

**Linguistic advergence and divergence in north-western Catalan: A dialectometric  
investigation of dialect leveling and border effects**

Esteve Valls\*

University of Barcelona

Martijn Wieling

University of Groningen

John Nerbonne

University of Groningen

\* Corresponding author: Esteve Valls, University of Barcelona, Spain, e.valls@ub.edu

## Abstract

This paper investigates several linguistic changes which are ongoing in north-western Catalan using a contemporary corpus. We take advantage of a range of dialectometric methods that allow us to calculate and analyze the linguistic distance between varieties in apparent time from an aggregate perspective. Specifically, we pay attention to the process of structural dialect loss due to linguistic advergence to standard and eastern Catalan in many north-western Catalan dialects located in Catalonia (Spain) and Andorra. We also provide evidence that the dialect leveling taking 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, where Catalan is not an official language. These opposite sociolinguistic situations (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 varieties, 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 one of the first attempts to study the border effects not only between regions belonging to different countries, but also between different administrative regions *within* the same country. In addition, we investigate the different roles of urban versus rural areas, providing support for the view that the spatial and hierarchical diffusion patterns are complementary.

## 1. Introduction<sup>1</sup>

The point of departure of this study arises in the work of Viaplana (1999), who stated that the north-western Catalan varieties of Catalonia and Andorra were undergoing a gradual process of orientalizacion<sup>2</sup> that might lead to their definitive de-dialectalization:

It seems clear [...] that the north-western varieties are subject to a gradual process of language leveling toward the eastern varieties. [...] As the leveling is deeper for the morphological features than for the phonological ones, the north-western dialects might lose, in the course of this process, their idiosyncratic morphological features and become, as a consequence, mere accentual varieties (Viaplana 1999: 115-116).

This paper investigates several linguistic changes which are now taking place in north-western Catalan, where, according to Viaplana, three different dialect areas can be established depending on their degree of orientalizacion: an (almost) orientalized area, to the east; a transition central area where the dialects have already adopted several features from the eastern dialects; and a more conservative western area (see Fig. 1). This study is innovative because it is the first dialectometric attempt to investigate linguistic change in apparent time using a corpus which contains data from all the north-western dialects, both from urban and rural areas. The corpus allows us to quantitatively study dialect leveling and the effects of borders, both of which have been changing this part of the Catalan-speaking area for some time.

We shall attempt to establish that the present-day situation among the north-western Catalan varieties is marked by a two-way split along the Catalan-Aragonese border, where, moreover, the Andorran varieties group with those in Catalonia. We suggest that this situation has emerged as a result of both standardization efforts in Catalonia on the one hand and linguistic advergence toward the eastern varieties on the

other. Finally, we point to evidence that suggests that both hierarchical diffusion (from more to less populous communities) but also contagious diffusion (from geographically nearby communities) has played a role in the advergence processes.



*Fig. 1 Map of the north-western Catalan domain with the three areas established by Viaplana (1999) according to their degree of orientalizacion: an orientalized area, to the east ('zona orientalitzant'); a central transition area ('zona de transició'); and a more conservative western area ('zona occidentalitzant'). Reprinted from Viaplana (1999: 95).*

In addition, this paper tries to comply with wishes of several researchers who have commented on recent dialectological studies. Gerritsen (1999) points out that too few studies analyze the attrition of dialects due to the influence of a standard. This is a focus of the present paper. Furthermore, we proceed from a contemporary corpus, designed in particular for this purpose, and we attend to the spatial diffusion of standard features among the north-western dialects. Our aim in this is to engage Britain (2002), who finds almost no analyses of this sort in the literature and points out that most dialectal research is based on relatively old corpora. Finally, we try to contribute to the studies on border effects by paying attention not only to the impact of a border between

regions belonging to different countries, but also to a second border between different administrative regions *within* the same country. This is a crucial addition as these types of borders have been scarcely studied at all (Woolhiser 2005).

