C., Hiligsmann, Ph., & Rasier, L. “Het klemtoonpatroon in de tussentaal van Franstalige leerders van het Nederlands”, XII. Internationaler Germanistenkongress Warschau 2010: Vielheit und Einheit der Germanistik weltweit, Warsaw, 321-332, 2012.
- [12] Bui, A.V. *De woordklemtoon in de tussentaal van Franstalige leerders van het Nederlands: De beklemtoning van samenstellingen*, unpublished thesis, Université catholique de Louvain, 2012.
- [13] Trommelen, M. & Zonneveld, W. *Klemtoon en metrische fonologie*. Muidergerg: Coutinho, 1989.
- [14] Goldman, J.-Ph. “EasyAlign: an automatic phonetic alignment tool under Praat”, *Proceedings of Interspeech*, 3233-3236, 2011.
- [15] Bigi, B. & Hirst, D. “Speech phonetization alignment and syllabification (SPPAS): a tool for the automatic analysis of speech prosody”, *Proceedings of Speech Prosody ‘12*, 19-22, 2012.
- [16] Brognaux, S., Roekhaut, S., Drugman, T. & Beaufort, R. “Train&Align: A new online tool for automatic phonetic alignments”, *IEEE Workshop on Spoken Language Technologies*, 410-415, 2012a, Online: http://central.fltr.ucl.ac.be/train_and_align/
- [17] Brognaux, S., Roekhaut, S., Drugman, T. & Beaufort, R., “Automatic phone alignment. a comparison between speaker-independent models and models trained on the corpus to align”, *Lecture Notes in Computer Science*, 7614, 300-311, 2012.
- [18] Christodoulides, G. “Praaline: integrating tools for speech corpus research”, *Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland, 26-31 May 2014.
- [19] Bartlett, S., Kondrak, G. & Cherry, C., “On the syllabification of phonemes”, *Proceedings of NAACL’09*, 308-316, 2009.
- [20] Mertens, P., “The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model” in B. Bel & I. Marlien [Eds], *Proceedings of Speech Prosody 2004*, Nara, Japan, 23-26 March, 549-552, 2004.
- [21] Geertzen, J. (2012). “Inter-Rater Agreement with multiple raters and variables”, accessed on 23 December 2013, from <https://mlnl.net/jg/software/ira/>
- [22] Landis, J. R., & Koch, G. G. “The measurement of observer agreement for categorical data”, *Biometrics*, 33 (1), 159-174, 1977.
- [23] Goldman, J.-Ph., Avanzi, M., Auchlin, A. & Simon, A.C., “A Continuous Prominence Score Based on Acoustic Features”, *Proceedings of Interspeech*, Portland, Oregon, USA, 9-13 September, 2454-2457, 2012.
- [24] Avanzi, M., Lacheret, A., Obin, N. & Victorri, B., “Vers une modélisation continue de la structure prosodique: le cas des proéminences”, *Journal of French Language Studies* 21, 53-71, 2011.
- [25] Arnold, D., Wagner, P. & Möbius, B., “Obtaining prominence judgments from naïve listeners. Influence of rating scales, linguistic levels and normalisation”, *Proceedings of Interspeech*, Portland, Oregon, USA, 9-13 September, 2012.