

Chapter 38

Morphology changes faster than phonology¹

Esteve Valls

University of Barcelona

In Valls, Wieling & Nerbonne (2013), we proved that the north-western Catalan varieties of Catalonia and Andorra (in contrast to those located in Aragon) have undergone a process of dedialectalization due to linguistic advergence to standard Catalan. In this paper, we intend to analyze whether this process affects the morphological and phonological components of language differently. To do so, we have performed a generative analysis of the corpus that has allowed us to discriminate the unpredictable (i.e. morphophonological) elements from the predictable (i.e. strictly phonological) components of language.² Therefore, we have been able to calculate the morphological and phonological distance among the north-western varieties separately and to visualize (by means of multidimensional analysis) the different evolutions of the morphological and the phonological components of these varieties throughout four generations. We attempt to shed light on the debate on which language components are more likely to be influenced by standard languages. We conclude that, at least in north-western Catalan, the hierarchy of instability of linguistic elements is clear: morphology has leveled faster than phonology. As a result, most of these varieties have undergone a process of accentualization.

1 Introduction

In Valls, Wieling & Nerbonne (2013) we took advantage of a range of dialectometric methods that allowed us to calculate and analyse the linguistic distance between 41 varieties in apparent time from an aggregate perspective. Specifically, we paid attention to the process of structural dialect loss due to linguistic advergence to standard Catalan undergone by most north-western Catalan dialects located in Catalonia

¹ This work is part of the project FFI2013-46987-C3-1-P (www.ub.edu/GEVAD), funded by the Spanish Ministerio de Economía y Competitividad.

² In this paper, we will refer to *morphophonological components* simply as *morphological components* in order to clearly contrast the differences between morphophonological underlying elements and strictly phonological elements (i.e. the phonological rules that produce the phonetic outputs).

(Spain) and Andorra. We also provided evidence that the dialect levelling that takes place in these two areas strongly contrasts with the relative stability of the Catalan dialects on the other side of the Catalan-Aragonese border in Spain, due to the fact that Catalan is not an official language in the Autonomous Community of Aragon. These opposite sociolinguistic situations (i.e. Catalonia and Andorra have strong language policies to support Catalan, whereas Aragon does not) have triggered a twofold process of vertical advergence between the Catalan spoken in Catalonia and Andorra towards the prestigious variety, on the one hand, and of horizontal divergence between these dialects and those located in Aragon, on the other hand. This situation has notably strengthened the border differences between Aragon and Catalonia during the last 80 years.

This paper is a step forward in the analysis of the abovementioned process of dedialectalization, because we intend to study the different *speeds* of change depending on the *nature* of language components from an aggregate perspective. Our hypothesis is that morphological aspects of language tend to change faster than phonological ones, a situation that would lead to *accentualization* according to Chambers & Trudgill (1998: 5). Because of space limitations, we will only mention five hierarchies of instability of linguistic components to show that there is no consensus about which of these two categories tend to change faster due to prestige reasons in situations of language or dialect contact:

- a) According to Viaplana (1999), who also focussed on north-western Catalan: morphology > phonology.
- b) According to Sankoff (2002), who focussed on several language contact situations: lexicon > phonology > morphology / syntax.
- c) According to Berruto (2005), who focussed on the influence of standard Italian on Italo Romance languages: lexicon > phonetics / phonology > morphology / syntax.
- d) According to van Coetsem (1988), who focussed on several language contact situations: lexicon > syntax and phonology > morphology.
- e) According to van Bree (1985), who focussed on the influence of Netherlandic Dutch on two Netherlandic dialects: lexicon > morphology > phonology > syntax.

2 Corpus

The dataset used in this paper was conceived as a corpus of contemporary north-western Catalan and covers the whole area where this dialect is spoken: all of Andorra and two dialect areas within Spain, specifically the western half of the Autonomous Community of Catalonia (with the exception of the Occitan-speaking Val d'Aran) and the eastern counties of the Autonomous Community of Aragon. Fieldwork was carried out in forty villages (two in Andorra, eight in Aragon, and thirty in

Catalonia) located in twenty counties. We added an artificial variety, standard Catalan, to these forty localities; hence, on the whole, we examine forty-one varieties. We interviewed 320 informants, 8 per locality, divided into four age groups. We selected a subset of 363 glosses per informant from the general questionnaire. These are distributed in eight (regular) morphological categories: articles, locative adverbs, verbs, and clitic, demonstrative, neuter, possessive and personal pronouns. The final corpus contains 113,749 items and 680,639 sound segments. More details about it can be found in Valls, Wieling & Nerbonne (2013).

