CAPÍTULO 2

GEOPROSODY

QUANTITATIVE APPROACHES OF PROSODIC VARIATION ACROSS DIALECTS

Albert Rilliard (LIMSI, CNRS, Université Paris Saclay; UFRJ)

ABSTRACT

This chapter aims at summarizing approaches to dialectometry proposed in the literature since Seguy and Goebl pioneering works, as well as on the basis of metrics used in sociolinguistics. Based on the literature review, it discusses these different metrics under a still under studied angle: their application to prosodic variation, especially for studying dialectological changes. Possible objective measures are regrouped in two broad categories, depending on whether they uses descriptions based on melodic contours or on feature sets. Proposals for use of such metrics are made, in relation to existing phonological and phonetic approaches to prosody description. Examples of application of these metrics are then given on a small subset, and their output discussed and reconciled. It shows several approaches are available, that basically gives convergent objective information, and can be applied to studies based on different theoretical backgrounds.

INTRODUCTION

Since the pioneering work of Séguy (1971), the idea of quantitative measurements of linguistics differences across dialects has flourished. Séguy was working on lexical differences, and pleads for a computational approach due to the complexity of the task. Building on this founding stone, Goebl (e.g. 1982, 2006, 2010) laid the bases of computational dialectometry, introducing notably measures of divergences between linguistic varieties based on similarity and dissimilarity measurements (a similarity being easily converted into a dissimilarity, we'll use mostly the term "divergence"), their grouping into "distance matrices" the content of which is then represented on choropleth maps (or variations of it representing for example boundaries, see INOUE 1996). Representations of language variation may be made in relation to a reference point, so to display the changes occurring in language from a given point of view. Maps may also present main zones characterized by stable feature sets, by enhancing the transition regions between main dialectal varieties (HEERINGA; NERBONNE 2013). These dialectometric approaches target various levels of linguistic description, including its lexical, diachronic, orthographic, phonetic, phonologic, or morpho-syntactic aspects (see GOEBL, 2003; HEERINGA, GOOSKENS, 2003; PEIRSMAN et al. 2010 for different examples). Changes in the prosodic structures of dialects are seldom addressed in these dialectometric approaches. Even a reference work such as Chambers & Trudgill's "dialectology" (2004), or the review article by Wieling & Nerbonne (2015) on "advances in dialectometry", does not mention the terms "prosody" and "intonation". Prosodic changes were used in several works, but mostly within perceptual approaches (GOOSKENS 1997; MASE 1999; KUIPER 1999), a fact that underlines its importance in the reception of language variation, and the significance it may have to take it into account for dialectometry. An early dialectometric work on prosody was proposed by Gooskens & Heeringa (2006), who took into account stress and tonemes to calculate a "prosodic distance".

Changes within a language are not only linked to the geographic distance across population speaking dialectal varieties. Factors such as education, exposure to audio-visual media, self-affirmation and cultural empowerment, may have important effects on language use (CHAMBERS & TRUDGILL 2004), without being necessarily best analysed on the basis of geographic spread. Dialects may also be used diastratically by competent speakers, according to the communication situations – a view that is coherent with Séguy's claim that dialects serve both to communicate and to differentiate the speaker from other social groups.

Whatever the complexity of the various sources of variation, similar techniques designed to quantify linguistic differences may equally be applied; this shall not be the case for representation methods – but several proposals exist: see for example Inoue's glottograms (2016) used to represent age and geographic spread of (lexical) changes in language use.

The central notions of dialectometry are: (i) measures of differences or similarities across language varieties, for one or several linguistic functions; (ii) agglomeration and clustering algorithms to organize large datasets according to their proximity/differences; (iii) tools for the representation of measured variations, across time, space and social groups. From these methods, the first ones are the most dependent on the types of data one wants to analyse, the two others being mostly similar whatever the linguistic analysis and mostly dependant on datasets and research aims.

Measuring of differences between linguistic features has often been based on several kinds of string metrics (typically Hamming or Levenshtein distances: SEGUY 1971; GOOSKENS & HEERINGA 2004). It basically consists in counting the number of differences between two series of labels representing features (letters, phonemes, phonetic features, etc.) that may change along the dialectal continuum. Other measures, notably inspired by sociolinguistics methodology, are based on frequency differences between kinds: the relative frequencies of studied items (lexical, grammatical, phonetic, etc.) may be used as a metric to evaluate differences in language usage (in these cases, the frequency of a phenomenon is used rather a divergence based on the comparison between two versions of the phenomenon). Such approach may compare different varieties for their relative use of a given phenomenon (e.g. WEINER et al. 1983), but more recent method may also apply the approach mixing the social and geographic dimensions (WIELING et al. 2014). The approach proposed by Speelman & Geeraerts (2008) groups lexical variants of the same concept under one "profile", so to take into account a finer description of lexical use, which is not necessarily binary, as it may be seen in atlases or other approaches; profiles are then compared across language varieties or across time, based on several metrics.

The next step, once speakers' productions have been compared for their varying linguistic characteristics, consists in agglomerating the divergences into matrices representing the comparisons between all pairs of the considered entities (speaker, village, time, etc.) in the dataset. These matrices (often referred to as "distance matrices") are then subjected to a multidimensional analysis, typically a multidimensional scaling (MDS; see Baayen, 2008, for details): the output of these methods will give a reduced set of abstract dimensions that best represent the variation between the individuals represented in the input matrix. In the case of MDS, the solution is forced on two dimensions, a practice that is particularly interesting to study variations that are supposed to take place along geographic space. The output of the multidimensional analysis allows extrapolating distances (i.e. linguistics distances) between the compared items (ROMNEY et al. 2000). Two points (i.e. the items of the input) that are the closest according to this measurement may be supposed to share most of their linguistic competences; two points that are far removed shall have marked distinctive linguistic practices. It is possible to regroup these points according to these distances, using whatever clustering method best fit the purpose of the study. Hierarchical clustering approaches are often used (GOEBL 2003, HEERINGA & GOOSKENS 2003), and produce dendrogram representations of the distances between items, allowing a simple comparison between groups.

Then, the divergences between items or groups may be represented so to allow meaningful analysis of the results. Representations in the case of dialectology are often based on maps. Goebl (2006) heralded the use of similarity maps: first Delaunay-Voronoi polygonal structures allow partitioning map with spread inquiry points into a continuum, each polygon can then be filled according to the distance the related point has with a reference point; other maps may enhance the boundaries between inquiry points (so to mark strong dialectal variations), or the link between two point (so to marks zones of similarity). Such cartography has been evolved with the progress of computer graphic technology (Wieling & Nerbonne 2015).

The aim of this chapter is to discuss methods suited to measure prosodic changes across linguistic varieties. The suitability of the different measures to several types of analyses, their robustness to measurement noise, their applicability to different types of datasets will be discussed. The discussion as well as the works presented will have a focus on Romance languages, especially on the application part; specific aspects linked to such kind of prosodic measurements for tonal languages, particularly, will not be addressed. Existing solutions will be presented and challenges for better approaches and representations of such measurements will be presented.