## 2. Corpus

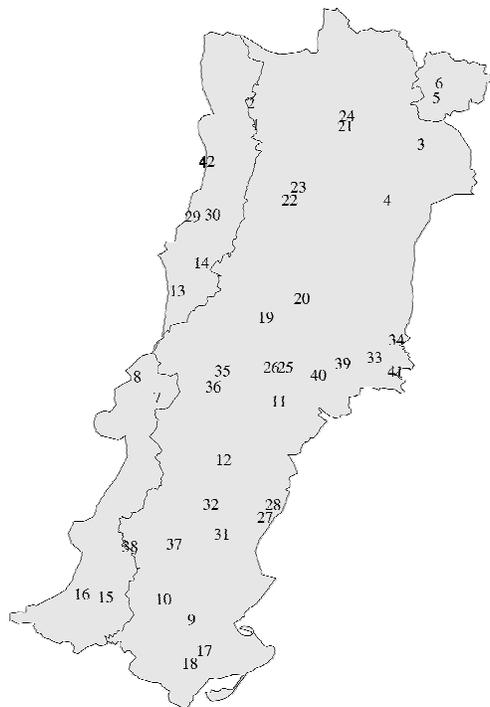
The data set used in this paper was conceived as a corpus of contemporary north-western Catalan (a context map is shown in Fig. 2) and covers the whole area where this dialect is spoken: all of Andorra and two dialect areas within Spain, specifically the western half of the Autonomous Community of Catalonia (with the exception of the Val d'Aran, where Occitan is spoken) and the eastern counties of the Autonomous Community of Aragon. Fieldwork was carried out in 40 villages (2 in Andorra, 8 in Aragon and 30 in Catalonia) located in 20 counties. We added an artificial variety, standard Catalan, to these 40 localities, so in total we examine 41 varieties. Figure 3 shows the distribution of the varieties studied (including the artificial placement of standard Catalan in the East). Table 1 lists all data collection sites together with their counties and regions. The numbers in this table correspond to the numbers in Fig. 3. The standard Catalan variety is the one sanctioned by the *Institut d'Estudis Catalans* (Fabra 1918).

Since we wanted to measure the differences between the urban and the rural areas, half of the interviews were conducted in the most populated localities of each county, i.e. their capitals, while the other half were conducted in small villages from the same counties. Thus, our sample includes 20 urban localities (the populations of which vary from the 1,177 inhabitants of Benavarri to the 137,387 of Lleida, with a mean of 17,787 inhabitants) and 20 rural localities varying from 171 inhabitants in Tolba to 4,396 in Ordino (with a mean of 641 inhabitants). Note that Ordino is regarded here as a

rural area in spite of having more inhabitants than some urban areas, such as Benavarri, because it is the smallest community in Andorra.



*Fig. 2 Context map of the Catalan-speaking area, including some important cities: (1) Perpinyà (France); (2) Andorra la Vella (Andorra); (3) Girona, (4) Barcelona, (5) Tarragona, (6) Lleida (Autonomous Community of Catalonia, Spain); (7) Fraga (Autonomous Community of Aragon, Spain); (8) Castelló de la Plana, (9) València, (10) Alacant (Autonomous Community of the Valencian Country, Spain); (11) Eivissa, (12) Palma, (13) Maó (Autonomous Community of the Balearic Islands, Spain); and (14) l'Alguer (Sardinia, Italy). The area where north-western Catalan is spoken (see Fig. 3) has a darker shade of grey.*



*Fig. 3 Distribution of the 40 varieties (plus standard). The corresponding variety names are shown in Table 1. The internal line almost running north to south separates Catalonia in the east from Aragon in the west. Andorra is located in the north-east corner.*