3 Methodology

To examine the different speed of linguistic change on morphology and phonology, we first needed to discriminate, by means of a generative analysis, the unpredictable components of language (i.e. the underlying morphophonological differences) from its predictable elements (i.e. the regular phonological phenomena that generate the phonetic outputs). The aim was to calculate the linguistic distance for both categories separately in order to analyse how they evolved from the older speakers (F4) to the younger speakers (F1).

We illustrate this point by looking at the complete paradigm of the first person singular pronominal clitic (in bold in the examples below) in Cervera, Catalonia (see Table 1). The first two forms are proclitics followed by a consonant (a) and a vowel (b); the last two forms are enclitics preceded by a consonant (c) and a vowel (d).

In the examples below, the older speakers (F4) from Cervera use the form [me] to refer to the first person singular pronominal clitic before a verb starting with a consonant: [me'rento] 'I wash myself'. Younger speakers (F1), instead, use a more standardized form in the same context: [em]. Therefore, we might think that the underlying morphological forms of this clitic are /me/, /em/ and /m/. The fact is, however, that this clitic only appears with a vowel when the addition of the clitic to the verb creates a sequence that cannot be properly syllabified, that is: preceding a consonant ($_ \#C$: [me]/[em] *rento*) or preceded by a consonant ($C\#_$: *irritant*[me]). This is crucial to come to the conclusion that the underlying morphological form of the clitic is /m/ for both F4 and F1 speakers and that, being [e] the unmarked epenthetic vowel of north-western Catalan, in clitic clusters it is inserted to satisfy syllabic requirements.

Under this view, the difference between F4 and F1 is not morphological but merely phonological, as it lies in the position of the epenthesis. In F4, the epenthetic vowel is always placed to the right of the clitic, i.e. [me] in (a) and (c). In F1, instead, it always appears at the periphery of verb-clitic sequences, i.e. at the beginning in (a) but at the end in (c). That indicates that there has been a phonological change between F4 and F1 due to the different strategies that speakers undergo to repair syllabification. As a consequence, during the generative analysis we decided to assign a *Phonological Rule 1: epenthesis to the right* to the F4 speakers and a *Phonological Rule 2: peripheral epenthesis* to the F1 speakers. Thus, the morphological distance between F4 and F1 is 0, but the phonological distance between these two age cohorts is 1.

Table 1: Complete paradigm of the first person singular pronominal clitic in Catalan. It may appear in four contexts: $_ \#C$ (a), $_ \#V$ (b), $C\# _$ (c) and $V\# _$ (d).

	Cervera (F4)	Cervera (F1)
a. <i>em rento</i> ‘I wash myself’	[meˈrento]	[emˈrento]
b. <i>m’irrita</i> ‘he irritates me’	[miˈrita]	[miˈrita]
c. <i>irritant-me</i> ‘irritating me’	[iriˈtamme]	[iriˈtamme]
d. <i>renta’m</i> ‘wash me’	[rentam]	[rentam]

This is the method we followed to discriminate the unpredictable components of language (i.e. the underlying morphological differences) from its predictable elements (i.e. the regular phonological rules that produce the phonetic outputs). As a result, we obtained two different databases: one containing the underlying morphological forms and another one comprising the phonological phenomena involved. To simplify the analysis, each phenomenon has been assigned a number. In addition, all the inputs (both the underlying forms and the processes) have been manually aligned. The measure of distance we used is the following:

$$\text{dist}(i, j) = \frac{\sum_{k=1}^{\text{long}} \text{dif}_k(i, j)}{\text{long}} \times 100 \quad (1)$$

That is, the linguistic distance between two varieties (i, j) is the result of the summation of their differences (each having a value of 1) with regard to a linguistic variable k and dividing them by *long*, which is the length of each item compared. This dialectometric method was developed by Viaplana (1999) at the University of Barcelona. More detailed information about it can be found in Clua & Lloret (2006) and Valls et al. (2011).