DESCRIPTION OF PROSODIC CHANGES

As prosody may be viewed as the domain of gradual variation, as opposed to phonemic categories, it has long been approached using different means to control its continual changes. One of the most fruitful such approach may be found in the theoretical and practical works made at IPO that culminates in t'Hart et al. book (1990). The process of stylisation that is advocated by these researchers, simplifying the fundamental frequency (F_0) curve as a series of straight lines so to remove microprosodic and other involuntary changes from pitch contours, had a main influence on prosodic description works. The first step is linked with the calculus of a close-copy stylization that carries phonetic variations; the second one consists in creating an equivalent copy, the changes of which have phonological values. From these simplified curves (obtained through IPO's close copy stylization or similar processes, e.g. the MOMEL/INTSINT one, see HIRST et al. 2000), it is possible to extract two types of data: either parameters defining continuous contours that span the segments of interest (syllable, prosodic word…), or discrete feature sets. Both approaches are primarily descriptive, but may be used to extract quantitative descriptions and comparison between prosodic performances.

Description based on contours

Approaches of intonation have tried to describe the relations between F_0 changes and time in terms of patterns of smooth curves; in many cases these approaches were linked to signal processing aims. The idea to use functions so to describe the non-linear changes of intonation movements along time was tested by Levitt & Rabiner (1971). They proposed to use orthogonal polynomial bases to that aim, and introduced two levels of analysis (based on short-term windows and on groups of continuously voiced syllables). Their method lacks an analysis of utterance-level contours, and also uses a strong assumption on time normalization (that fit their particular case). Olive (1975) analysed sentence-level intonation contours, with systematic variation of the utterance syntactic structure and length. In that respect, his work is similar to the constrained prosodic data gathered within the dialectological AMPER project (Contini et al. 2002). But Olive averages the F_0 contours at the sentence level for different sentences with different phonemic content so to remove microprosodic effects. The averaged contours are then fitted by fourth-order orthogonal polynomials, and the models used for speech synthesis. Orthogonal polynomials were used for several studies targeting linguistic description; most of them target short linguistic elements that can be though of as comparable across speakers and phrases, such as the syllable for tones (ANDRUSKI & COSTELLO 2004), the last prominence for sentence-final contours in English (LAI 2014). The coefficients of the polynomial bases are used to describe categories of intonation shapes and groups productions into shape-coherent clusters. Similarly, Hadjipantelis et al. (2012) use a functional data analysis to fit smooth functions to the syllabic tones of Mandarin, and extract the most relevant shapes associated to each one.

Such approaches aim at summarizing a complex shape of time-varying F_0 contour by a reduced number of parameters (in an information-theoretic approach); the Fujisaki Model has a similar background: fit a smooth curve, derived from a reduced set of parameters, to F_0 variations (FUJISAKI 1983, 1988, 2004). Fujisaki approach is notably different from previous ones because the parameters are grounded in the physiological process of voice production, and thus each parameter has a specific interpretation (e.g. baseline F_0 value, phrase or accent commands; see also Kochanski & Shih 2003 for an approach with similar physiological motivations). The Fujisaki model was implemented for production purposes; Mixdorff's implementation (2000) allows an automatic analysis of measured F_0 variations in recorded datasets, so to automatically estimate the model's parameters from the raw measurements. It is then possible to use Fujisaki's model in a descriptive way, founding descriptions on sets of accent commands extracted from the analysis, and relating them to the production by speakers of voluntary melodic variations (in that respect, it does tie with IPO's philosophy of analysis, targeting meaningful intonation movements). Mixdorff's (2000) process starts by fitting the F_0 values with a continuous smooth curve, based on MOMEL's spline fitting of raw F_0 measures (HIRST & ESPESSER 1993). The MOMEL algorithm is similar to the functional fitting presented in the preceding paragraph, but it differs at least for the use of quadratic splines as fitting functions (MOMEL serves as a basis to calculate INTSINT's phonological features, cf. infra). From this smooth curve, the Fujisaki model estimates, inter alia, a set of accent commands (with their position in time, duration, and amplitude) that may be used as variables to describe the F_0 contours, or in a similar way as feature sets (e.g. number of accented syllables, localization regarding the tonic syllable, etc.), so to describe prosodic variations – as Mixdorff & Pfitzinger (2005) have demonstrated.

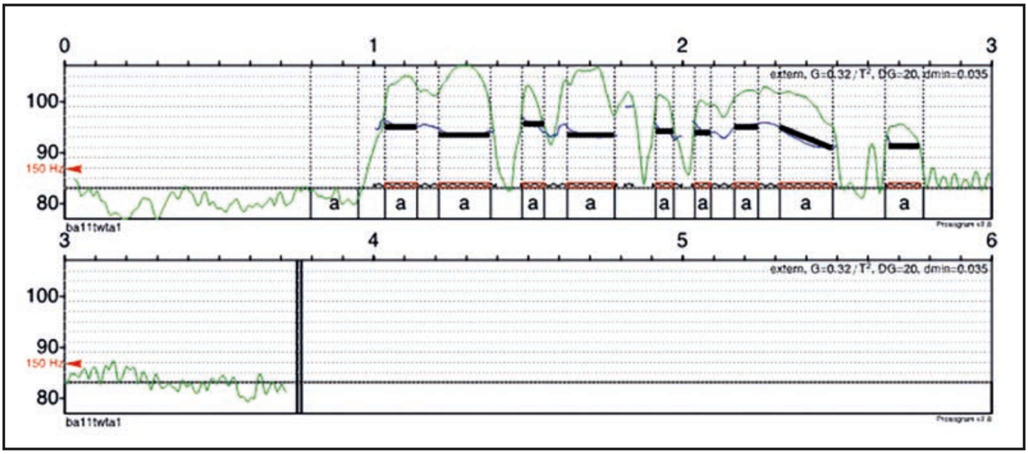
Description based on feature sets

Describing prosodic changes in terms of features is typical of linguistic approaches to prosody – and follows the classical principle of feature sets for phonemic descriptions, as opposed to the mainly signal processing approach in the case of smooth contours. An early proposition for using prosodic features sets was made by Martin (1975, 1982), and improved in Martin (1987). He proposes the use of feature matrices to describe prosodic markers. Another infamous approach to prosodic description, the ToBI system (SILVERMAN et al. 1992), and notably its application to Portuguese and Spanish varieties (FROTA & MORAES 2016; HUALDE & PRIETO 2015), may be seen as a series of features transcribing relevant phonological structures of prosody, or as categories the frequency of which vary across dialects (similarly to lexical approaches described in the introduction; CRUZ et al. 2017). In the same line, other systems of phonological transcription of prosody may serve the same purpose (e.g. INTSINT as in HIRST et al. 2000; Polytonia by MERTENS 2014).

Specifically targeting prosodic variation for dialectometric purposes, Contini & Profili (1989) build on Martin and IPO proposals to set up a 13-feature set describing variations of intonation and duration (see also Contini 1992 for a programmatic view). The features are established after an IPO-style stylisation process that allows straight lines representation of F_0. The features proposed for F_0 changes on each syllable are ±rising, ±falling, ±steep, ±gentle, ±wide, ±narrow, plus two features indicating if the syllable is ±above or ±below the speaker's mean F_0. Five other features are attributed to one syllable, relatively to either the prosodic groups or the complete sentence: four features characterize the highest (vs. lowest) F_0 point of the group or of the sentence, the strongest lengthening of the group or of the sentence; the last feature characterizes the ±falling global F_0 contour of the sentence. These binary features are applied to each vowel of sentences; that way, it is possible to create matrices describing a sentence's prosodic patterns, with features on lines, and vowels on columns. A column that gathers positive evaluations on most features is described as a "hot spot" (a "*point chaud*", in Contini's terms) of prosodic variation for this sentence.