<b>Region</b>	<b>County</b>	<b>Variety</b>
<i>Andorra</i>	<i>Andorra</i>	<i>Andorra la Vella–Escaldes (5)</i> <i>Ordino (6)</i>
<i>Aragon</i>	<i>El Baix Cinca</i>	<i>Fraga (7)</i> <i>Vilella de Cinca (8)</i>
	<i>La Llitera</i>	<i>Tamarit de Llitera (13)</i> <i>Camporrells (14)</i>
	<i>El Matarranya</i>	<i>Vall-de-roures (15)</i> <i>La Portellada (16)</i>
	<i>La Ribagorça</i>	<i>Benavarri (29)</i> <i>Tolba (30)</i>
<i>Catalonia</i>	<i>L'Alta Ribagorça</i>	<i>El Pont de Suert (1)</i> <i>Vilaller (2)</i>
	<i>L'Alt Urgell</i>	<i>La Seu d'Urgell (3)</i> <i>Coll de Nargó (4)</i>
	<i>El Baix Ebre</i>	<i>Tortosa (9)</i> <i>Alfara de Carles (10)</i>
	<i>Les Garrigues</i>	<i>Les Borges Blanques (11)</i> <i>Bovera (12)</i>
	<i>El Montsià</i>	<i>Amposta (17)</i> <i>Freginals (18)</i>
	<i>La Noguera</i>	<i>Balaguer (19)</i> <i>Cubells (20)</i>
	<i>El Pallars Sobirà</i>	<i>Sort (21)</i> <i>Rialp (24)</i>
	<i>El Pallars Jussà</i>	<i>Tremp (22)</i> <i>Salàs de Pallars (23)</i>
	<i>El Pla d'Urgell</i>	<i>Mollerussa (25)</i> <i>Sidamon (26)</i>
	<i>El Priorat</i>	<i>Falset (27)</i> <i>Porrera (28)</i>
	<i>La Ribera d'Ebre</i>	<i>Móra d'Ebre (31)</i> <i>Vinebre (32)</i>
	<i>La Segarra</i>	<i>Cervera (33)</i> <i>Sant Ramon (34)</i>
	<i>El Segrià</i>	<i>Lleida (35)</i> <i>Montoliu de Lleida (36)</i>
	<i>La Terra Alta</i>	<i>Gandesa (37)</i> <i>Caseres (38)</i>
	<i>L'Urgell</i>	<i>Tàrrrega (39)</i> <i>Preixana (40)</i>
		<b><i>Standard (41)</i></b>

*Table 1. List of all varieties (right) with corresponding regions and counties. The numbers correspond with those in Fig. 3. The first and the second variety of each county are urban and rural, respectively.*

We interviewed 320 informants, 8 per locality, who had to meet the following basic criteria: they had to be middle class citizens, they had to be descendants of parents born in the same locality and they had to have lived there for their entire lives. In a few places this was not possible, and there we used informants whose parents came from nearby villages from the same county. Our informants are divided into 4 age groups: (1) F1: 80 informants, 39 male and 41 female, born between 1991 and 1996 (median year of birth: 1994); (2) F2: 80 informants, 37 male and 43 female, born between 1974 and 1982 (median year of birth: 1979); (3) F3: 80 informants, 36 male and 44 female, born between 1946 and 1960 (median year of birth: 1954); and (4) F4: 80 informants, 63 male and 17 female, born between 1917 and 1930 (median year of birth: 1922). Consequently, 54.6% of the informants are male and 45.4% are female. Women are a bit more numerous than men in F1, F2 and F3, whereas men are much more numerous in F4. The informants are distributed evenly so that there are 2 informants in each of the four age groups for each locality. The first, youngest group was interviewed roughly fourteen years after the other three. We shall explain this in more detail in the next paragraph.

Several facts must be kept in mind when examining these age groups. First, the data corresponding to the oldest three age groups were gathered by one fieldworker between 1995 and 1996, whereas the youngest age cohort was interviewed between 2008 and 2011 by a second fieldworker (the first author of this paper, who also transcribed and processed the entire corpus). There is also a greater gap between the median birth year of ages F2, F3 and F4 (25 and 32 years) than between F1 and F2 (only 15 years). Although a more evenly distributed set of birth years would have had its benefits, two reasons led us to select the youngest age group as we did. On the one hand we wished to interview adolescents who were in the last two years of secondary

(compulsory) school. This was a crucial criterion because both F1 and F2 were about to finish their compulsory schooling *in Catalan* at the time of their interviews, whereas F3 and F4 had been taught exclusively *in Spanish*, and we expected the language of education to influence the degree of standardization of their dialects. On the other hand, we assume that children's speech does not become stable until late adolescence (Bailey 2002: 320) or early adulthood (Nakhola & Saanilahti 2004: 87). Working with younger speakers (which would have resulted in a more even distribution of birth years) would have been problematic as we might have tapped into unstable idiolects and drawn ill-founded conclusions.

