CALCU-DISTA: A TOOL FOR DIALECTOMETRIC ANALYSIS OF INTONATIONAL VARIATION[*]

CALCU-DISTA: UNA HERRAMIENTA PARA EL ANÁLISIS DIALECTOMÉTRICO DE LA VARIACIÓN ENTONATIVA

PAOLO ROSEANO
Universitat de Barcelona
paolo.roseano@ub.edu

WENDY ELVIRA-GARCÍA
Universitat de Barcelona
wendyelvira@ub.edu

ANA Mª FERNÁNDEZ-PLANAS
Universitat de Barcelona
anamariafernandez@ub.edu

ABSTRACT

During the last few decades, the size of databases containing data about prosodic diatopic variation has increased dramatically. For this reason, researchers need adequate statistical methods and tools to cope with such huge amount of information. This paper aims at describing Calcu-Dista, a tool whose aim is carrying out a dialectometric analysis of acoustic intonational data within the framework of the AMPER project. In addition to describing how Calcu-Dista works, this paper provides validation of its results.

Keywords: Calcu-Dista, dialectometry, diatopic variation of intonation, geoprosody.

RESUMEN

Durante las últimas décadas, el tamaño de las bases de datos que contienen datos sobre la variación diatópica de la prosodia ha aumentado considerablemente. Por esta razón, los investigadores necesitan instrumentos y métodos estadísticos adecuados para hacer frente a esa enorme cantidad de información. Este trabajo pretende describir Calcu-Dista, una herramienta cuyo objetivo es llevar a cabo un análisis dialectométrico de datos entonativos

acústicos en el marco del proyecto AMPER. Además de describir cómo funciona Calcu-Dista, este artículo proporciona la validación de sus resultados.

1. INTRODUCTION

1.1. Objectives

This paper aims at presenting a new dialectometric tool, Calcu-Dista, which was designed to analyze diatopic variation of intonational data. In order to explain why such tool was needed, first of all one needs to remind some basic characteristics of diatopic intonational data (See § 1.2.) and of diatopic intonational databases of Romance languages (See § 1.3.). In the following section, we shall present a short history of dialectometric analyses of intonational data before the creation of Calcu-Dista (See § 2.). Calcu-Dista will then be presented in detail in § 3. Finally (See § 4.), we shall provide some conclusive remarks.

1.2. Data of diatopic intonational variation

Data used in phonetic and phonological research, in very broad terms, belong to two main families: quantitative data and categorical data. Usually, such data types are not equally used in the different traditional branches of speech studies. In articulatory and acoustic phonetics, in fact, researchers usually deal with numeric data (e.g. the number of electrodes activated by contact with the tongue in electropalatographic studies, the formant values of a vowel or the F0 values in spectrographic analyses, and so forth). In auditory phonetics, on the other hand, the most common data types are not strictly quantitative, insofar as they usually are categorical (e.g. "A/B" answers in identification tasks) or scale (e.g. Likert-scale answers in discrimination tasks). In phonological research data tend to be categorical, insofar as the most usual representations of speech phenomena at this level are either symbols (like IPA symbols or ToBI labels) or features (like, for example, ±ATR, ±voice, and so on).

The above-mentioned scheme is quite evident in segmental speech studies, but can be applied to suprasegmental studies, too. In fact, intonational diatopic variation of Romance languages has been approached from all four the disciplinary standpoints we have previously mentioned (articulatory phonetics, acoustic

phonetics, auditory phonetics, phonology). This notwithstanding, it has not been paid the same attention in all four branches. As a matter of fact, diatopic intonational variation in Romance languages has hardly been studied from an articulatory point of view: to the best of our knowledge, Vanrell, Stella, Gili-Fivela and Prieto (2013) experiment represents, as of today, the only attempt to describe possible geographic differences in F0 movements among Romance languages basing on articulatory analyses.

Acoustic approaches to geoprosodic differences among varieties of this linguistic family are definitely more common. In addition to the acoustic F0 analyses carried out within the framework of the AMPER project (See § 1.3. for more details), there is a considerable number of other acoustic studies of geoprosodic varieties (like Henriksen, 2010, for Spanish; Font-Rotchés, 2007, for Catalan; Endo and Bertinetto, 1997, for Italian; Apopei, Jitcă and Turculeţ, 2006, for Rumanian, among many others).

Auditory studies (that include perception tests) are common in the field of intonation of Romance languages, especially within Laboratory Phonology approaches (like D'Imperio, 2000; Vanrell, 2006; Gili-Fivela, 2009; Crespo-Sendra, 2011; Cruz and Frota, 2011; or Borràs-Comes, Vanrell and Prieto, 2014, among many others). Nevertheless, these perception studies in most cases do not focus on diatopic variation, but rather on the intonational phonology of one single dialect of a language. The few remarkable exceptions are researches whose aim is establishing the perceived distances between intonational patterns of distinct dialects or languages (like Woehrling and Boula de Mareüil, 2006; Martínez, Fernández Planas, Dorta and Fernández Rei, 2006; Simon, Hambye, Bardiaux and Boula, 2012; Gili-Fivela, 2013; Boula de Mareüil and Akissi Boutin, 2011; Boula de Mareüil, Rilliard, Ivent and Kozhevina, 2016; Fernández Planas, Roseano, Elvira-García, Carrera and Román, 2017).

Phonological descriptions of the intonation of a single dialect or of several (if not all) diatopic varieties of a language are more and more frequent. Whereas until the end of the 20th century one could find a limited number of phonological descriptions of the intonation of non-standard geographic varieties of Romance languages (like, for Italian regional varieties, the studies by Grice, 1995; Caputo, 1996; Sorianello, 1996; D'Imperio, 1997; Sardelli, 1998; Besana, 1999; Marotta and Sorianello, 1999, and very few more), in the following decades the amount of publications of this kind has increased notably (for Italian, for example, Gili-Fivela et al., 2015, contains a review of the literature that gives an idea of the boost in the number of publications about the intonational phonology of the regional varieties of that language since the beginning of the century). In the second decade of the 21st century, moreover, different researchers have carried out the first attempts to describe the diatopic intonational variation of several Romance languages from a phonological point of view (see Prieto and Roseano, 2009-2013, and Hualde

and Prieto, 2015, for Spanish; Carrera, Fernández and Martínez, 2010; Prieto and Cabré, 2007-2012, and Prieto et al., 2015, for Catalan; Gili Fivela et al., 2015, for Italian; Delais-Roussarie et al., 2015, for French; Roseano, Vanrell and Prieto, 2015, for Friulian; Sichel-Bazin, Meisenburg and Prieto, 2015, for Occitan; Frota et al., 2015, for Portuguese; Jitcă, Apopei, Păduraru and Marusca, 2015, for Romanian; Vanrell, Ballone, Schirru and Prieto, 2015, for Sardinian).

The increase in the number of acoustic and phonological studies about intonational diatopic variation since the mid-Nineties has gone hand in hand with two other processes. On one hand, researchers have begun to build databases of intonational diatopic variation (See § 1.3. for more details). And, on the other hand, the willingness to give comprehensive interpretations of diatopic intonational variation of the several Romance languages has led the researchers to try and draw the borders of intonational dialects[1]. In order to do so, intonologists have made use of both qualitative and quantitative techniques to help them define the prosodic dialectal clusters (in § 2.2. and 2.3. we shall describe more in detail some of the techniques in question).

1.3. Databases for diatopic intonational variation

Thanks to the field researches carried out during the last two decades, we nowadays have access to several databases that were specifically designed to collect data about the prosody of Romance languages and of their dialects. There are currently two main groups or "families" of open databases about the prosody of geographic varieties of Romance languages: the AMPER family (Contini, 1992) and the IARI family (Prieto, Borràs-Comes and Roseano, 2010-2014; Frota and Prieto 2015). Both projects are not only similar in their aims, but also in their organization and in their results. Nevertheless, they show differences in the data that are usually analyzed. As far as the organization is concerned, both projects have a central steering committee and research groups for each language (and sometimes, in the case of geographically vast languages, regional research groups). Data are made public online in the central webpage of the each project, with samples of all languages (See Prieto et al., 2010-2014, for IARI), and on the websites of the research groups that collect data of specific languages (Martínez and Fernández Planas, 2003-2016a; and Prieto and Cabré, 2007-2012, for Catalan; Prieto and Sichel-Bazin, 2014, for Occitan; Prieto and Roseano, 2009-2013, and Martínez and Fernández Planas, 2003-2016b, for Spanish; Frota and Cruz, 2012-2015, and

---

[1] One of the interesting conclusions of the overwhelming majority of geoprosodic studies is that intonational isoglosses do not usually coincide with other kinds of isoglosses (segmental, morphological or lexical).