Other approaches that extract sets of patterns or features from the F_0 contours may be used to similar goals: the model of tonal perception proposed by d'Alessandro & Mertens (1995), and its implementation in Praat (BOERSMA & WEENINK 2018) as the Prosogram (MERTENS 2004) allow to automatically extract information on the pitch patterns composing sentences. Figure 1 shows an example of a sentence with F_0 contour on vowels being stylized by the Prosogram (the example comes from the work of Rebollo Couto et al., 2017, that will be presented in more details latter). In that case, only the penultimate syllable (here the final tonic syllable) is stylized with a dynamic tone, all the other having contours simplified as a flat tone as their F_0 variation is below the default threshold of perception for glissando used in the model of tonal perception (d'Alessandro & Mertens 1995). Such simplification of the raw F_0 estimation (the blue line) allows straightforwardly estimating features such as the F_0 movement on each syllable (flat, raising, etc.), the amplitude of this movement, etc., so to construct features matrices similar to Contini & Profili (1989) proposition. The table presented at the bottom part of figure 1 is such a feature matrix. The eleven features are derived from Contini & Profili (1989) ones. The first three features describe the F_0 movement on the vowel, if its flat or dynamic (raise, fall, raise-fall, etc.), the second the speed of these movements, and the third its amplitude; these features or not binary ones, contrary to Contini & Profili. Features 4 to 10 describe the vowel in relation with others: if its F_0 is above or below the mean; if it is the highest of the prosodic word or of the utterance, if the syllable is the longest of the prosodic word or of the utterance, etc. The last one describes the direction of the sentence's slope. These features are presented in Rilliard (2014: p. 45ff); in figure 1, each column corresponds to (and is vertically aligned with) one vowel, and each line contains the description of one feature.

Figure 1: ***Top*** Prosogram stylization (thick black straight lines)) of the F_0 (blue line) of each of the 10 syllables of an assertive sentence from a female speaker of Rio de Janeiro (REBOLLO COUTO et al. 2017); each vowel is symbolized by an "a", the first one not being produced in that case. The green line represents intensity. ***Bottom*** Feature matrix constructed from the Prosogram's stylisation (see text).



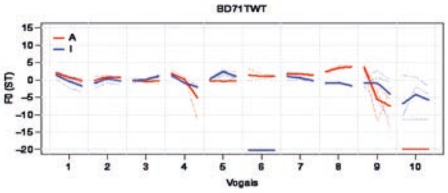
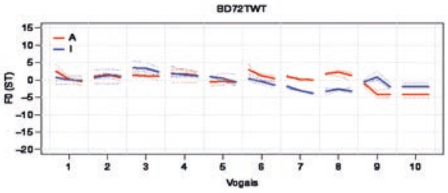
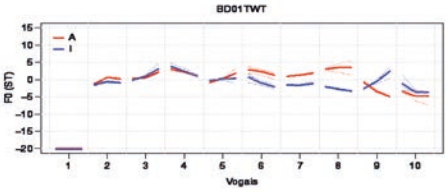
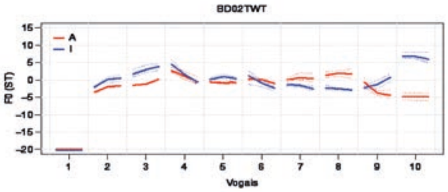
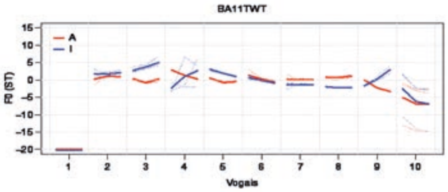
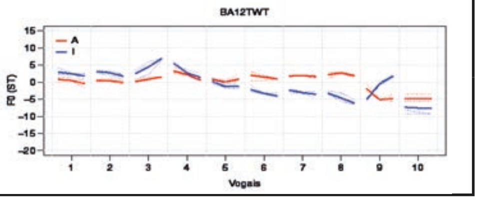
Features	1	2	3	4	5	6	7	8	9	10
F_0 movement shape	0	0	0	0	0	0	0	0	-	0
F_0 movement speed	0	0	0	0	0	0	0	0	+	0
F_0 movement range	0	0	0	0	0	0	0	0	+	0
F_0 position ~ mean	0	0	0	0	0	0	0	0	+-	-
Highest (prosodic word)	0	0	-	+	-	+	0	0	+	-
Longest (prosodic word)	-	0	0	+	+	-	0	0	+	-
Most salient F_0 (prosodic word)	0	0	+	-	+	-	+	0	-	0
Highest (utterance)	0	0	0	0	0	0	0	0	+	-
Longest (utterance)	0	0	0	0	0	0	0	0	+	-
Most salient (utterance)	0	0	0	0	+	0	0	0	-	0
Sentence's slope	+	0	0	0	0	0	0	0	0	-

Grabe et al. (2007) proposed a detailed account on both types of approaches, based on feature sets and on smooth curves, so to evaluate the adequacy of the first approach (based on ToBI) to describe prosodic strategies across English urban varieties. The shapes of nuclear accents were analysed thanks to Legendre polynomials, with the F_0 contours weighted by the signal's loudness and periodicity. Six out of the seven autosegmental-metrical label shapes were related to significantly different F_0 contour shapes, with a potential limit due to data sparsity. This result enhances the coherence of both approaches, showing their coherent conclusions.

QUANTITATIVE MEASUREMENT OF PROSODIC VARIATION

Once prosodic variation has been described in a systematic manner, whatever the model selected, the next task for a geolinguistic of prosody is to quantify its variation. Several propositions have been made for comparing other types of linguistic changes (reviewed in the introductory part) but can readily be applied to prosody; other proposals have been designed specifically for quantifying prosodic differences.

Table 1: reproduction of prosodic variations observed on sentences produced by one female and one male speaker from three cities, with two sentence modes (assertion: Ass., interrogation: Int.). Prosody is described in terms of autosegmental-metric labels (ToBI) and as AMPER stylized curves (red for assertion, blue for interrogation, 3 F_0 [measured in semitones] points per vowel, devoiced vowels marked by a low straight line).

City	Gender	Mode	ToBI	AMPER
Fortaleza	Female	Ass.	L+H* H+L*L%	
		Int.	L+H* H+H*L%	
	Male	Ass.	L+H* H+L*L%	
		Int.	L+H* L+H*L%	
Salvador	Female	Ass.	L+H* H+L*L%	
		Int.	L+H* L+H*L%	
	Male	Ass.	L+H* H+L*L%	
		Int.	L+H* L+H*H%	
Rio de Janeiro	Female	Ass.	H* H+L*L%	
		Int.	H* L+H*L%	
	Male	Ass.	H* H+L*L%	
		Int.	H* L+H*L%	

To apply different measurements of prosodic divergences as example of their application, I'll use the results of Rebollo Couto et al. (2017) who described prosody in three varieties of Brazilian Portuguese (Salvador, Fortaleza and Rio de Janeiro) following the AMPER methodology (Contini et al. 2002) and in terms of autosegmental-metric labels adapted to Portuguese (Frota & Moraes 2016). I'll take advantage of this versatility to show that several approaches are possible to get quantitative measurements of prosodic variations that may fit different research questions. In Rebollo Couto et al. (2017), the sentence "*O Renato gosta do Renato*" (Renato likes Renato) is presented for two speakers in each of these three cities, with assertive and interrogative modes. Table 1 summarizes its prosodic characteristics as seen within the autosegmental-metric framework (for the pre-nuclear and nuclear parts), and within the phonetic approach recommended by the AMPER project (in terms of the F_0 contour). As an example, the feature matrix of one sentence is presented in figure 1.