A third factor to keep in mind is that the data were collected through a questionnaire of 712 glosses plus the recordings of approximately fifteen-minute samples of informal speech. From this questionnaire, which included phonetic data, morphological data, lexical data and syntactic data, we selected a subset of the same 363 glosses per informant. These are distributed in 8 morphological categories: articles (16 items), clitic pronouns (81 items), demonstrative pronouns (12 items), neuter pronouns (3 items), locative adverbs (3 items), verbs (220 items), possessive pronouns (20 items) and personal pronouns (8 items). The items are listed in Appendix A. As can be observed, our corpus is mainly based on verbal and nominal inflections, i.e. on common features of the language.

As Appendix A shows, our corpus includes elements from the five regular paradigms of the three verbal classes in Catalan<sup>3</sup> into account: *cantar* 'to sing' (I = first conjugation verbs), *perdre* 'to lose' (II[-EXT] = second conjugation verbs without an infix), *beure* 'to drink' (II[+EXT] = second conjugation verbs with an infix), *sentir* 'to hear' (III[-EXT] = third conjugation verbs without an infix) and *servir* 'to serve' (III[+EXT] = third conjugation verbs with an infix). This classification of paradigms is

exclusively based on the distribution of some morphological features that appear regularly on all the verbs of a class: the so-called inflections. For instance, the presence or absence of the palatal extension [éjʃ] in some verbal tenses and persons is crucial to differentiating the verbs of the third class: those belonging to III[-EXT] will never contain the extension [éjʃ], whereas those belonging to III[+EXT] will regularly show this extension in PI 1, 2, 3 [-PLU], 3 [+PLU]; PS 1, 2, 3 [-PLU], 3 [+PLU]; and IMP 2, 3 [-PLU], 3 [+PLU] as illustrated in Table 2.<sup>4</sup>

Several methods were employed to elicit the dialectal pronunciations, but only two of them were used in the subset: for the verbs, the informants were given a sentence with a gap that they had to fill in with the correct verbal tense; for the remaining words, the informants were asked to translate sentences from Spanish into their own dialects.

The final corpus contained 113,749 items (i.e. one item represents a specific pronunciation of one speaker for one gloss, such as [séŋto] ‘I hear’) and, as every pronunciation consisted of about six sound segments (e.g., [o]), a total of 680,639 sound segments. Note that the transcription alphabet consisted of 35 unique sound segments.

In order to show the clearest effect of age, we will focus in the presentation of our results on the youngest (F1) and oldest speakers (F4), but note the pattern of F2 and F3 is intermediate between F1 and F4 (but not shown).

	Present Indicative (PI)		Present Subjunctive (PS)		Imperative (IMP)	
	III [-EXT]	III [+EXT]	III [-EXT]	III [+EXT]	III [-EXT]	III [+EXT]
1[-PLU]	séŋto	serβe <b>j</b> so	séŋti	serβe <b>j</b> si		
2[-PLU]	séns	serβe <b>j</b> ses	séŋtis	serβe <b>j</b> sis	sén	serβe <b>j</b> ʃ
3[-PLU]	sén	serβe <b>j</b> ʃ	séŋti	serβe <b>j</b> si	séŋti	serβe <b>j</b> si
1[+PLU]	seŋtím	serβím	seŋtím	serβím	seŋtím	serβím
2[+PLU]	seŋtíw	serβíw	seŋtíw	serβíw	seŋtíw	serβíw
3[+PLU]	séŋten	serβe <b>j</b> ʃen	séŋtin	serβe <b>j</b> ʃin	séŋtin	serβe <b>j</b> ʃin

Table 2. The verbs sentir ‘to hear’ and servir ‘to serve’ conjugated in PI, PS and IMP. Note that the palatal extension [éjʃ] (in bold) appears only in the verbs of the III[+EXT] verbal class.

### 3. Methods

#### 3.1 Obtaining aggregate differences using the Levenshtein distance

The Levenshtein distance (Levenshtein 1965, also known as *edit distance*) is a string comparison procedure that calculates the distance between two phonetic strings. To obtain this distance, the Levenshtein algorithm seeks the least costly set of basic operations (insertions, deletions and substitutions) needed to transform one string into another. In the simplest version of the algorithm, these three operations have the same cost, as can be seen in the example below, based on two pronunciations of a conjugated form of the Catalan verb *servir* ‘to serve’, specifically *servís* ‘(if I) served’. In this case, the final distance between the two pronunciations is 3:

(1)