Castro Moutinho, 2008-2013, for Portuguese; Fernández Rei, 2016, for Galician; Roseano and Fernández Planas, 2009-2013, for Friulian; Turculeţ, 2010-2013, for Romanian, among others). Table I offers a summary of the available online databases for each language within both projects (databases that are not available online are not included).

Table I. Online databases within the AMPER and IARI frameworks.

	AMPER	IARI
Portuguese	Yes	Yes
Galician	Yes	
Asturian	Yes	
Spanish	Yes	Yes
Aragonese		
Catalan	Yes	Yes
Occitan		Yes
French		Yes
Francoprovençal		
Sardinian		Yes
Italian		Yes
Romansh		
Ladin		
Friulian	Yes	Yes
Romanian	Yes	Yes
GENERAL	Yes	Yes

As far as the similarities between the corpora are concerned, one has to point out that the databases of both projects contain sentences obtained by means of a Discourse Completion Task (henceforth DCT; Blum-Kulka, House and Rasper, 1989; Billmyer and Varghese, 2000; Félix-Brasdefer, 2010), by means of Map Task (henceforth MT; Anderson et al., 1991), by means of an interview and, in AMPER only, also a corpus of acted laboratory speech sentences, called "textual elicitation corpus" within the framework of the project (Romano, Lai and Roullet, 2005). Within AMPER, this laboratory speech corpus has been paid the most attention, whereas DCT, MT and interviews have been analyzed much less extensively. On the contrary, DCT sentences are the kind of items that have been given priority in the IARI family (while MT and interviews have not been studied in depth). This represents the first relevant difference between the two projects.

The second important difference between AMPER and IARI is the method of analysis that, on its turn, determines the kind of information that is contained in the above mentioned databases. Whereas within the IARI family the predominant method of analysis is an auditory and phonological one, within the AMPER family the approach is mainly phonetic and spectrographic. This has important consequences on the type of data that constitute the final dabatase. Whereas IARI online databases usually offer the F0 contour and the phonological labeling of one single example of a comparative large number of sentence-types per locale (from 18 to 47, depending on the language), AMPER databases offer the numeric values (in Hz and st, ms and dB) of the prosody of several utterances (from 378 to 567, depending on the language) of two sentence-types only (broad focus statements and information-seeking yes-no questions). Figure 1 contains an example of intonational data offered in the Spanish online atlas of the IARI family (Prieto and Roseano, 2009-2013).
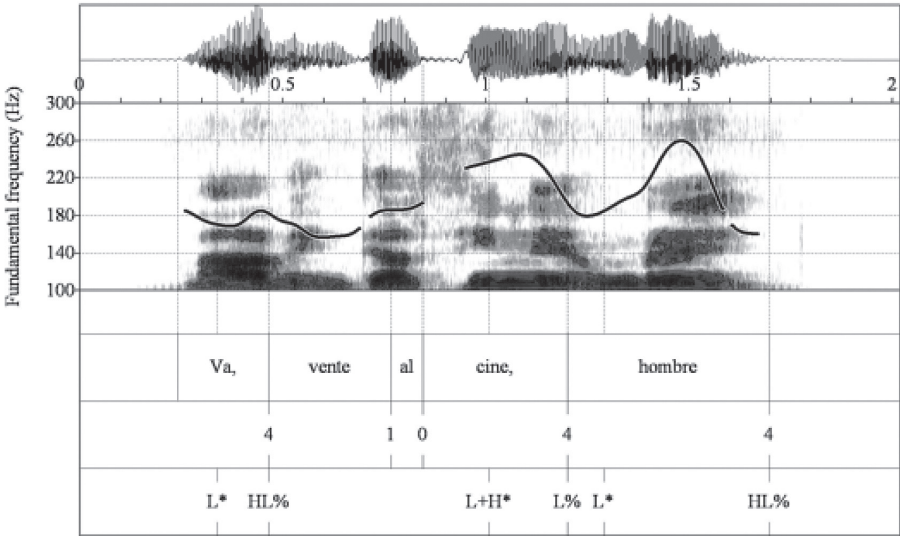
Figure 1. Example of intonational data contained in the Interactive Atlas of Spanish Intonation database (F0 contour, phonological transcription of intonation).

On the other hand, the AMPER databases do not usually provide any phonological labeling, but they contain selected numerical values that summarize the prosody of each sentence; basically, for each syllable of the sentence the database contains three F0 values (in Hz and st), three intensity values (in dB) and the duration (in ms). Basing on these numeric values, a transcriber called AmperEti

(Roseano and Fernández Planas, 2013) can be used to add a narrow phonetic transcription of the contour of each sentence (Figure 2 represents the three steps of analysis in the AMPER framework). This means that, since very few years ago, the AMPER database may contain, in addition to its usual numeric prosodic phonetic values, also a set of alphabetic autosegmental-metrical representations of phonetic data.
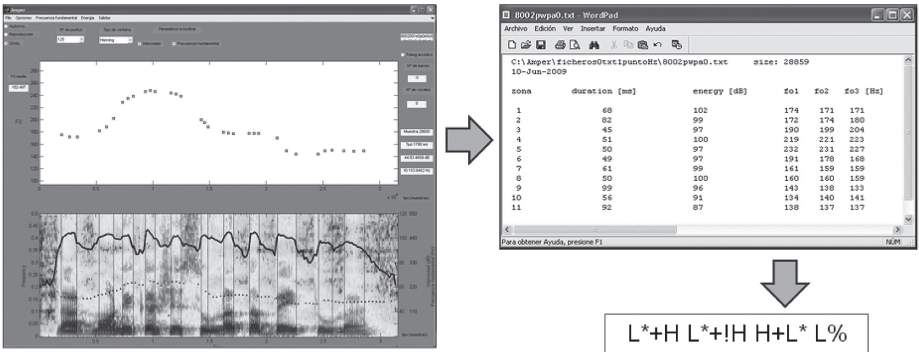
Figure 2. Main steps of the AMPER analysis (spectrographic analysis, creation of numeric database, creation of database of narrow phonetic transcriptions). The example in the picture is a Friulian broad focus statement.

Last but not least, the above-mentioned databases of diatopic variation in Romance languages share another characteristic, that it also common to all other big linguistic atlases: the analysis of their huge amount of data becomes problematic because it is very difficult, for a human being, to cope with such an enormous amount of information (just to give an idea of the quantity of data, the Catalan AMPER database contains data about 2.610 acted speech sentences form 30 speakers, and the IARI-family atlas for Catalan contains 3.332 DCT sentences from 71 speakers). In order to overcome this problem, intonational dialectologists have developed a set of specific statistical tools, which will be discussed in the following sections.