4 Results

In this section we present six plots which show the results of applying multidimensional scaling to three distance matrices: Figs. 1 and 2 show the linguistic distance among varieties on the basis of phonetic data; Figs. 3 and 4 show the results of calculating the linguistic distance using morphological data; finally, Figs. 5 and 6 show the linguistic distance among varieties taking into account phonological data only.

To investigate dialect change, we will contrast the oldest and youngest age groups (i.e. F4 and F1). If we look at the MDS plots dealing with these age groups in Figs. 1 and 2, we can observe two noteworthy facts. First, the varieties that, based on the data of older speakers’ pronunciation (F4), were regularly spread on the left side of the plot have undergone a process of homogenization in the younger (F1) speakers, i.e. a gradual reduction of their original differences. However, this new, more homogeneous grouping is still located in Fig. 2 *at some distance* from the standard. Second, the va-

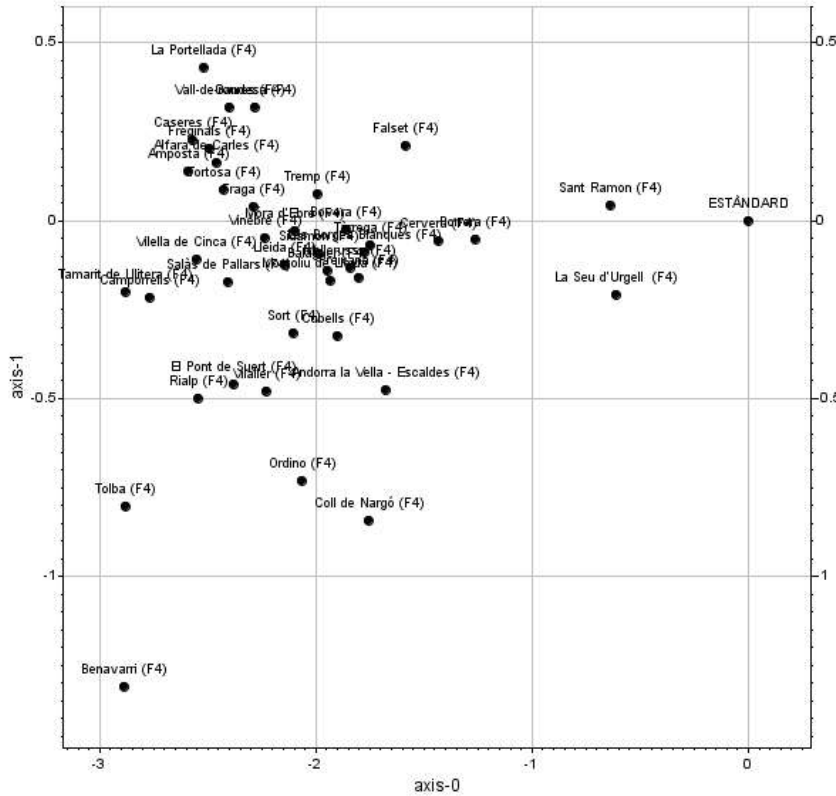


Figure 1: Multidimensional analysis based on the pronunciation of the older speakers (F4). The plot visualizes 52% of variance.

rieties from Aragon (Benavarri, Tolva, Tamarit de Llitera, Camporrells, Fraga, Vilella de Cinca, Vall-de-roures and la Portellada) seem to have remained stable, as they have not substantially moved their positions in the two plots. As a consequence, the linguistic distance between the two groups of varieties located on either side of the political border has increased considerably, resulting in a linguistic boundary where there was a clear dialect continuum previously. There are only three extremely conservative varieties (*ribagorçà*: el Pont de Suert and Vilaller; *pallarès*: Rialp, e.g.; and *tortosí*: Caseres or Tortosa, e.g.) where the impact of the border effect has been less important.