Variety 1	s e r β i s k é s	delete s	1
	s e r β i k é s	substitute k/γ	1
	s e r β i γ é s	insert ε	1
Variety 2	s e r β i γ é s ε		
<hr style="border: 0.5px solid black;"/>			
Total			3

From a different perspective, the procedure can also be seen as the result of aligning two strings of phonetic segments. In these alignments, phonetic overlap is binary, so that non-identical phones contribute to phonetic distance, whereas identical ones do not. In order to increase accuracy, we used a common modification of the Levenshtein algorithm, not allowing alignments of vowels with (non-sonorant) consonants. The following example illustrates the alignment of the two pronunciations compared in (1):

(2)

Variety 1	s	e	r	β	i	s	k	é	s	
Variety 2	s	e	r	β	i		γ	é	s	ε
<hr style="border: 0.5px solid black;"/>										
							1	1		1

The total distance between two pronunciations is subsequently divided by the alignment length in order to treat every segment the same. In the above example, the normalized distance would be 0.3 (3 divided by 10).

While there have been methods developed which obtain more sensitive word pronunciation distances (by using linguistically sensitive sound segment distances; Wieling *et al.*, 2012), this does not appear to influence results at an aggregate level (which we investigate here) greatly (Wieling, 2012). Consequently, we use the simple Levenshtein distance here.

In calculating the linguistic distances of the verbs, we took only the inflections into consideration, and we deleted all stems. We made this decision because the inflections are morphological features that reappear regularly in *all* Catalan verbs, but the stems merely contain the idiosyncratic lexical information of *one* verb required to differentiate one verb from another in the lexicon. If we had included them, we would have weighted them too heavily, as they appeared 44 times in each conjugated verb. Working only with the inflections had a drawback, however, as differences in inflections might give rise to higher normalized distances than with the base form included (e.g., when an inflection would completely differ, this would result in the maximum normalized distance of 1.0, which would obviously be too high).. We therefore corrected the length of these words by adding a constant four-segment stem to all inflections and took these into account in calculating the phonetic distances (in this case, when the inflection would completely differ and occupy 4 positions in the alignment, the equal stem would result in a normalized distance of 0.5, as 4 positions in

the alignment are equal and 4 positions are different). This is a novel step in measuring pronunciation differences.<sup>5</sup>

### 3.2 Techniques to visualize aggregate distances<sup>6</sup>

#### 3.2.1 Reference point maps

The reference point maps (RPM) display the linguistic distance between a reference point and all the other varieties investigated. After a reference point is selected, the other points are coloured lighter or darker depending on their distance with respect to the reference point (the more distant, the darker the colour).

This visualization technique (see Fig. 7a and 7b) has been used extensively by Goebel and is also available in the programme *Visual Dialectometry* (VDM) developed at the University of Salzburg. As Goebel (1991: 285) points out, the main scope of the RPM (known as *similarity maps* in Salzburg):

[...] lies –at least for Romance linguistics– in giving a heuristically comprehensive answer to the one question asked for a hundred years in many different ways about the *position of a local dialect in its geolinguistic environment* [emphasis in original].

#### 3.2.2 Stable clustering: probabilistic dendrograms and maps

Clustering is a well-known procedure to identify groups of close varieties, and has been used in dialectometry since Shaw (1974). Hierarchical agglomerative clustering is an iterative procedure that selects the shortest distance in a matrix and fuses the two corresponding data points. As these two points form a new cluster, the distance between this cluster and the remaining elements in the matrix is recalculated. In the end, hierarchical clustering produces a hierarchically structured dendrogram. The clustering algorithm employed in this study used was a combination of the so-called UPGMA

(Unweighted Pair Group Method with Arithmetic Mean) and the WPGMA (Weighted Pair Group Method with Arithmetic Mean) as those are the more reliable clustering algorithms (Prokić & Nerbonne 2008).

Although the use of regular clustering has become more and more popular among linguists interested in dialectometry, it is also broadly accepted that it lacks stability. This is caused by the fact that clustering procedures look for the minimum distance between two points in a matrix, and sometimes several pairs of elements may show similar distances. As a consequence, small differences in the input data matrix can lead to considerably different clusters (Prokić & Nerbonne 2008).