2. TOOLS TO ANALYZE DATABASES

2.1. Dialectometry

The set of statistical methods to analyze dialectal databases (usually atlases focused mostly on lexical variation and, to a minor extent, on morphological variation) is called dialectometry. As it is known, dialectometry was born in the 1970s and

80s, mainly as a result of the work carried out by Séguy (1971) and Goebl (1981). Dialectometry can be summarized in two main steps, as in Figure 3: 1) calculating numeric distances between alphabetic data (by means of ad hoc metrics, like Levenshtein's distance), 2) representing such numeric distances by means of easily interpretable graphics (like, for example, dendrograms). At present, the most known dialectometric software packages are Visual DialectoMetry (Haimerl, 2006), Gabmap (Nerbonne, Colen, Gooskens, Kleiweg and Leinonen, 2011) and DiaTech (Aurrekoetxea, Fernández-Aguirre, Rubio, Ruiz and Sánchez, 2013).
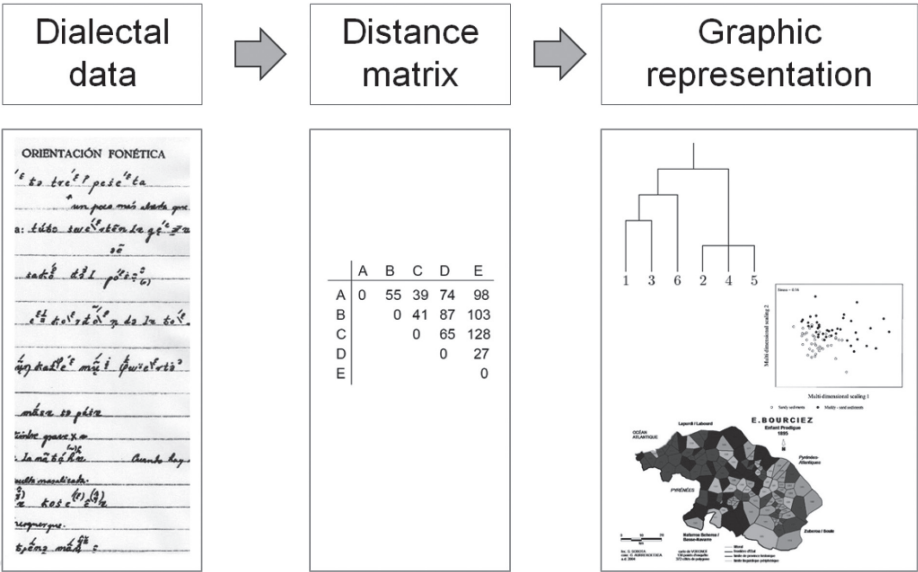
Figure 3. Main steps in dialectometry: 1) calculating numeric distances between alphabetic data, 2) representing such numeric distances by means of easily interpretable graphics.

2.2. Dialectometry of numeric intonational data

When, at the end of the first decade of the 21th century, researchers interested in diatopic variation of intonation first tried to carry out a dialectometric analysis of geoprosodic data, they had to face important problems. In fact, any dialectometric analysis requires 1) a database and 2) a software, and at the end of that decade the only existing geoprosodic database for Romance languages was the numeric part of AMPER (which was far from being complete) and the only dialectometric software was Visual DialectoMetry, which could not manage numeric data. For this

reason, intonologists needed a new dialectometric tool, which had to be capable of managing numeric intonational data. As an answer to this need, the research group headed by the Center of Dialectology of Grenoble presented a dialectometric tool known as Amper-StatDistances, created by A. Rilliard (Rilliard and Lai, 2008; Romano, Contini, Lai and Rilliard, 2011), that computed the numeric distance between two acoustic prosodic curves using a simple formula suggested by Hermes (1998a; 1998b) a few years before. Amper-StatDistances, which was used in a few studies thereafter (Castro Moutinho, Coimbra, Rilliard and Romano, 2011; Fernández Planas, Roseano, Martínez and Romera, 2011), proved to be effective in calculating prosodic distances among the locales of the AMPER project, but it had some shortcomings. On one hand, it was soon discontinued, so that the only released version was a beta one. On the other hand, due to the fact that it was a beta version, some of its statistical aspects were incomplete: it did not show the correlation matrix, it did not state which clustering technique was used to create dendrograms and MDS plots, it did not validate clusters, and it did not provide the Stress and RSQ values for MDS plots. As we shall see later, one of the reasons why Calcu-Dista was born is exactly to provide this indispensable basic statistic information.

2.3. Dialectometry of string intonational data

A few years after the first explorations in the field of intonational dialectometry of acoustic numeric values within the AMPER framework (resumed in § 2.2.), the Grup d'Estudis de Prosòdia, a research group belonging to what we informally called the IARI-family, carried out the first dialectometric analysis of intonational phonological (i.e. categorical) data. The variables used for the dialectometric analysis were the nuclear configurations[2] of 20 sentences-types in 70 Catalan dialects. The tool chosen was Gabmap (Nerbonne et al., 2011) and the results were published in Prieto and Cabré's (2013) book on Catalan intonation. Gabmap proved to be an adequate tool for basic intonational dialectometry, but it had a shortcoming: it did not allow choosing the string comparison method to be used.

2.4. Dialectometry of combined intonational data

Basing on the previous experiences (resumed in § 2.2 and § 2.3.), the team of the Phonetics Laboratory of the University of Barcelona started combining quantita-

---

[2] Since Gabmap did not allow some symbols commonly used in ToBI labels (like +, * or %), they had to be replaced with letters (e.g. L+H*L% → LpHsLb, L*H% → LsHb).

tive and categorical dialectometric analysis of geoprosodic variation. The pilot research of this dual approach was carried out on Friulian data (Roseano, 2012) by means of Amper-StatDistances (that was used to dialectometrize acoustic numeric data) and of Gabmap (that was chosen to dialectometrize the narrow phonetic transcriptions of the contours of each sentence of the Friulian corpus in question). Although the outcome of the pilot study was encouraging, insofar as the results of both kinds of dialectometric analyses coincided, the research group soon felt the need to look for new dialectometric tools that could help to avoid the limitations of Amper-StatDistances and Gabmap (See § 2.2 and § 2.3.).

3. CALCU-DISTA

3.1. The procedure

Calcu-Dista is not properly a dialectometric tool insofar as it is not a single software. It could be defined more exactly as a set of procedures, scripts and softwares, whose ultimate aim is dialectometrizing intonational data collected within the AMPER project. In the following sections we shall describe more in detail some methodological characteristics of Calcu-Dista (§ 3.2.) and, later on, we shall compare and validate the results of Calcu-Dista with those obtained by a more traditional dialectometric analysis carried out by means of DiaTech (§ 3.3). In order to follow the description, it is useful bearing in mind that both Calcu-Dista and the DiaTech-based dialectometric analysis start from the same numeric input, which is analyzed along two pathways: the strictly numeric one, or Calcu-Dista (represented by the left branch of the flowchart in Figure 4), and a string one, the AmperEti+DiaTech procedure (represented by the right branch of the flowchart in Figure 4). The second branch (i.e. the string one) relies on tools (AmperEti and DiaTech) that have already demonstrated to be reliable; for this reason the results of the string analysis will be used to validate the results obtained with the numeric procedure (i.e. with Calcu-Dista).
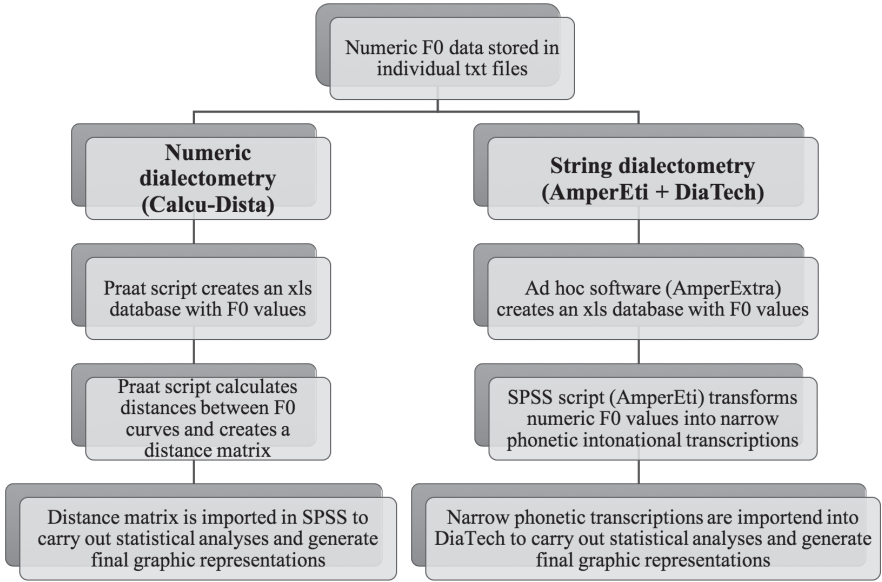
Figure 4. Flowchart representing the main steps of the numeric analysis branch (by means of Calcu-Dista) and of the string analysis branch (by means of AmperEti + DiaTech).