Divergence measures based on features

Classical dialectometric approaches use different versions of edit distance metrics to compare sets of features; such distance measurement count the number of transformations (possibly deletion, insertion, substitution and transposition) required to convert one string of characters into another; many variants exist in the literature, allowing notably to weight the different operations (see Navarro 2001 for a technical survey, Heeringa 2004 for their use for dialectometry and Heeringa et al. 2006 for an evaluation of several forms of them). Such metrics apply directly to measures of prosodic features – being based on matrices or sets of labels.

Assertive sentences in table 1 all have the same nuclear ToBI pattern (H+L*L%), differing only for the prenuclear part (L+H* or H*); their Damerau-Levenshtein distance (hereafter DL) equals two (that correspond to the missing "L+" at the beginning of Rio de Janeiro prenuclear accent). Thus for assertive mode, the Fortaleza and Salvador varieties have the same phonologic patterns (i.e. DL (Fortaleza, Salvador) = 0), while they both have a distance of 2 with the Rio de Janeiro variety. For interrogative sentences, the sum of differences between all pair of speakers from two different cities equals 10 for Rio de Janeiro / Fortaleza and Rio de Janeiro / Salvador, but 4 for Salvador / Fortaleza, as variations in the pre-nuclear part are systematic in Rio compared to the other two cities.

This string metric also applies to the feature matrices: the first feature (i.e. the pattern of F_0 movement on each vowel) for interrogative sentences is respectively “0 0 + 0 0 0 0 0 + -”, “0 0 0 0 0 0 0 0 ++ 0”, “0 0 0 0 + 0 0 0 0+ -+” (with spaces separating features of each vowel) for the female and male speakers from Rio de Janeiro, and the female from Salvador; the two carioca speakers’ prosodic shapes diverge by a DL distance of 3, while each carioca speaker’s prosody has a DL distance of 4 with Salvador’s female production. The DL distance may be applied in a similar way to each feature (i.e. each pair of matrix’s lines), the sum of these distances across features giving an idea of the quantitative difference between prosodic patterns of two sentences.

Another possible approach would build on corpus based frequency analyses, looking at the frequency ratio between types of accents (or shapes of contours, etc.) occurring in each dialect under investigation. The approach defended for lexical analysis by Speelman & Geeraerts (2008) may here prove interesting: “profiles” of accentual patterns may be created, that list possible performances for a given prosodic function (e.g. assertion, interrogation), and extract the relative frequency of occurrence for each types of accentual patterns. In the current example, interrogation would possibly be actualized as “L+H* H+H*L%”, “L+H* L+H*L%”, “L+H* L+H*H%”, or “H* L+H*L%” under their ToBI description. A particular dialect will use these variants with a specific frequency (the Rio de Janeiro variety possibly focusing on one pattern only, while the two others cities may show a more diverse profile). The patterns of frequency distribution actually observed in a corpus for several prosodic functions may serve as a dependant variable for a statistical analysis that will compare such prosodic “profiles” across dialects (see Wieling & Nerbonne 2015 for a review). Details of the methodology are discussed, for lexical variation and its application to geolinguistic variation, in Speelman & Geeraerts (2008) – I don’t know any use of this concept for prosody, but it easily apply to, and may prove efficient as it comes with a robust methodology.

Application of regression modelling to the parameters of orthogonal polynomial decompositions allows evaluating the relative role of a set of factors (including geographic origin) to prosodic characteristic (see Kochansky et al. 2005; Grabe et al. 2007; Lai 2014). In such and approach, bounded contours are described by their shape components and their proximity may be evaluated and compared across geographic and/or social variation, showing the similarities between feature-based and some contour-based approaches. The parameters extracted from an application of the Fujisaki model may be treated in a similar way so to analyse prosodic variation (MIXDORFF & PFITZINGER 2005).

Divergence measures based on contours

When prosody is described as contours, other approaches are generally used to quantitatively compare them. An emblematic method was proposed and evaluated by Hermes (1998b), who shows the correlation between two F_0 contours, weighted by the maximum amplitude of the subharmonic sumspectrum (a measure of the intensity in the speech signal that contributes to F_0 perception; the measure was also applied elsewhere with signal intensity, or other loudness measurement), is the measure that best fit the perceived difference between two contours (HERMES 1998a). The measure is described in equation 1 (adapted from Hermes 1998b to the case of discrete F_0 measures, as in d'Alessandro et al. 2011), were *r* is the weighted correlation of two vectors of *n* F_0 measures f_1 and f_2, w is a weighting vector of length n, and μ_1 and μ_2 are the respective means of f_1 and f_2.

(1)

$$\boxed{r_{f_1 f_2} = \frac{\sum_{i=1}^{n} w(i)\left(f_1(i) - \mu_1\right)\left(f_2(i) - \mu_2\right)}{\sqrt{\sum_{i=1}^{n} w(i)\left(f_1(i) - \mu_1\right)^2 \sum_{i=1}^{n} w(i)\left(f_2(i) - \mu_2\right)^2}}}$$

This correlation measurement was applied to the comparison of prosodic performances for synthetic speech evaluation (HIRST et al. 1998; D'ALESSANDRO et al. 2011). The same measure was also used for the observation of prosodic divergences across dialects, in a seeding article on geoprosody (MOUTINHO et al. 2011). The measure is similar to others proposed for the same aim (see ROMANO et al. 2011 for an historic) but that propose unweighted measures (correlation or covariance) between two contours. Has shown by Hermes (1998b), considering the voicing strength adds much to the perceptual relevance of quantitative measurement, minimizing the influence of possibly ample F_0 movements (e.g. linked to microprosodic effects or F_0 rises after vowels, that may not be intended nor controlled) that have few or none functional relevance, as performed at very low intensity levels (and so are almost not perceived).

Some tools do implement the weighted correlation measures for dialectometric purposes; Martínez Calvo & Fernández Rei (2015) proposes an implementation in the R software; Elvira-García et al. (2018) proposes an integrated tool based on this distance, and targeting AMPER-style data, but also data under a raw acoustic form. Their implementation of the distance allows varied forms of weight, interestingly one that possibly includes segmental duration as a weighting factor.

The weighted correlation measures a similarity between two sets of points; this raises several difficulties. First, the two F_0 vectors to be compared shall have the same length, which requires a normalization so to deal with time differences between utterances (see Xu 1999 for an example of such time normalization), such linear normalizations can be contemplated only for cases of similar structure or length. Second, missing values raise difficulties when comparing two continua; such missing values are particularly frequent, but not systematic, for post-tonic final vowels in Brazilian Portuguese[1]. Third, its output is bounded between [-1; 1], so Hermes (1998b, p. 75) proposes to apply Fisher's Z-transform so to get values in the [0; +∞] range – the higher this measure, the more similar the two compared contours: it is thus (as mentioned by Hermes) a *similarity* measure that inversely correlate with *dissimilarity* measures between contours that have described up to now (e.g. string metric; on similarity and distance, see also Heeringa et al. 2006).