To overcome this instability, two methods have been suggested and tested during the last years, *noisy clustering* (Kleiweg *et al.* 2004) and *bootstrapping* (Nerbonne *et al.* 2008). We applied noisy clustering to our data set to obtain robust clustering results. Briefly, noisy clustering can be viewed as a procedure in which different small amounts of random noise (i.e. the increase or decrease of the aggregate pronunciation distances between two sites with a small random value lower than a certain threshold, in our case 0.2) are added to the distance matrix before clustering. Every distance matrix modified in this way is then used to calculate a new clustering. Only clusters which are observed in many of these runs, are judged to be robust. Bootstrapping is similar, but consists of varying the input data set (instead of the distance matrix) in several clustering iterations, allowing some words to be repeated, while others are deleted. Consequently, the set of words used to determine the aggregate pronunciation distances differs somewhat in each run, resulting in different distance matrices. Each distance matrix is then used (similar to noisy clustering) to obtain a new clustering, and only clusters which are observed in many of these runs are judged to be

robust. The result of both techniques is a consensus (or probabilistic) dendrogram, which can be projected onto a probabilistic map (see Fig. 6a and 6b).

### 3.2.3 *Multidimensional Scaling (MDS)*

A very suitable alternative technique to investigate dialect distances is multidimensional scaling (MDS), which aims at reducing a large distance matrix containing many dimensions (i.e. in our case as many as there are varieties) to only a few dimensions (e.g. 2 or 3). Having only a few dimensions is highly beneficial as two or three dimensional data can be visualized easily. In most cases using only 2 or 3 dimensions is already sufficient to explain most of the variation in the original distance matrix. MDS was first introduced to linguistics by Black (1976), who measured the distances among several dialects of four linguistic groups, located in the Philippines, Africa and North America. There are two reasons for using MDS. First, it is a stable method to analyze the linguistic distances, unlike normal clustering, and second, it provides us with the possibility of examining the relations between varieties in more detail than by using a probabilistic dendrogram (see Prokić and Nerbonne 2008: 163ff). To visualize the MDS results we will use 2-dimensional MDS plots (see Fig. 4 and 5).

## **4. Results**

Section 4.1 establishes that advergence is taking place on the eastern side of the Catalan-Aragonese border by examining the varieties in apparent time, i.e. as they are spoken by the oldest and the youngest age group. Section 4.2 then considers the role standardization is playing. We maintain that both advergence and standardization have been influential. In Section 4.3 we examine the diverse paths along which diffusion has progressed, and in the final Section (4.4) we argue that Aragon's independent

development may be appreciated in the distinctive role of analogy in the development of the verbal paradigms, which are developing in a novel direction in the west.

#### *4.1 Linguistic advergence and divergence: the two sides of the border effect*

To investigate dialect change, we will contrast the oldest and youngest age groups (F4 and F1). If we look at Fig. 4 and 5, the MDS plots dealing with these age groups, we can observe two remarkable facts. First, varieties which were regularly spread on the right side of the plot based on the data of older speakers' pronunciation (F4) have undergone a process of homogenization in the younger (F1) speakers, i.e. a gradual reduction of their original differences.<sup>7</sup> This new, more homogeneous grouping is located on the right of the plot in Fig. 5, at some distance from the standard. Second, the varieties from Aragon seem to have remained stable, as they have not moved their positions substantially in the two plots.

Another way to approach this dialect leveling process is to investigate Fig. 6a and 6b, which visualize the probabilistic clustering results of the older speakers (Fig. 6a) and younger speakers (Fig. 6b) on the map. Similar colours in these maps indicate that these varieties were clustered frequently in the same cluster. These maps clearly show that in the past it was much easier to identify a speaker's provenance based on how he or she spoke. For instance (see Fig. 6a), it was possible to know if they spoke *tortosí* (in blue, to the south), *lleidatà* (in light blue, in the central area), *pallarès* or *ribagorçà* (in green, to the north), or some sort of "Aragonese central dialect", with a clear distinction between Benavarri and Tolba (in violet) and the other Aragonese varieties (in dark green). Benavarri and Tolba always appear very isolated from the others due to their proximity to the transition area between Catalan and Aragonese languages, and probably also due to the fact that they have several *phonemic isoglosses*;

they do not have, for example, voiced fricatives and affricates in their phonemic inventories. Figure 6b, in contrast, shows that these distinctions cannot be made anymore *from an aggregate perspective*.<sup>8</sup> The differences must either be infrequent, or they are counterbalanced by similarities. This is due to the fact that most of the northern dialects and a few southern dialects have converged toward (or have already merged with) the central varieties. As a consequence, it is much more difficult to know if a speaker comes from one or another traditionally different linguistic area.