3.2. The common input

The common input for both the numeric and the string branches of analyses are the txt files where the prosodic information of each single utterance of the AMPER corpus are stored. Such txt files are generated by Amper06 (López Bobo et al., 2007), the acoustic analyzer created within the framework of this project, which provides a tab-separated file in txt format that contains duration, intensity and F0 values for each vowel of the utterance. Figure 5 contains an example of such txt for an 11-syllable sentence, where each row corresponds to a syllable and each column to a prosodic value (duration, energy, F0 at the beginning, middle and end of the vowel).

```
C:\Amper\ficheros0txt1puntoHz\ja01pwpi.txtsize: 29992
28-Oct-2010

zona          duration [ms]        energy [dB]        fo1   fo2   fo3 [Hz]

  1               54                  100               95    95    96
  2              106                   96              100   102   106
  3               45                   92              110   113   116
  4               49                   91              121   125   127
  5               42                   89              127   123   117
  6               61                91                 99    92    88
  7               56                   87               86    86    86
  8               56                   89               87    87    88
  9               64                   92               88    88    86
 10              112                   91               81    70    66
 11              122                   82               61    62    60

values at:
2449 2879 3310 5138 5987 6836 7559 7920 8280 9187 9580 9974 11527 11862 12197 13686 14169
14653 15290 15734 16178 17272 17718 18165 19279 19789 20298 22442 23337 24232 26024 26998
27973
```

Figure 5. Txt file with the prosodic data of a sentence of the AMPER corpus.

The AMPER corpus contains utterances of two sentence-types (broad focus statements and information-seeking yes/no questions). Both sentence-types usually have a SVO structure[3], like for example the Friulian broad focus statement La dipendent si nete la cotule 'The saleswoman cleans her skirt' or the Italian broad focus statement Il viverè mangiava la fragola 'The viceroy was eating the strawberry'. The S and the O may be stressed on the last syllable (like Italian viverè), on the penultimate syllable (like Italian bambina), or on the syllable before the penultimate (like Italian vedova), whereas the V always has the accent on the penultimate syllable (like Italian mangiava). The S and the O usually are lexical words with 3 syllables each (plus, usually, a proclitic element like an article or a preposition), while the V is a phonological word with 3 syllables. This means that, in general, in each language there will be a broad focus statement with the same number of syllables and the same accentual positions as in any other language. For example, the Friulian broad focus statement La dipendent si nete la cotule has the same number of syllables and the same accentual positions of the Italian broad focus statement Il viverè mangiava la fragola. The same holds for yes/no questions. For example, the Northern Catalan yes/no question El copista portava la caputxa? has the same number of syllables and the same accentual positions as the Italian yes/no question La bambina mangiava la banana? Thanks to these common characteristics, the corpora of the different languages in the project are easily comparable.

---

[3] There are a few exceptions to the SVO-structure rule, due to the syntactic characteristics of certain Romance languages. In addition thereto, one has to remember that the AMPER corpus also includes sentences with syntactically complex S and O (but since they have not been used in dialectometric studies with Calcu-Dista, they are not described in this paper).

3.3. The numeric analysis: Calcu-Dista

Calcu-Dista is the name chosen for a combination of Praat scripts (Elvira-García, 2014) and SPSS statistical analyses whose aim is dialectometrizing numeric acoustic intonational data. The first step of such analysis of numeric F0 values is carried out by a Praat script that extracts from the txt files all F0 values of all sentences that have to be dialectometrized and saves them in a single database. This database is then used by another Praat script that computes the prosodic distance between each utterance of every locale and all other utterances with the same sentence-type and accentual structure of the same locale and of all other locales. For example, the script computes the distance between the Friulian broad focus statement *Il mulinâr al mas̲ane la s̲emule* as uttered by a speaker form Agrons with the same statement uttered by all other Friulian speakers (including the same speaker), but also with all the utterances of the Italian broad focus statement *Il vicerè mangiava la fra̲gola*, as well as with all utterances of the Spanish broad focus statement *El saxofón se to̲ca con pánico,* of the Galician broad focus statement *O ferrolán fala̲ba co cómico,* and so on.

One of the key aspects in the dialectometric acoustic analysis carried out in Calcu-Dista is how the intonational distances between the different sentences are computed. The distance between two F0 contours (named x and y) is computed by means of the RMS of the difference between the values of F0 in the contour x and y at each measurement point, which is possibly the most simple and best known formula to compute the distance between two curves and had already been proposed and tested by Hermes (1998a; 1998b). For two sets of F0 values {f0x1, f0x2, ..., f0xN} and {f0y1, f0y2, ..., f0yN}, where N is the number of measurement points of F0 in each of the two sentences, while f0xi and f0yi are the F0 values in semitones[4] at each measurement point, the RMS is given by the formula shown below:

$$RMS = \sqrt{\frac{\sum_{i=1}^{N}(f0x_i - f0y_i)^2}{N}}$$

One of the main methodological and technical issues in the preparation of the Praat script that computes the distances with the above-mentioned formula has to do with some "irregularities" in the AMPER corpus. In fact, the number of syllables is, unfortunately, not always the same in all sentences and languages of the corpus, for different reasons. In some cases, for example, the number of syllables is theoretically right, but due to syntactic-induced vowel contact it is not possible to segment the sound chain and, as a consequence, two separate syllable end up

---

[4] Using semitones instead of Hz allows comparing the data of male and female speakers.

being analyzed as one (like in the Northern Catalan VOS yes/no question Que portava la caputxa, el copista?, where there is a vowel contact between caputxa and el, which in Catalan produces the elision of one of the two vowels). Whenever the number of syllables is different from the expected, the Praat script fixes the problem by adding an unstressed syllable where the missing unstressed syllable would be. This added syllable is assigned, by default, the same F0 values as the neighboring syllables (namely, the initial F0 value is the same as the final value of the previous syllable, the final F0 value is the same as the initial value of the following syllable, and the central F0 value is the mean between the other two). By means of this rudimentary interpolation, the problem represented by the different number of syllables is solved without altering the F0 contour[5].

Once the problem of the sentences with different syllable numbers is solved, and after computing the distance between every single utterance of a locale and all other utterances (with the same characteristics) of other locales, the script also computes the median of the distance between locales (the median was chosen instead of the mean because the distance values do not show a normal distribution). The medians of the distances among locales are then used to build a distance matrix among locales (an example is contained in Figure 6) that is saved as a txt file and, thus, can be imported into SPSS as a distance matrix.

Matriz de distancias

Caso	1: Tenerife	2: Gran_Canaria	3: La_Palma	4: Lanzarote	5: Palencia	6: Salamanca	7:Madrid	8: Barcelona	9:Lleida	10: La_Habana	11: Santiago_Cuba	12:Bullas	13: Mallorca
						distancia euclídea							
1:Tenerife	,000	,865	1,813	,687	2,587	3,789	2,734	2,049	2,699	1,681	,797	2,785	2,911
2:Gran_Canaria	,865	,000	1,775	,853	2,139	3,659	2,873	2,394	2,738	1,434	,921	2,715	2,950
3:La_Palma	1,813	1,775	,000	1,339	3,123	3,541	2,024	2,033	2,034	2,476	1,586	2,420	2,656
4:Lanzarote	,687	,853	1,339	,000	2,662	3,720	2,504	2,066	2,491	1,815	,760	2,674	2,837
5:Palencia	2,587	2,139	3,123	2,662	,000	3,827	3,784	3,449	3,530	1,916	2,503	3,398	3,562
6:Salamanca	3,789	3,659	3,541	3,720	3,827	,000	2,125	2,549	2,089	3,032	3,451	1,394	1,328
7:Madrid	2,734	2,873	2,024	2,504	3,784	2,125	,000	1,274	,888	2,783	2,410	1,273	1,094
8:Barcelona	2,049	2,394	2,033	2,066	3,449	2,549	1,274	,000	1,288	2,411	1,962	1,685	1,447
9:Lleida	2,699	2,738	2,034	2,491	3,530	2,089	,888	1,288	,000	2,832	2,493	1,317	1,140
10:La_Habana	1,681	1,434	2,476	1,815	1,916	3,032	2,783	2,411	2,832	,000	1,330	2,325	2,558
11:Santiago_Cuba	,797	,921	1,586	,760	2,503	3,451	2,410	1,962	2,493	1,330	,000	2,435	2,618
12:Bullas	2,785	2,715	2,420	2,674	3,398	1,394	1,273	1,685	1,317	2,325	2,435	,000	,807
13:Mallorca	2,911	2,950	2,656	2,837	3,562	1,328	1,094	1,447	1,140	2,558	2,618	,807	,000

Esta es una matriz de disimilaridades

**Figure 6.** Intonational distance matrix among different locales (source: Fernández Planas et al. 2015).

[5] Luckily enough, this correction turned out to be needed in a very limited percentage of sentences of the corpus (<0.5%).