The numerical application of the weighted correlation gives the *similarities* (Z-transformed of the correlations) reported in table 2, for the AMPER-stylization reported in table 1. All pairs of sentences with the same modality have been compared, which makes three types of pairs: same-speaker pairs, same-city pairs (two speakers from the same city), and different-city pairs. Means of these measures are reported, aggregating all pairs from one city together (same speaker or not), separately for each modality. The mean similarity for assertive sentences is 1.06, and 0.59 for interrogative ones: there is much more cross-dialectal variation in interrogative performances, as the results obtained from ToBI labels had already shown.

---

[1] Solutions exist, such as interpolation, but have to be adapted specifically to each situation.

Table 2: mean of the weighted correlation's Z-transforms between each pair of sentences presented in table 1 (same modality) between speakers from the three cities (Rio de Janeiro, Salvador and Fortaleza). See text for details.

Mode	City	Rio de Janeiro	Salvador	Fortaleza
Assertion	Rio de J.	1.26	1.10	0.87
	Salvador		1.58	0.83
	Fortaleza			0.85
Interrogation	Rio	1.05	0.36	0.51
	Salvador		1.08	0.19
	Fortaleza			0.61

The correlation measure between speakers of the same city gives information on the coherence of the speakers in terms of prosodic strategy, and thus on variability of prosodic patterns there. For example, the two speakers from Fortaleza show low similarity rating, compared to what is observed in other cities (especially for interrogatives: 0.61 vs. values superior to 1). Comparing intra-individual variation (i.e. coherence of one speaker to reproduce similar pattern across repetitions) and inter-speaker measures (for the same city), is possible to propose inferences on the representativeness of measured contours as dialect-specific, while inter-city measurements are indicative of variation between varieties[2] (for such an approach, even if based on another measure, see Grabe et al. 2007).

In our example, interrogative productions from Salvador are more dissimilar to those of other cities. Note that the results found here are not completely coherent with what was calculated on the basis of the ToBI labels (*supra*). The large difference between Rio de Janeiro and the two north-eastern cities observed with ToBI labels is mostly related to the different encoding of the prenuclear part, while this part (also taken into account in the weighted correlation) sounds much less important in its magnitude than the discrepant changes in the nucleus. One may also apply the LD metric on nuclear accent only, but the greater difference between Fortaleza and Salvador (compared to Rio de Janeiro) would remain, as one speaker of both cities diverge from the phonological pattern described for Rio de Janeiro (Moraes 1998), but each in a specific way. Knowing how these

---

[2] The example data presented here are not sufficient for proposing a statistical approach, but examples exist in the literature: see Grabe et al. 2007 or Wieling et al. 2014.

variations are perceived, which variant is perceived as most distant from the others, would allow tuning quantitative metrics so they may more accurately reflect perception (by weighting specific parameters or operation: a string metric may weight differently each operation, for example by putting more load to deletion of a character than on its substitution). It is also important to note the example here is based on a ridiculously small subset; using more data allows a more robust analysis, and may reconcile the different measures.

As it was said earlier, measures of correlation (or similar ones) have problems linked with variations in the duration of spoken utterances to be compared. The most common solution consists in time normalization, as described in Xu (1999), and that acts at a phonemic level. Meanwhile, this approach can hardy solve duration differences linked with variation in structure (different sets of phonemes, even with equivalent number of syllables or other targeted structures). Rilliard et al. (2011) did propose the use of a non-linear adaptation of prosodic contour before applying the correlation measure. The non-linear process is based on the application of a dynamic-time-warping (DTW) algorithm (as in Jouanelle et al. 1981) to vectors of prosodic characteristics that includes the intonation contours but also duration and intensity patterns. The results show non-linear alignment allows reaching higher similarity measurements for shape-similar, but time varied contours. The DTW algorithm show strong similarities with string metrics, in an approach applied to signal processing: both target an optimal path to compare or transform pairs of vectors according to their similarities; DTW works on numeric values and string metrics on character. Both of these approaches aim at finding the relation between two sequences that could be represented as acoustic signals or as strings. Heeringa & Gooskens (2003) application of the Levenshtein distance to acoustic data is an excellent demonstration of this convergence.

AGGREGATION OF DISTANCES

Once a metric has been applied to the comparison of prosodic performances within and across speakers and dialectal areas, the dialectometric process consists in an agglomeration of these measurements so to reach representation of the variation at hand before descriptive and interpretative works. This part is not specific to prosody: similar methods can be applied to prosodic divergences that were already used since Goebl (1993, 2003, 2006). It basically consists in aggregating the output of objective metrics into a “distance matrix” (eventually converting similarity measures into divergence ones); this matrix is a symmetric table the cells of which present the metric's values that separate two *individuals* (individuals may be speakers, sentences, localities etc.): the diagonal presents the internal divergence within each individual (if any, it may be zeros), while the other cells represent between-individuals divergences.

This matrix, which may be very large, is then subjected to a multidimensional analysis (e.g. HUSSON et al. 2017 for a theoretical and practical account). This analysis will spread all individuals on a few abstract dimension that best structure the cloud of individuals. The Multidimensional Scaling (MDS) method is very often used, as it allows forcing the spread of individuals on the bidimensional output (i.e. a plane) that best represent the observed variation, and that easily fits on a map. Other favourite algorithms used to deal with distance matrices are clustering ones – and typically hierarchical clustering algorithm, that produces dendrogram representation of the individuals by iteratively grouping (or splitting) them in wider (or smaller) groups (see Husson et al. 2017, chapter 4, for details).

Such outputs are ideally used by a geographic information system (GIS) so to plot the result on a map – one frequent requirement for dialectology. A few GIS specialized in the representation of dialectological data: GabMap is a prominent one (many other flavours exist as a search of "GIS dialectology" in a search engine may show) that also comes with an R interface (http://www.let.rug.nl/~kleiweg/L04/) and allows geographic representation of the variation captured in a metric. The type of dialectological information to be represented varies: distance from a reference point (as a shade of colour), boundaries between variants or links within variants, or aggregated patches of clustered variants (e.g. Goebl 2006, Nerbonne 2009).

Another approach, much typical of sociolinguistics, is based on regression analysis, and may allow the observation of several levels of variations – social and geographic. The notion of distance is not necessarily summarized in a matrix, but still can be represented on a map, as the work presented in Wieling et al. (2014) may demonstrate. Wieling & Nerbonne (2015) proposes an excellent review of methods and approaches with lists of existing tools.

DISCUSSION & CONCLUSIONS

This article proposes a review of methods used to measure objective divergences between prosodic performances – typically in relation with dialectological variation. We have seen the prosodic aspect of speech is still under-resourced in the field of dialectology and sociolinguistic. Both field proposes some solutions, which may be seen as different, but do recently show their possibility to converge, and their potential application to other type of data. The main challenge (as prosody is concerned) certainly remains reliable objective measures of prosodic variation; other tools and methodologies (statistical aggregations and analysis, multidimensional analyses, and representation tools) have been already developed for the other levels of linguistic description and are more or less readily applicable to the case of prosody.