There is a crucial question that arises from Figs. 1 and 2: Why this situation of language levelling among younger speakers does not entail an approach to the standard variety? We hypothesize that this distance is due to the maintenance of the majority of phonological rules that characterize these varieties, a situation that would

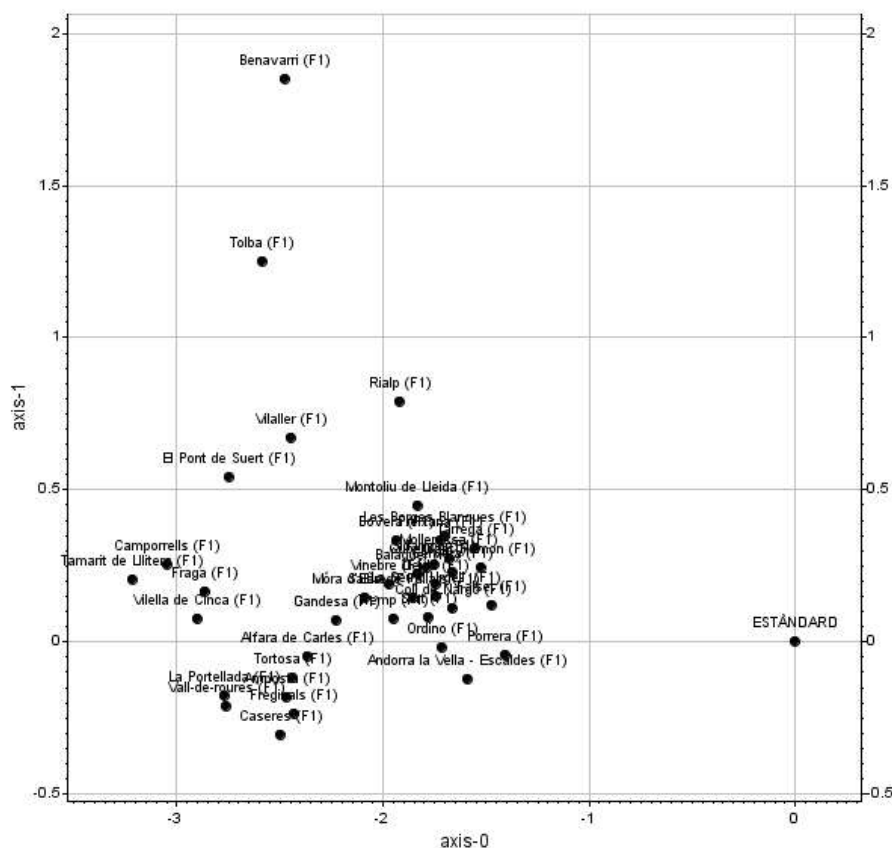


Figure 2: Multidimensional analysis based on the pronunciation of the younger speakers (F1). The plot visualizes 58% of variance.

counteract the strong morphological advergence to the standard undergone by these dialects.

Figs. 3 and 4 should help us prove this hypothesis. These figures display the linguistic distance among varieties on the basis of morphological data. The results are revealing: first, we can see that the morphological distance to the standard is smaller than the phonetic distance both among the F4 and the F1. As a consequence, it is likely to assume that the weight of the phonological rules is more important than the weight of the morphological components in the linguistic distance of Figs. 1 and 2. In fact, the process of *structural dialect loss* (Hinskens, Auer & Kerswill 2005: 11) undergone by most north-western dialects is clearly visible among the F1 speakers in Fig. 4. Their position in the plot indicates that they have massively adopted many standard morphological features. Second, we can see that *tortosí* and *ribagorçà* are