Once the distance matrix has been imported into SPSS, it can undergo any kind of appropriate statistical analysis[6]. The kind of statistical analyses that can be carried out are, basically, two:

a) a cluster analysis, where the researcher (differently from what happened in Amper-StatDistances) can choose the appropriate data clustering methods (e.g. complete distance, that has proved to be a simple and effective method for intonational data within the Calcu-Dista procedure). S/he may also choose an appropriate graphic representation of the results of cluster analysis (usually a dendrogram, that is the most commonly used graph in dialectometry), and s/he may also select the appropriate indexes to validate the results of the cluster analysis;

b) a multidimensional scaling analysis (MDS), which generates a graphic output called MDS plot. For MDS analyses the researcher may also choose the appropriate indexes to validate the results, like stress and RSQ (See Matas Crespo, 2006, for details about the appropriate indexes).

Figure 7 shows one of the graphic outputs of this cluster analysis of intonational numeric distances: a dendrogram where some Spanish-speaking locales were grouped in two main clusters, basically due to the shape of their yes/no questions contours. On one side we have locales with L* H% rising yes/no questions (Central Peninsular Spanish and some dialects spoken in the Eastern Mediterranean area of the Peninsula), and on the other side we have locales with ¡H* L% falling questions (Leonese Spanish, Canarian Spanish, and Cuban Spanish) (See Fernández Planas et al., 2015, for further details about this dialectometric study and for an explanation of the characteristics of the intonation patterns of these clusters).

---

[6] Depending on the SPSS version, one might need to change some formal aspects of the matrix to carry out either the cluster analysis or the multidimensional scaling analysis. These changes are, nevertheless minor (e.g. organization of data in rows vs. columns) and do not have any effect on the procedure on the whole.
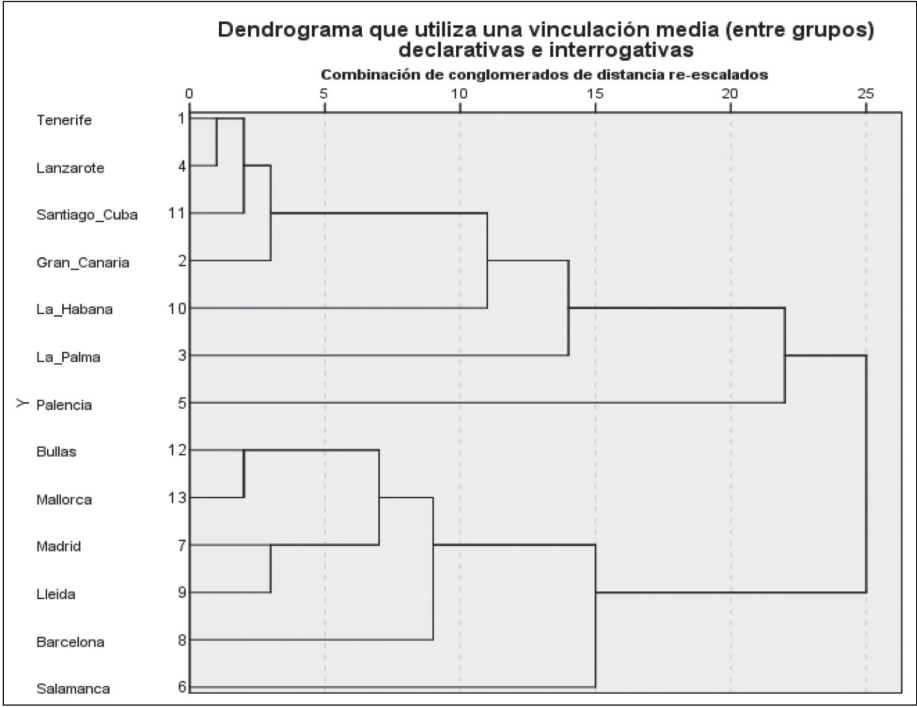
Figure 7. Dendrogram representing the result of the SPSS cluster analysis of prosodic distances computed basing on numeric acoustic intonational data (source: Fernández Planas et al., 2015).

Figure 8 shows the MDS plot obtained with the same set of data. One can observe that the clusters that one may single out are the same as those in the dendrogram in Figure 7: a group of dialects with falling questions vs. a group of geoprosodic varieties with rising questions. The two analyses, thus, provide cohesive results, as expected.
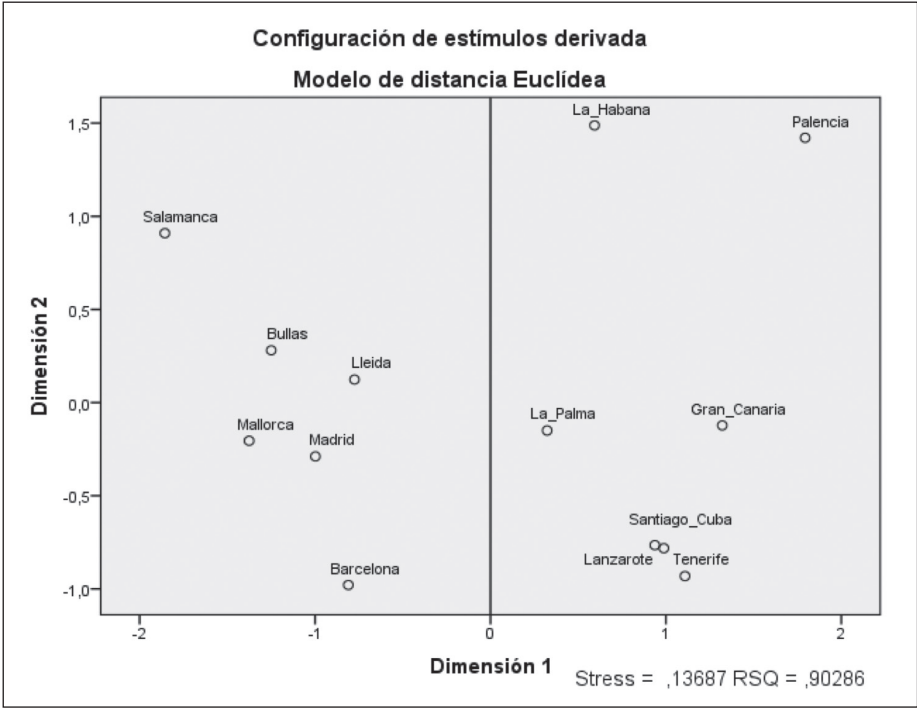
Figure 8. Plot representing the result of the SPSS multidimensional scaling of prosodic distances computed basing on numeric acoustic intonational data (source: Fernández Planas et al., 2015).

3.4. Validating Calcu-Dista by means of the string analysis

The results obtained with Calcu-Dista needed to be validated by means of an analysis carried out with a different and non-related tool. Unfortunately, using Amper-StatDistances was not a viable option due to the limitations mentioned above (See § 2.2.). For this reason, we decided to validate them by means of a non-numeric dialectometric analysis. For this purpose, we envisaged a set of operations to be carried out by means of an existing phonetic intonational transcriber called AmperEti (already mentioned in § 1.3.) and of DiaTech (the dialectometric tool cited in § 2.1.).