Propositions for such metrics have been reviewed, that show representation of prosody is the basis to set up divergence measurements. As many representation of prosody are linked to theoretical models (e.g. the ToBI annotation scheme to the phonological autosegmental-metric theory), the choice of one solution will generally not be neutral, and may be dictated by the special interest of a particular research: the AMPER project for example as a strong phonetic focus that rely on its aims to observe variation as a first step to describe it, and then having the capacity to categorize it. Grabe et al. (2007) have shown approaches based on very different materials may show a strong convergence; thus a readily applicable choice is certainly the better solution for a given research question, that will depend on the researcher familiarity with the various aspects of such approaches, and access to / familiarity with the related tools. The presentation of different measurements based on the same example dataset shows that there are similarities to be found between approaches, but also potentially differences linked to the importance placed on different aspects of the prosodic changes.

One such difference is related to the set of parameters used to describe prosody, and notably on the use of duration related measurements. Most models of prosody focus on intonation, which has certainly a prime importance – but rhythm and lengthening also carry critical information on the linguistic message (see Moraes 2008 for a discussion on some lengthening effects). The matrices of features advocated by Contini (1992) in his proposition of a geoprosodic approach to dialectal variation may include such durational level of information; Mixdorff & Pfitzinger (2005) proposes another possible solution based on Pfitzinger's (1998) proposal of a "perceptual local speech rate", that gives a continuous measure of rhythm. This approach shares similarities with the models of duration proposed by Campbell (1993) and Barbosa (2007). Such continuous representations of duration may be taken into account together with F_0 by approaches based on functional modelling – as well as other parameters as loudness or aperiodicity – as was done in Kochanski & Shih (2003) and Kochanski et al. (2005).

So to get more reliable tools, problems still need to be further investigated. The main one certainly is linked with the relations between the acoustic dimension of speech and its perceptual interpretation that is not done in a unique and systematic way for a given set of audio parameters, and neither is similar across language varieties. Complex challenges include setting up objective measures able to take into account pragmatic variables so to deal with changes linked to expressive speech (because of e.g. hierarchical relations between interlocutors; vocal effort control due to situation: noise, distance, need for discretion; expected behaviour in relation to social norms). Another challenge is related to the limited set of prosodic patterns compared to the complexity of pragmatic interpretation: prosodic meaning consists in a small set in comparison to pragmatic uses of language, but there is still few knowledge on the composition of this set (see Mixdorff et al. 2017 for a discussion).

A difficulty linked to these challenge is related to the complexity of some prosodic modelling, that involves complex statistical approaches, and that are not necessarily proposed within user-friendly computer program. Comparatively, the success of the varbrul analysis (CEDERGREN & SANKOFF 1974) in sociolinguistics is certainly due to the efficient computer programs that allow efficient applications of an otherwise quite complex mixed effects logistic regression. Some tools are available to apply dialectometric measures (see Wieling & Nerbonne 2015 for a review) and a few tools specialized in measuring prosodic divergences (notably Martínez Calvo & Fernández Rei 2015; Elvira-García et al. 2018), but there are still spaces for much work in that direction.

Finally, there is also questions linked with the varying sources of prosodic variation and its representation: if dialectology focuses mostly on the diatopic aspect, there is certainly much differences linked to the diastratic composition of the society, especially in urban areas and in cities where important difference in education are found within the population. Geographic changes are efficiently represented through techniques pioneered by Goebl (see the introduction), but they are not necessarily adapted to take into account important differences in the special density of survey points, and especially cases where variation is more social than geographic – a fact that is more and more common with population mobility and important urban concentrations. Reflexion and interdisciplinary collaborations (notably with geographers and computer vision scientists) are needed so to propose and adapt solutions to concrete cases of geoprosody.

ACKNOWLEDGMENTS

The author wants to thank the invaluable support of UFRJ in the completion of this article; I'm also indebted to Leticia Rebollo Couto to let me use her data as examples for this article, and Jussara Abraçado for discussions on onomasiological profiles.

REFERENCES

D'ALESSANDRO, Christophe; MERTENS, Piet. Automatic pitch contour stylization using a model of tonal perception. Computer Speech and Language, v. 9, n. 3, p. 257, 1995.

D'ALESSANDRO, Christophe; RILLIARD, Albert; LE BEUX, Sylvain. Chironomic stylization of intonation. The Journal of the Acoustical Society of America, v. 129, n. 3, p. 1594-1604, 2011.

ANDRUSKI, J. E.; COSTELLO, J. Using polynomial equations to model pitch contour shape in lexical tones: an example from Green Mong. Journal of the International Phonetic Association, v. 34, n. 02, p. 125-140, 2004.

BAAYEN, R. Harald. Analyzing linguistic data: A practical introduction to statistics using R. Cambridge University Press, 2008.

BARBBOSA, Plinio. From syntax to acoustic duration: A dynamical model of speech rhythm production. Speech Communication, v. 49, n. 9, p. 725–742, 2007.

BAYLEY, Robert. The quantitative paradigm. In Chambers, J. K., Peter Trudgill and Natalie Schilling-Estes (eds) The handbook of language variation and change, p. 117-141, 2003.

VAN BEZOOIJEN, Renée; GOOSKENS, Charlotte. Identification of language varieties: The contribution of different linguistic levels. Journal of language and social psychology, v. 18, n. 1, p. 31-48, 1999.

BOERSMA, Paul; WEENINK, David. Praat: doing phonetics by computer [Computer program]. Version 6.0.35. Online: http://www.praat.org/. Retrieved in 2018.

CAMPBELL, Nick. Automatic detection of prosodic boundaries in speech. Speech Communication, v. 13, n. 3, p. 343–354, 1993.

CEDERGREN, Henrietta J.; SANKOFF, David. Variable rules: Performance as a statistical reflection of competence. Language, v. 50, n. 2, p. 333-355, 1974.

CHAMBERS, Jack K.; TRUDGILL, Peter. Dialectology (second edition). Cambridge University Press, 2004.

CONTINI, Michel. Vers une géoprosodie. In: Proceedings of the Nazioarteko Dialektologia Biltzarra, IKER 7, Euskaltzaindia. Bilbao: Real Academia de la Lengua Vasca, p. 83-109, 1992.

CONTINI, Michel; PROFILI, Olga. L'intonation de l'italien régional – un modèle de description par traits. In BOTHOREL, A. et al. (Eds.), Mélanges de phonétique générale et expérimentale offerts à Péla Simon, Strasbourg: Publications de l'Institut de Phonétique de Strasbourg, p. 855-870, 1989.

CONTINI, Michel; LAI, Jean-Pierre; ROMANO, Antonio; ROULLET, Stefania; MOUTINHO, Lurdes de Castro; COIMBRA, Rosa Lídia; BENDIHA, Urbana Pereira; RUIVO, Suzana Secca. Un projet d'atlas multimédia prosodique de l'espace roman. In: Proceedings of the International Conference of Speech Prosody, p. 227-230, 2002.

CRUZ, M.; OLIVEIRA, P.; PALMA, P.; NETO, B.; FROTA, S. Building a prosodic profile of European Portuguese varieties: The challenge of mapping intonation and rhythm. In BARBOSA, P.; PAIVA, M. C.; RODRIGUES, C. (Eds.), Studies on Variation in Portuguese. Amsterdam: John Benjamins, p. 81-110, 2017.