38 Morphology changes faster than phonology

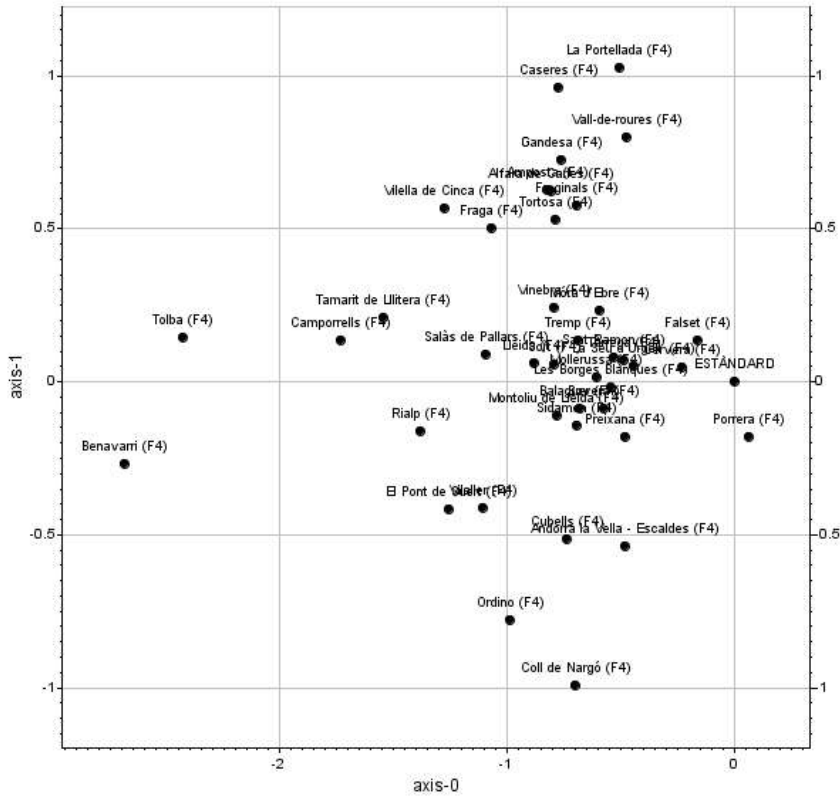


Figure 3: Multidimensional analysis based on the morphological components of the older speakers (F4). The plot visualizes 48% of variance.

the most conservative areas in Catalonia both among the F4 and the F1 speakers; the younger speakers of *pallarsès*, instead, have abandoned most of their morphological peculiarities. Third, the comparison of these two figures shows again the impact of the border effect between Catalonia and Aragon, especially in the central area, where the Catalan varieties have converged more with the standard.

If we agree with the fact that the morphological levelling of the majority of the north-western varieties with the standard has been very important, we will have to agree that the linguistic distance with respect to the standard that could be observed in Figs. 1 and 2 has to be attributed to the maintenance of the phonological rules that characterize these varieties. This hypothesis is confirmed by Figs. 5 and 6, which show the linguistic distance among varieties taking into account only phonological data.

A quick glimpse to the results shows that the levelling undergone by most north-

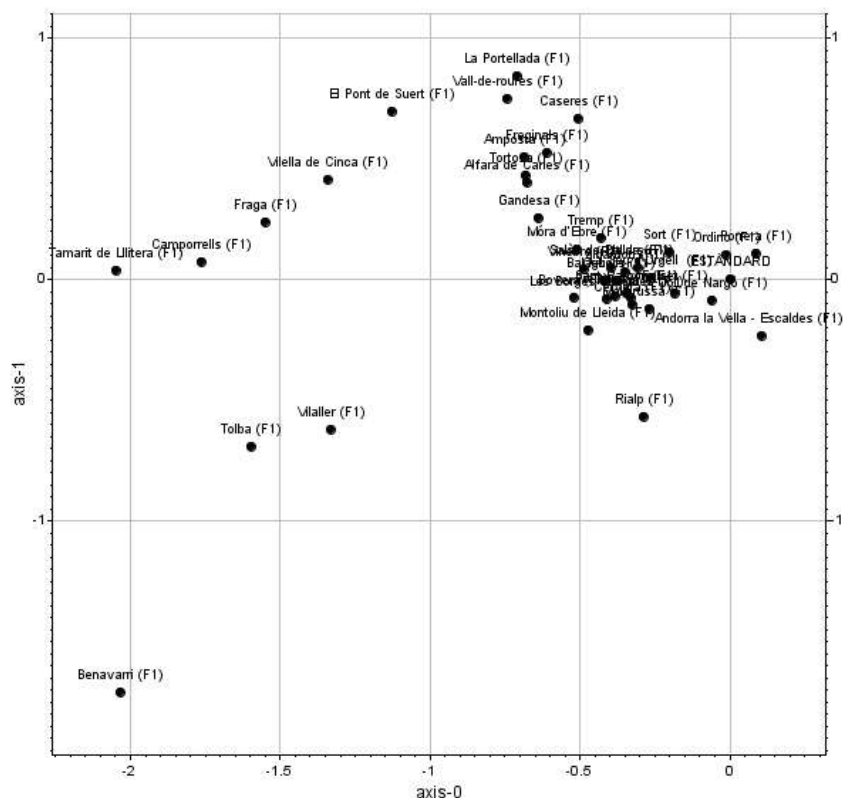


Figure 4: Multidimensional analysis based on the morphological components of the younger speakers (F1). The plot visualizes 58% of variance.