The first step of the analysis of string intonational data was the creation of the strings themselves. To this aim, the F0 values contained in each txt file described in § 3.2. had to be converted into a single string. This transformation can be thought of in terms of a *narrow phonetic* intonational transcription based on F0 values. Phonetic transcription of intonation and automatic transcription of

intonation are not trivial issues and they exceed by far the purposes and limits of this paper. For this reason, in this section we shall only summarize very briefly the characteristics of the automatic transcriber that was used (AmperEti) and of its labeling system. We shall then describe how the strings generated with AmperEti are used as the input for a dialectometric analysis with DiaTech. Finally, we shall briefly present the results of the analysis and compare them to those obtained with Calcu-Dista (more details are contained in Fernández Planas et al., 2015).

AmperEti (that is described in detail in Roseano and Fernández Planas, 2013; Fernández Planas et al., 2015) carried out a narrow phonetic intonational transcription based on the metrical-autosegmental assumption that relevant intonational events are anchored to prominent syllables. For this reason, it analyzed F0 values in a three-syllable window around each stressed syllable, as well as in the final unstressed stretch. Basing on F0 values, AmperEti provided a simple phonetic transcription that made use of the L and H symbols. For example, an F0 movement where F0 is low in the pretonic syllable and the tonic syllable and it rises in the postonic syllable was transcribed as LLH[7]. Each sentence in the corpus, thus, ended up being represented by a string formed by twelve characters (three for each stressed syllable and three for the final unstressed stretch), like, for example LLHLLHHLLLLL, which is the narrow phonetic transcription of the intonation of a Friulian broad focus statement whose phonological representation is L*+H L*+H H+L* L% according to Fri_ToBI (Roseano et al., 2015).

The database containing all numeric strings of that kind was then imported in DiaTech, which was used to dialectometrize the intonational data (using Levenshtein's distance to compute the distances between strings and Ward's method to build the clusters). The results of this intonational dialectometry of string data coincided with those of the numeric intonational dialectometric analysis carried out with Calcu-Dista. In fact, as one can see in Figure 9, the two clusters are the same (locales with L* H% rising yes/no questions in Group 2 vs. locales with ¡H* L% falling questions in Group 1).

The fact that the results obtained with a well-known dialectometrical tool like DiaTech coincide with those obtained with the new tool presented in this paper (Calcu-Dista) confirms that the analyses carried out by the latter are accurate.

---

[7] Labels need to have the same length (three characters) because the metrics used in dialectometric analyses (like Levenshtein's distance) require that the strings to be compared have a similar length. If the strings have a number of characters that is too different, the result of the dialectometric analysis is unreliable. If we used labels with different length (like H, LH, LHL) for pitch accents, no string dialectometric analysis would have been possible. See Fernández Planas et al. (2015) for further details.
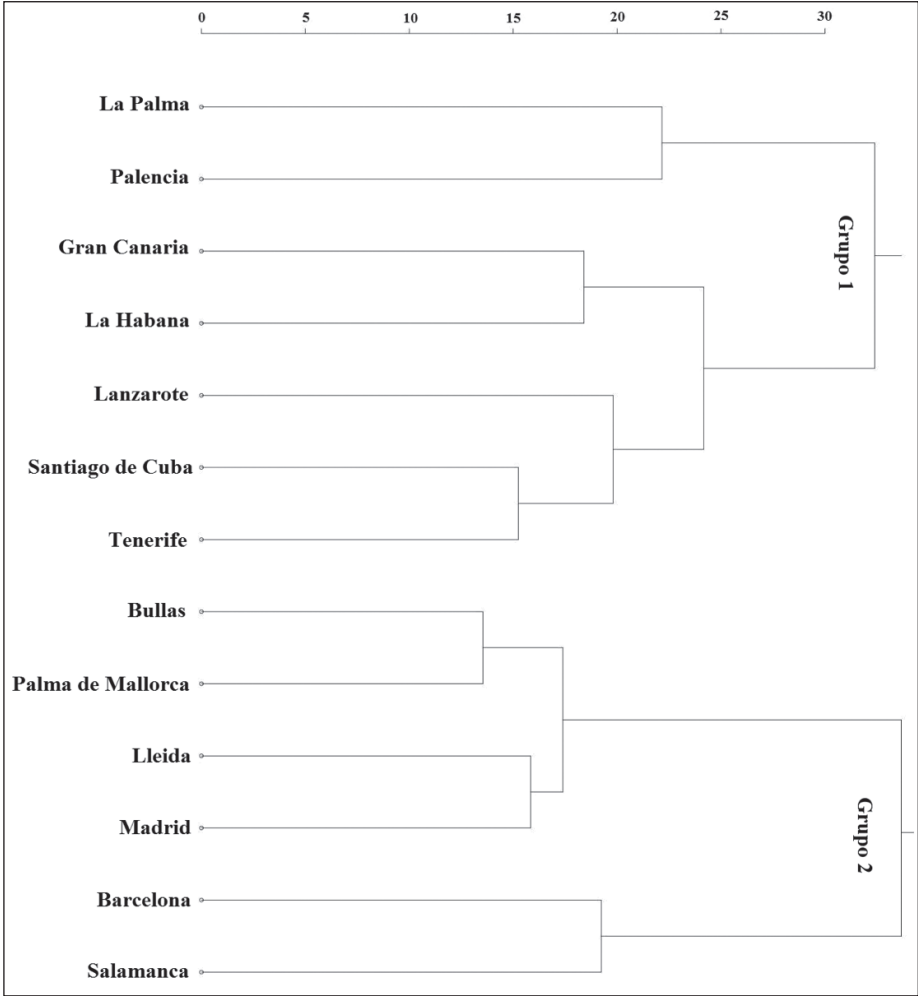
Figure 9. Dendrogram representing the result of the DiaTech cluster analysis of string intonational data (source: Fernández Planas et al., 2015).

4. CONCLUSIONS

Calcu-Dista has proved to be a reliable procedure for dialectometrical analysis of numeric intonational data. The main advantage in comparison with the previously existing tool, Amper-StatDistances, is that it makes explicit some statistical aspects that Amper-StatDistances did not. Calcu-Dista, in fact, provides the distance matrix, it allows choosing among different methods to create dendrograms (and MDS plots), it allows cluster validation, and it provides the Stress and RSQ

values for MDS plots.

This notwithstanding, Calcu-Dista shows a shortcoming that will have to be dealt with in the future: the Praat script that creates the database and calculates prosodic distances requires some time to process all data (from several hours to a couple of days, depending on the quantity of data), due also to the technical limitations of the RAM of most PCs. Hopefully, the quick advances in PC technologies will help overcome this inconvenient.

Beyond the considerations about the advantages and limitations of Calcu-Dista, a more general final remark can be made. If we look back at the last two or three decades, we can say that studies about the diatopic intonational variation in Romance languages have increased very quickly. This has led to the birth of some multilingual databases specifically dedicated to Romance prosody (AMPER, IARI). Although these databases are still under construction, their size has increased dramatically over the last few years. This implies that researchers will need more and more powerful and flexible tools to manage the information overload contained in such databases, and dialectometric tools like Calcu-Dista might indeed help greatly.

REFERENCES

Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry Thompson, and Regina Weinert. 1991. The HCRC Map Task Corpus. *Language and Speech* 34(4). 351-366.

Apopei, Vasile, Doina Jitcă and Adrian Turculeţ. 2006. Intonational structures in Romanian yes-no questions. *Computer Science Journal of Moldova* 14.1. 113-37.

Aurrekoetxea, Gotzon, Karmele Fernández-Aguirre, Jesús Rubio, Borja Ruiz and Jon Sánchez. 2013. DiaTech: A new tool for dialectology. *Literary and Linguistic Computing* 28(1). 23-30.

Besana, Sveva. 1999. *Towards an analysis of Turinese Italian intonation and theoretical implications for intonational phonology*. Cambridge, MA: MIT dissertation.

Billmyer, Kristine and Manka Varghese. 2000. Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests. *Applied Linguistics* 21(4). 517-552.