ELVIRA-GARCÍA, Wendy; BALOCCO, Simone; ROSEANO, Paolo; FERNÁNDEZ-PLANAS, Ana Ma. ProDis: A dialectometric tool for acoustic prosodic data. Speech Communication, 2018, vol. 97, p. 9-18.

FROTA, Sónia; MORAES, João Antônio. Intonation in european and brazilian portuguese. In: Wetzels, W.L; Menuzzi, S.; Costa, J. (Eds.), Handbook of Portuguese Linguistics. 1ed. Hoboken: Wiley-Blackwell, p. 141-166, 2016.

FUJISAKI, Hiroya. Dynamic characteristics of voice fundamental frequency in speech and singing. In MACNEILAGE, P. (Ed), The production of speech. New York: Springer, p. 39–55, 1983.

FUJISAKI, Hiroya. A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In FUJIMURA, O. (Ed.), Vocal fold physiology: voice production, mechanisms and functions. New York: Raven, p. 347–355, 1988.

FUJISAKI, Hiroya. Information, prosody, and modelling - with emphasis on tonal features of speech. In: Proceedings of the International Conference on Speech Prosody, Nara, Japan, p. 1-10, 2004.

GOEBL, Hans. Dialectometry: a short overview of the principles and practice of quantitative classification of linguistic atlas data. In: Reinhard Köhler; Burghard B. Rieger (Org.) Contributions to quantitative linguistics. Springer, Dordrecht, p. 277-315, 1993.

GOEBL, Hans. Regards dialectométriques sur les donées de l'Atlas linguistique de la France (ALF): Relations quantitatives et structures de profondeur. Estudis Romànics, p. 60-117, 2003.

GOEBL, Hans. Recent advances in Salzburg dialectometry. Literary and linguistic computing, v. 21, n. 4, p. 411-435, 2006.

GOOSKENS, Charlotte Stenkilke. On the role of prosodic and verbal information in the perception of Dutch and English language varieties. PhD Thesis, Katholieke Universiteit Nijmegen, 1997.

GOOSKENS, Charlotte; HEERINGA, Wilbert. Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. Language variation and change, v. 16, n. 3, p. 189-207, 2004.

GOOSKENS, Charlotte; HEERINGA, Wilbert. The relative contribution of pronunciational, lexical, and prosodic differences to the perceived distances between Norwegian dialects. Literary and Linguistic Computing, v. 21, n. 4, p. 477-492, 2006.

GOOSKENS, Charlotte. Non-linguists' judgments of linguistic distances between dialects. Dialectologia: revista electrònica, n. 9, p. 27-51, 2012.

GOOSKENS, Charlotte. Experimental methods for measuring intelligibility of closely related language varieties. The Oxford handbook of sociolinguistics, p. 195-213, 2013.

GOOSKENS, Charlotte. Methods for measuring intelligibility of closely related language varieties. In: BAYLEY, Robert; CAMERON, Richard; LUCAS, Ceil (eds.). Handbook of sociolinguistics. Oxford University Press, p. 195-213, 2013.

GRABE, Esther; KOCHANSKI, Greg; COLEMAN, John. Connecting intonation labels to mathematical descriptions of fundamental frequency. Language & Speech, v. 50, n. 3, p. 281-310, 2007.

HADJIPANTELIS, Pantelis Z.; ASTON, John AD; EVANS, Jonathan P. Characterizing fundamental frequency in Mandarin: A functional principal component approach utilizing mixed effect models. The Journal of the Acoustical Society of America, v. 131, n. 6, p. 4651-4664, 2012.

‘T HART, Johan; COLLIER, René; COHEN, Antonie. A perceptual study of intonation: an experimental-phonetic approach to speech melody. Cambridge: Cambridge University Press, 1990.

HEERINGA, Wilbert Jan. Measuring dialect pronunciation differences using Levenshtein distance. PhD, University Library Groningen. 2004.

HEERINGA, Wilbert; NERBONNE, John. Dialect areas and dialect continua. Language Variation and Change, v. 13, n. 3, p. 375-400, 2001.

HEERINGA, Wilbert; GOOSKENS, Charlotte. Norwegian dialects examined perceptually and acoustically. Computers and the Humanities, v. 37, n. 3, p. 293-315, 2003.

HEERINGA, Wilbert; KLEIWEG, Peter; GOOSKENS, Charlotte; NERBONNE, John. Evaluation of string distance algorithms for dialectology. In: Proceedings of the workshop on linguistic distances. Association for Computational Linguistics, p. 51-62, 2006.

HEERINGA, Wilbert; NERBONNE, John. Dialectometry (Chap. 33). In HINSKENS, Frans; TAELDEMAN, Johan (eds.): Language and Space. An International Handbook of Linguistic Variation, Volume III: Dutch. (Series: Handbook of Linguistics and Communication Science (HSK)). Berlin and New York: Walter de Gruyter, p. 624-646, 2013.

HERMES, Dik J. Auditory and visual similarity of pitch contours. Journal of Speech, Language, and Hearing Research, vol. 41, no 1, p. 63-72, 1998a.

HERMES, Dik J. Measuring the perceptual similarity of pitch contours. Journal of Speech, Language, and Hearing Research, vol. 41, no 1, p. 73-82, 1998b.

HIRST, Daniel; ESPESSER, Robert. Automatic modelling of fundamental frequency using a quadratic spline function. Travaux de l’Institut de Phonétique d’Aix, v. 15, p. 75–85, 1993.

HIRST, Daniel; RILLIARD, Albert; AUBERGÉ, Véronique. Comparison of subjective evaluation and an objective evaluation metric for prosody in text-to-speech synthesis. In: Proceedings of the Third ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis, p. 1-4, 1998.

HIRST, Daniel; DI CRISTO, Albert; ESPESSER, Robert. Levels of representation and levels of analysis for the description of intonation systems. In: Merle Home (Ed.) Prosody, theory and experiment: studies presented to Gŏsta Bruce. Springer, Dordrecht, p. 51-87, 2000.

HUALDE, José Ignacio; PRIETO, Pilar. Intonational variation in Spanish: European and American varieties. In: FROTA, Sónia; PRIETO, Pilar (Eds.), Intonation in Romance. Oxford: Oxford University Press, p. 350-391, 2015.

HUSSON, François; LÊ, Sébastien; PAGÈS, Jérôme. Exploratory multivariate analysis by example using R. Chapman and Hall/CRC, 2017.

INOUE, Fumio. Computational dialectology (2). Area and Culture Studies, v. 53, p. 115-134, 1996.

INOUE, Fumio. A century of language change in progress. New dialect in Tsuruoka, Japan. Dialectologia: revista electrònica, v. 17, p. 71-89, 2016.

JOUANNELLE, G.; BOË, L. J.; BELLET, G. Variation intra et inter locuteur de F0: procédures de normalisation – étude préliminaire. In: Proceeding of the 12es JEP, Montréal, 246–262, 1981.

KOCHANSKI, Greg; SHIH, Chilin. Prosody modeling with soft templates. Speech Communication, v. 39, p. 311-352, 2003.

KOCHANSKI, Greg, GRABE, Esther, COLEMAN, John, et al. Loudness predicts prominence: Fundamental frequency lends little. The Journal of the Acoustical Society of America, v. 118, n. 2, p. 1038-1054, 2005.