western varieties (with the exception of the most conservative dialects: *ribagorçà*, *tortosí* and the Aragonese varieties) is almost complete. There is an astonishing difference, however, between these plots and those based on morphological data: here the distance between north-western varieties and standard is very important. This fact confirms that nowadays most of these varieties have undergone a process of *accentualization*, as they are currently characterized by a small bunch of phonological rules.

5 Conclusions and prospects

In this paper, we have presented a way to calculate linguistic distance in the morphological and the phonological components of language separately. To do so, we used a dialectometric approach developed at the University of Barcelona. This ap-

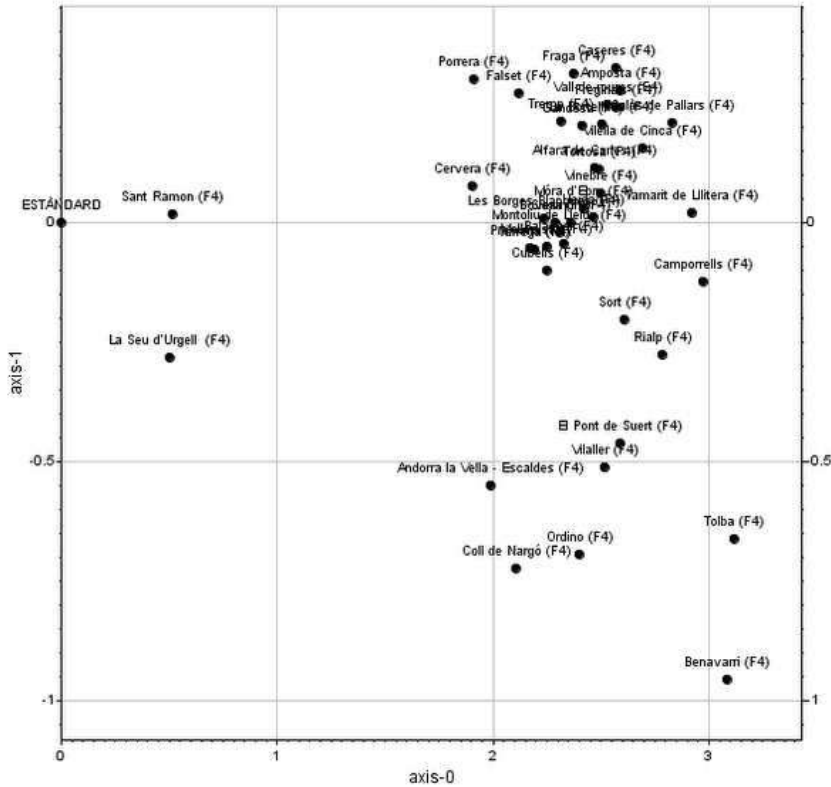


Figure 5: Multidimensional analysis based on the phonological rules of the older speakers (F4). The graph visualizes 55% of variance.

proach claims that it is possible to increase the accuracy of the final groupings by capturing the differences among varieties not only quantitatively but also qualitatively, by means of analysing the underlying differences that remain invisible in the phonetic data. Our purpose, however, has been more theoretical than methodological: we have tried to demonstrate from an aggregate perspective that the process of standardization undergone by most north-western varieties of the Catalan language affects the morphological aspects and the phonological aspects in a different way. We have actually given evidence that, at least in these varieties, morphology tends to level with the standard faster than phonology. This process has resulted in an accentualization of the north-western varieties of Catalonia (with the exception of two conservative dialects), and has allowed us to support the hierarchy of instability of linguistic elements proposed in Viaplana (1999): morphology > phonology. Next step should be to analyse whether this hierarchy is valid in other Catalan varieties and in