Blum-Kulka, Shoshana, Juliane House and Gabriele Kasper. 1989. Investigating cross-cultural pragmatics: An introductory overview. In Shoshana Blum-Kulka, Juliane House and Gabriele Kasper (eds.). *Cross-cultural pragmatics: Requests and apologies*, 1-34. Norwood, NJ: Ablex.

Borràs-Comes, Joan, Maria del Mar Vanrell and Pilar Prieto. 2014. The role of

pitch range in establishing intonational contrasts. *Journal of the International Phonetic Association* 44(1). 1-20.

Boula de Mareüil, Philippe and Béatrice Akissi Boutin. 2011. Identification et évaluation perceptives d'accents ouest-africains en français. *Journal of French Language Studies* 21(3). 361-379.

Boula de Mareüil, Philippe, Albert Rilliard, Fanny Ivent and Varvara Kozhevina. 2016. Une étude prosodique comparative des questions en français en contact avec l'occitan et le catalan. *Langages* (forthcoming).

Caputo, Maria Rosaria. 1996. *Le domande in un corpus di italiano parlato. Analisi prosodica e pragmatica*. Napoli: Università degli Studi di Napoli "Federico II" dissertation.

Carrera Sabaté, Josefina, Ana Ma. Fernández Planas and Eugenio Martínez Celdrán. 2010. Declaratives i interrogatives absolutes del català en el marc del projecte internacional Atles Multimèdia de Prosòdia de l'Espai Romànic, *Caplletra* 49. 133-167.

Castro Moutinho, Lurdes de (ed.). 2008-2013. *AMPER-POR. Atlas Multimédia Prosódico do Espaço Românico - Língua Portuguesa*. http://pfonetica.web.ua.pt/AMPER-POR.htm. (29 June, 2016.)

Castro Moutinho, Lurdes de, Rosa Lídia Coimbra, Albert Rilliard and Antonio Romano. 2011. Mesure de la variation prosodique diatopique en portuguais européen. *Estudios de Fonética Experimental* 20. 33-55.

Contini, Michele. 1992. Vers une géoprosodie romane. *Nazioarteko Dialektologia Biltzarra Agiriak*, 83-109. Bilbao: Real Academia de la Lengua Vasca.

Crespo-Sendra, Verònica. 2011. *Aspectes de l'entonació del valencià*. Barcelona: Universitat Pompeu Fabra dissertation.

Cruz, Marisa and Frota, Sónia. 2011. Prosódia dos tipos frásicos em variedades do português europeu: Produção e percepção. In Ángela Costa, Isabel Falé and Plinio Barbosa (eds.), *Textos seleccionados: XXVI Encontro nacional da Associação portuguesa de linguística*, 208-225. Porto: Associação Portuguesa de Linguística.

D'Imperio, Mariapaola. 1997. Breadth of focus, modality and prominence in Neapolitan Italian. *OSU Working Papers in Linguistics* 50. 19-39.

D'Imperio, Mariapaola. 2000. *The role of perception in defining tonal targets and their alignment*. Columbus: Ohio State University dissertation.

Delais-Roussarie, Elisabeth, Brechtje Post, Mathieu Avanzi, Carolin Buthke, Albert Di Cristo, Ingo Feldhausen, Sun-Ah Jun, Philippe Martin, Trudel Meisenburg, Annie Rialland, Rafèu Sichel-Bazin and Hi-Yon Yoo. 2015. Intonational phonology of French: Developing a ToBI system for French. In Sónia Frota and Pilar Prieto (eds.), *Intonational variation in Romance*, 63-100. Oxford: Oxford University Press.

Elvira-García, Wendy. 2014. *Calcu-Dista scripts package. Praat script*. Available on:

http://stel.ub.edu/labfon/en/praat-scripts (28 February, 2015.)

Endo, Reiko and Pier Marco Bertinetto. 1997. Aspetti dell'intonazione in alcune varietà dell'italiano. In Francesco Cutugno (ed.). *Atti delle VII Giornate di studio del Gruppo di fonetica sperimentale (Naples,1996)*, 27-49. Roma: Esagrafica.

Félix-Brasdefer, J. César. 2010. Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In Alicia Martínez-Flor and Esther Usó-Juan (eds.), *Speech act performance: Theoretical, empirical, and methodological issues*, 41-56. Amsterdam and Philadelphia: John Benjamins.

Fernández Planas, Ana Ma., Paolo Roseano, Eugenio Martínez Celdrán and Lourde Romera Barrios. 2011. Aproximación al análisis dialectométrico de la entonación en algunos puntos del dominio lingüístico catalán. *Estudios de Fonética Experimental* 20. 141-178.

Fernández Planas, Ana Ma., Paolo Roseano, Wendy Elvira-García, Josefina Carrera Sabaté and Domingo Román Montes de Oca. 2017 (forthcoming). From a perceptual point of view, is there continuity between languages in contact? In Yolanda Congosto and Laura Morgenthaler (eds.), *Prosody and language contact in the Romance languages*. Amsterdam and Philadelphia: John Benjamins.

Fernández Planas, Ana Ma., Josefa Dorta, Paolo Roseano, Xaxiraxi Díaz, Wendy Elvira-García, José Antonio Martín Gómez and Eugenio Martínez Celdrán. 2015. Distancia y proximidad prosódica entre algunas variedades del español: Un estudio dialectométrico a partir de datos acústicos. *Revista de Lingüística Teórica y Aplicada* 53(2). 13-45.

Fernández Rei, Elisa (ed.). 2016. *AMPER-Galicia, Atlas Multimedia Prosódico do Espazo Románico*. http://ilg.usc.es/amper/. (29 June, 2016.)

Font-Rotchés, Dolors. 2007. *L'entonació del català*. Barcelona: Publicacions de l'Abadia de Montserrat.

Frota, Sónia and Marisa Cruz (eds.). 2012-2015. *Interactive Atlas of the Prosody of Portuguese*. http://labfon.letras.ulisboa.pt/InAPoP/. (29 June, 2016.)

Frota, Sónia and Pilar Prieto (eds.). 2015. *Intonation in Romance*. Oxford: Oxford University Press.

Frota, Sónia, Marisa Cruz, Flaviane Svartman, Gisela Collischonn, Aline Fonseca, Carolina Serra, Pedro Oliveira and Marina Vigario. 2015. Intonational variation in Portuguese: European and Brazilian varieties. In Sónia Frota and Pilar Prieto (eds.), *Intonational variation in Romance*, 235-283. Oxford: Oxford University Press.

Gili Fivela, Barbara. 2009. From production to perception and back: An analysis of two pitch accents. In Susanne Fuchs, Hélène Lœvenbruck, Daniel Pape and Pascal Perrier (eds.), *Some aspects of speech and the brain*, 363-405. Frankfurt am Main: Peter Lang.

Gili Fivela, Barbara. 2013. Varietà di italiano e differenze nella percezione degli accenti intonativi. In Antonio Romano and Mario Spedicato (eds.), *Sub voce*

sallentinitas: Studi in onore di G.B. Mancarella, 289-303. Lecce: Grifo.

Gili-Fivela, Barbara, Cinzia Avesani, Marco Barone, Giuliano Bocci, Claudia Crocco, Mariapaola D'Imperio, Rosa Giordano, Giovanna Marotta, Michelina Savino and Patrizia Sorianello. 2015. Intonational phonology of the regional varieties of Italian. In Sónia Frota and Pilar Prieto (eds.), *Intonational variation in Romance*, 140-197. Oxford: Oxford University Press.

Goebl, Hans. 1981. Eléments d'analyse dialectométrique (avec application à l'AIS). *Revue de Linguistique Romane* 45. 349-420.

Grice, Martine. 1995. *The intonation of interrogation in Palermo Italian: Implications for intonation theory*. Tübingen: Niemeyer.

Haimerl, Edgar. 2006. Database design and technical solutions for the management, calculation, and visualization of dialect mass data. *Literary and Linguist Computing* 21(4). 437-444.

Henriksen, Nicholas. 2010. *Question intonation in Manchego Peninsular Spanish*. Bloomington: Indiana University dissertation.