KUIPER, Lawrence. Parisian Perceptions of Regional French. In PRESTON, Dennis R. (Ed.), Handbook of perceptual dialectology, v. 1, p. 243-262, 1999.

LAI, C. Interpreting Final Rises: Task and Role Factors. In: Proceedings of Speech Prosody, Dublin, Ireland, p. 520-524, 2014.

LEVITT, Harry; RABINER, Lawrence R. Analysis of fundamental frequency contours in speech. The Journal of the Acoustical Society of America, v. 49, n. 2-2, p. 569-582, 1971.

MARTIN, Philippe. Analyse phonologique de la phrase française. Linguistics, v. 13, n. 146, p. 35-68, 1975.

MARTIN, Philippe. Phonetic realisations of prosodic contours in French. Speech communication, v. 1, n. 3-4, p. 283-294, 1982.

MARTIN, Philippe. Prosodic and rhythmic structures in French. Linguistics, v. 25, n. 5, p. 925-950, 1987.

MARTIN, Philippe. Structures prosodiques des langues romanes. In: Actes de la conférence conjointe JEP-TALN-RECITAL 2016, volume 1: JEP, p. 678-686, 2016.

MARTÍNEZ CALVO, Adela; FERNÁNDEZ REI, Elisa. Unha ferramenta informática para a análise dialectométrica da prosodia. Estudios de fonética experimental, v. 24, p. 289-303, 2015.

MASE, Yashio. On Dialect Consciousness – Dialect Characteristics Given by Speakers. In PRESTON, Dennis R. (Ed.), Handbook of perceptual dialectology, v. 1, p. 101-113, 1999.

MERTENS, Piet. The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In: Proceedings of the International Conference on Speech Prosody. 2004.

MERTENS, Piet. Polytonia: a system for the automatic transcription of tonal aspects in speech corpora. Journal of Speech Sciences, v. 4, n. 2, p. 17-57, 2014.

MIXDORFF, Hansjörg. A novel approach to the fully automatic extraction of Fujisaki model parameters. In: IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100). IEEE. p. 1281-1284, 2000.

MIXDORFF, Hansjörg; PFITZINGER, Hartmut R. Analysing fundamental frequency contours and local speech rate in map task dialogs. Speech Communication, v. 46, n. 3-4, p. 310-325, 2005.

MIXDORFF, Hansjörg, HÖNEMANN, Angelika, RILLIARD, Albert, LEE, Tan; MA, Matthew K. A. Audio-visual expressions of attitude: How many different attitudes can perceivers decode?. Speech Communication, v. 95, p. 114-126, 2017.

MORAES, J. A. de. Intonation in Brazilian Portuguese. In: HIRST, Daniel; DI CRISTO, Albert. (Org.). Intonation Systems: a Survey of Twenty Languages. Cambridge: Cambridge University Press, p. 179-194, 1998.

MORAES, J. A. de. The pitch accents in Brazilian Portuguese: analysis by synthesis. In: Proceedings of the International Conference of Speech Prosody, v. 4, p. 389-397, 2008.

MOUTINHO, L. de Castro; COIMBRA, R. L.; RILLIARD, A.; ROMANO, A. Mesure de la variation prosodique diatopique en portugais européen. Estudios de fonética experimental, v. 20, p. 33-55, 2011.

NAVARRO, Gonzalo. A guided tour to approximate string matching. ACM computing surveys (CSUR), v. 33, n. 1, p. 31-88, 2001.

NERBONNE, John. Data-driven dialectology. Language and Linguistics Compass, v. 3, n. 1, p. 175-198, 2009.

NERBONNE, John; COLEN, Rinke; GOOSKENS, Charlotte; KLEIWEG, Peter; LEINONEN, Therese. Gabmap-a web application for dialectology. Dialectologia: revista electronica, Special issue II, p. 65-89, 2011.

OLIVE, Joseph P. Fundamental frequency rules for the synthesis of simple declarative English sentences. The Journal of the Acoustical Society of America, 1975, v. 57, n. 2, p. 476-482.

PEIRSMAN, Yves; GEERAERTS, Dirk; SPEELMAN, Dirk. The automatic identification of lexical variation between language varieties. Natural Language Engineering, v. 16, n. 4, p. 469-491, 2010.

PFITZINGER, Hartmut R. Local speech rate as a combination of syllable and phone rate. In: Proceedings of the fifth international conference on spoken language processing, 1998.

REBOLLO COUTO, Leticia; GOMEZ DA SILVA, Carolina; DA SILVA MIRANDA, Luma. Prosódia de enunciados declarativos e interrogativos totais nas variedades de Salvador, Fortaleza e Rio de Janeiro. Revista de Estudos da Linguagem, v. 25, n. 3, p. 1105-1142, 2017.

RILLIARD, Albert. Prosodie et Interaction Homme-Machine : Étude de la variation démarcative, diatopique, diachronique & expressive. Mémoire d'Habilitation à Diriger des Recherches, Université Paris Sud, France, 2014.

RILLIARD, Albert; ALLAUZEN, Alexandre; BOULA DE MAREÜIL, Philippe. Using dynamic time warping to compute prosodic similarity measures. In: Proceedings of the twelfth Annual Conference of the International Speech Communication Association (INTERSPEECH), p. 2021-2024, 2011.

ROMANO, Antonio; CONTINI, Michel; LAI, Jean Pierre; RILLIARD, Albert. Distancias prosódicas entre variedades románicas en el marco del proyecto AMPER. Revista Internacional de Lingüística Iberoamericana, v. 9, n. 1, p. 13-25, 2011.

ROMNEY, A. Kimball; MOORE, Carmella C., BATCHELDER, William H., & HSIA, Ti-Lien. Statistical methods for characterizing similarities and differences between semantic structures. Proceedings of the National Academy of Sciences, v. 97, n. 1, p. 518-523, 2000.

SÉGUY, Jean. La relation entre la distance spatiale et la distance lexicale. Revue de linguistique romane, v. 35, p. 335-357, 1971.

SILVERMAN, Kim; BECKMAN, Mary; PITRELLI, John; OSTENDORF, Mari; WIGHTMAN, Colin; PRICE, Patti; PIERREHUMBERT, Janet; HIRSCHBERG, Julia. ToBI: A standard for labeling English prosody. In: Second International Conference on Spoken Language Processing (ICSLP), Banff, Canada, p. 867-870, 1992.

SPEELMAN, Dirk; GEERAERTS, Dirk. The role of concept characteristics in lexical dialectometry. International Journal of Humanities and Arts Computing, v. 2, n. 1-2, p. 221-242, 2008.

WEINER, E. Judith; LABOV, William. Constraints on the agentless passive. Journal of linguistics, v. 19, n. 1, p. 29-58, 1983.

WIELING, Martijn; MONTEMAGNI, Simonetta; NERBONNE John; BAAYEN, R. Harald. Lexical differences between Tuscan dialects and standard Italian: accounting for geographic and socio-demographic variation using generalized additive mixed modeling. Language, v. 90, n. 3, p. 669 – 692, 2014

WIELING, Martijn; NERBONNE, John. Advances in dialectometry. Annual Review of Linguistics, v. 1, n. 1, p. 243-264, 2015.

XU, Yi. Effects of tone and focus on the formation and alignment of F0 contours. Journal of phonetics, v. 27, n. 1, p. 55-105, 1999.