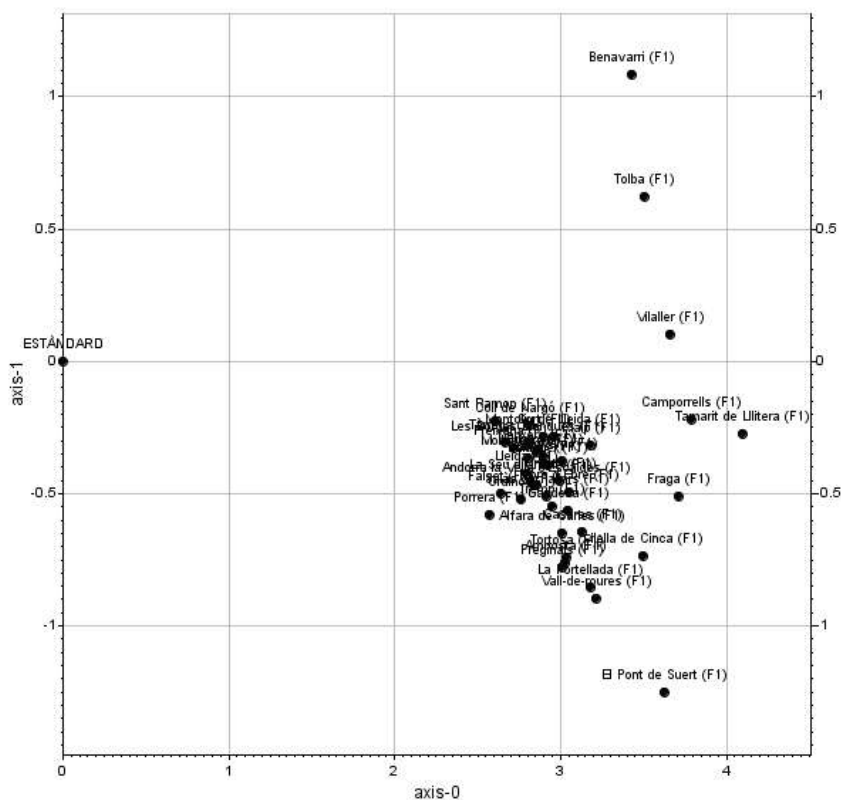


Figure 6: Multidimensional analysis based on the phonological rules of the younger speakers (F1). The graph visualizes 56% of variance.

other languages, but this is something that goes beyond the aims of this contribution.

References

- Berruto, Gaetano. 2005. Dialect/standard convergence, mixing, and models of language contact: the case of Italy. In Peter Auer et al. (eds.), 81–95.
- Chambers, Jack & Peter Trudgill. 1998. *Dialectology*. 2nd edn. Cambridge: Cambridge University Press.
- Clua, Esteve & Maria-Rosa Lloret. 2006. New tendencies in geographical dialectology: the Catalan Corpus Oral Dialectal (COD). In Jean-Pierre Y. Montreuil (ed.), *New perspectives on Romance linguistics*. Vol. 2: *Phonetics, phonology, and dialectology*. Amsterdam / Philadelphia: John Benjamins.

- Hinskens, Frans, Peter Auer & Paul Kerswill. 2005. The study of dialect convergence and divergence: conceptual and methodological considerations. In Peter Auer et al. (eds.), 1–48.
- Sankoff, Gillian. 2002. Linguistic outcomes of language contact. In Jack Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*, 638–668. Malden / Oxford / Victoria: Blackwell.
- Valls, Esteve, John Nerbonne, Jelena Prokić, Martijn Wieling, Esteve Clua & Maria-Rosa Lloret. 2011. Applying the Levenshtein distance to Catalan dialects: a brief comparison of two dialectometric approaches. *Verba. Anuario Galego de Filoloxía* 39. 35–61.
- Valls, Esteve, Martijn Wieling & John Nerbonne. 2013. Linguistic advergence and divergence in north-western Catalan: a dialectometric investigation of dialect leveling and border-effects. *Literary and Linguistic Computing* 28.
- van Bree, Cor. 1985. Structuurverlies en structuurbehoud in het dialect van Haaksbergen en Enschede, een onderzoek naar verschillen in resistentie. *Leuvense Bijdragen* 74. 1–35.
- van Coetsem, Frans. 1988. *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris.
- Viaplana, Joaquim. 1999. *Entre la dialectologia i la lingüística. La distància lingüística entre les varietats del català nord-occidental*. Barcelona: Publicacions de l'Abadia de Montserrat.