Hermes, Dik J. 1998a. Auditory and visual similarity of pitch contours. *Journal of Speech, Language and Hearing Research* 41. 63-72.

Hermes, Dik J. 1998b. Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language and Hearing Research* 41. 73-82.

Hualde, José Ignacio and Pilar Prieto. 2015. Intonational variation in Spanish: European and American varieties. In Sónia Frota and Pilar Prieto (eds.), *Intonational variation in Romance*, 350-391. Oxford: Oxford University Press.

Jitcǎ, Doina, Vasile Apopei, Otilia Pǎduraru and Samuil Marusca. 2015. Transcription of Romanian intonation. In Sónia Frota and Pilar Prieto (eds.), *Intonational variation in Romance*, 284-316. Oxford: Oxford University Press.

López Bobo, María Jesús, Carmen Muñiz Cachón, Liliana Díaz Gómez, Norberto Corral Blanco, Daniel Brezmes Alonso and Mercedes Alvarellos Pedrero. 2007. Análisis y representación de la entonación. Replanteamiento metodológico en el marco del proyecto AMPER. In Josefa Dorta (ed.), *La prosodia en el* ámbito *lingüístico románico*, 17-34. Santa Cruz de Tenerife: La Página.

Marotta, Giovanna and Patrizia Sorianello. 1999. Question intonation in Sienese Italian. *14th International Congress of Phonetic Sciences (ICPhS'99)*, 1161-1164. San Francisco, CA.

Martínez Celdrán, Eugenio and Ana Ma. Fernández Planas (eds.). 2003-2016a, *AMPER-CAT. Atles Multimèdia de la Prosòdia de l'Espai Romànic*. http://stel.ub.edu/labfon/amper/index_ampercat_cat.html. (29 June, 2016.)

Martínez Celdrán, Eugenio and Ana Ma. Fernández Planas (eds.). 2003-2016b, *AMPER España e Iberoamérica. Atlas multimedia de la prosodia del espacio románico*. http://stel.ub.edu/labfon/amper/cast/index.html. (29 June, 2016.)

Martínez Celdrán, Eugenio, Ana Ma. Fernández Planas, Josefa Dorta Luis and Elisa Fernández Rei. 2006. La intercomprensió d'interrogatives absolutes amb

tonemes descendents del català, de l'espanyol de Canàries i del gallec. *Estudis Romànics* 28. 7-28.

Matas Crespo, Josep. 2006. *La técnica del Escalamiento Multidimensional en el vocalismo: Un análisis comparativo*. Barcelona: Universitat de Barcelona dissertation.

Nerbonne, John, Rinke Colen, Charlotte Gooskens, Peter Kleiweg and Therese Leinonen. 2011. Gabmap: A web application for dialectology. *Dialectologia* Special issue II. 65-89.

Prieto, Pilar and Paolo Roseano (eds.). 2009-2013. *Atlas interactivo de la entonación del español*. http://prosodia.upf.edu/atlasentonacion/. (29 June, 2016.)

Prieto, Pilar and Rafèu Sichel-Bazin (eds.). 2014. *Atlàs interactiu de l'intonacion de l'occitan*. http://prosodia.upf.edu/atlasintonacion/. (29 June, 2016.)

Prieto, Pilar and Teresa Cabré (eds.). 2007-2012. *Atles interactiu de l'entonació del català*. http://prosodia.upf.edu/atlesentonacio/. (29 June, 2016.)

Prieto, Pilar, Joan Borràs-Comes and Paolo Roseano (eds.). 2010-2014. *Interactive Atlas of Romance Intonation*. http://prosodia.upf.edu/iari/. (29 June, 2016.)

Prieto, Pilar, Joan Borràs-Comes, Teresa Cabré, Verònica Crespo-Sendra, Ignasi Mascaró, Paolo Roseano, Rafèu Sichel-Bazin and Maria del Mar Vanrell. 2015. Intonational phonology of Catalan and its dialectal varieties. In Sónia Frota and Pilar Prieto (eds.), *Intonational variation in Romance*, 9-62. Oxford: Oxford University Press.

Rilliard, Albert and Jean-Pierre Lai. 2008. Outils pour le calcul et la comparaison prosodique dans le cadre du projet AMPER. L'exemple des varieties Occitane et Sarde. In Adrian Turculeţ (ed.), *La variation diatopique de l'intonation dans le domaine roumain et roman*, 217-229. Iaşi: Editura Universităţii Al. I. Cuza.

Romano, Antonio, Jean-Pierre Lai and Stephanie Roullet. 2005. La méthodologie AMPER. *Géolinguistique* Hors Serie 3. 1-5.

Romano, Antonio, Michel Contini, Jean-Pierre Lai and Albert Rilliard. 2011. Distancias prosódicas entre variedades románicas en el marco del proyecto AMPER. *Revista Internacional de Lingüística Iberoamericana* 9(1). 13-25.

Roseano, Paolo. 2012. *La prosòdia del friülà en el marc de l'Atles Multimèdia de Prosòdia de l'Espai Romànic*. Barcelona: Universitat de Barcelona dissertation.

Roseano, Paolo and Ana Ma. Fernández Planas (eds). 2009-2013. *AMPER-FRIÛL. Atlant multimediâl de prosodie des varietâts romanichis*. http://stel.ub.edu/labfon/amper/friul/index.html/. (29 June, 2016.)

Roseano, Paolo and Ana Ma. Fernández Planas. 2013. Transcripció fonètica i fonológica de l'entonació: Una proposta d'etiquetatge automàtic. *Estudios de Fonética Experimental* 22. 275-332.

Roseano Paolo, Maria del Mar Vanrell and Pilar Prieto. 2015. Intonational phonology of Friulian and its dialects. In Sónia Frota and Pilar Prieto (eds.), *Intonational variation in Romance*, 101-139. Oxford: Oxford University Press.

Sardelli, Elena. 1998. *Verso una grammatica dell'intonazione. Uno studio sperimentale dei contorni melodici di due varietà toscane (Siena-Pisa)*. Siena: Università degli Studi di Siena dissertation.

Séguy, Jean. 1971. La relation entre la distance spatiale et la distance lexical. *Revue de Linguistique Romane* 35. 335-357.

Sichel-Bazin, Rafèu, Trudel Meisenburg and Pilar Prieto. 2015. Intonational phonology of Occitan: Towards a prosodic transcription system. In Sónia Frota and Pilar Prieto (eds.), *Intonational variation in Romance*, 198-234. Oxford: Oxford University Press.

Simon, Anne Catherine, Philippe Hambye, Alice Bardiaux and Philippe Boula de Mareüil. 2012. Caractéristiques des accents régionaux en français: Que nous apprennent les approches perceptives? In Anne Catherine Simon (ed.), *La variation prosodique régionale en français*, 27-40. Bruxelles: De Boeck.

Sorianello, Patrizia. 1996. Intonazione e sintassi: Divergenze e convergenze sperimentali. *Quaderni del Dipartimento di Linguistica dell'Università degli Studi della Calabria, Serie Linguistica* 6. 35-51.

Turculeţ, Adrian (ed.). 2010-2013. *AMPROM. Atlas multimedia prozodic român*. http://amprom.uaic.ro/indexamprom.php. (29 June, 2016.)

Vanrell, Maria del Mar. 2006. A tonal scaling contrast in Majorcan Catalan interrogatives. *Journal of Portuguese Linguistics* 5/6. 147-178.

Vanrell, Maria del Mar, Antonio Stella, Barbara Gili-Fivela and Pilar Prieto. 2013. Prosodic manifestations of the Effort Code in Catalan, Italian and Spanish contrastive focus. *Journal of the International Phonetic Association* 43(2). 195-220.

Vanrell, Maria del Mar, Francesc Ballone, Carlo Schirru and Pilar Prieto. 2015. Sardinian intonational phonology: Logudorese and Campidanese varieties. In Sónia Frota and Pilar Prieto (eds.), *Intonational variation in Romance*, 317-349. Oxford: Oxford University Press.

Woehrling, Cécile and Philippe Boula de Mareüil. 2006. Identification d'accents régionaux en français: Perception et analyse. *Revue PArole* 37. 25-65.