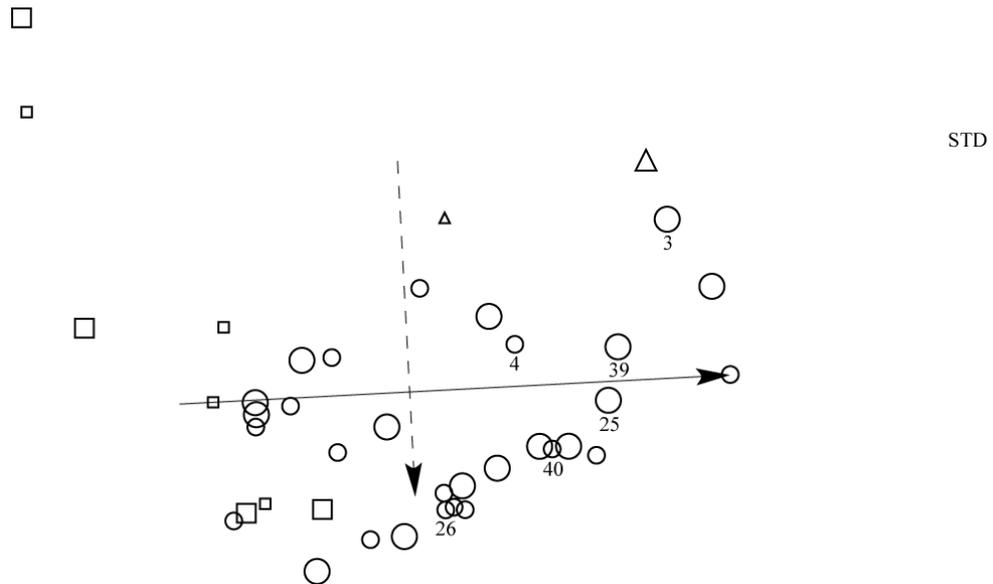
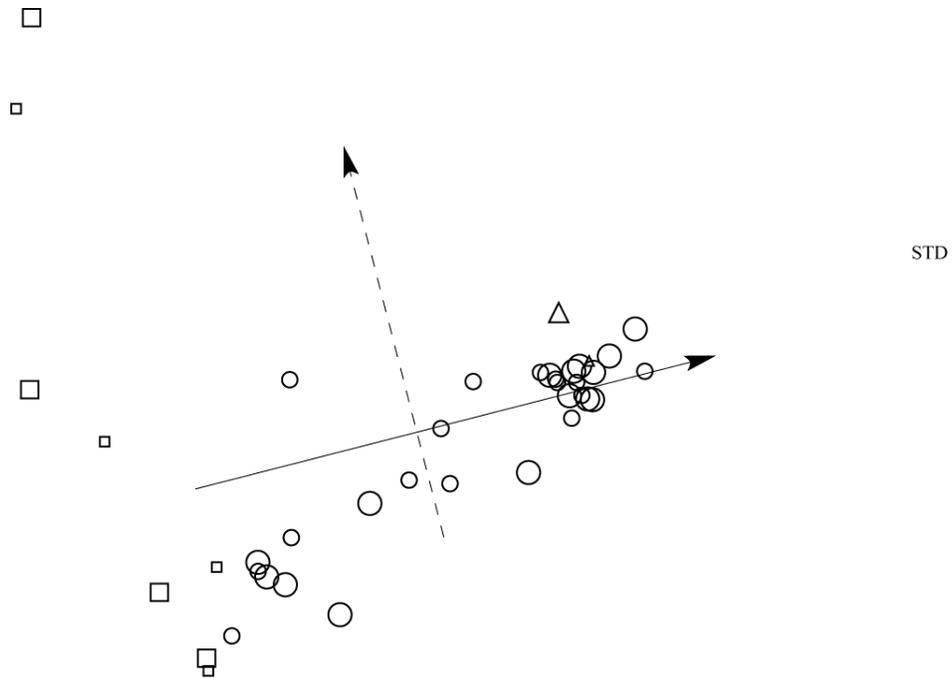
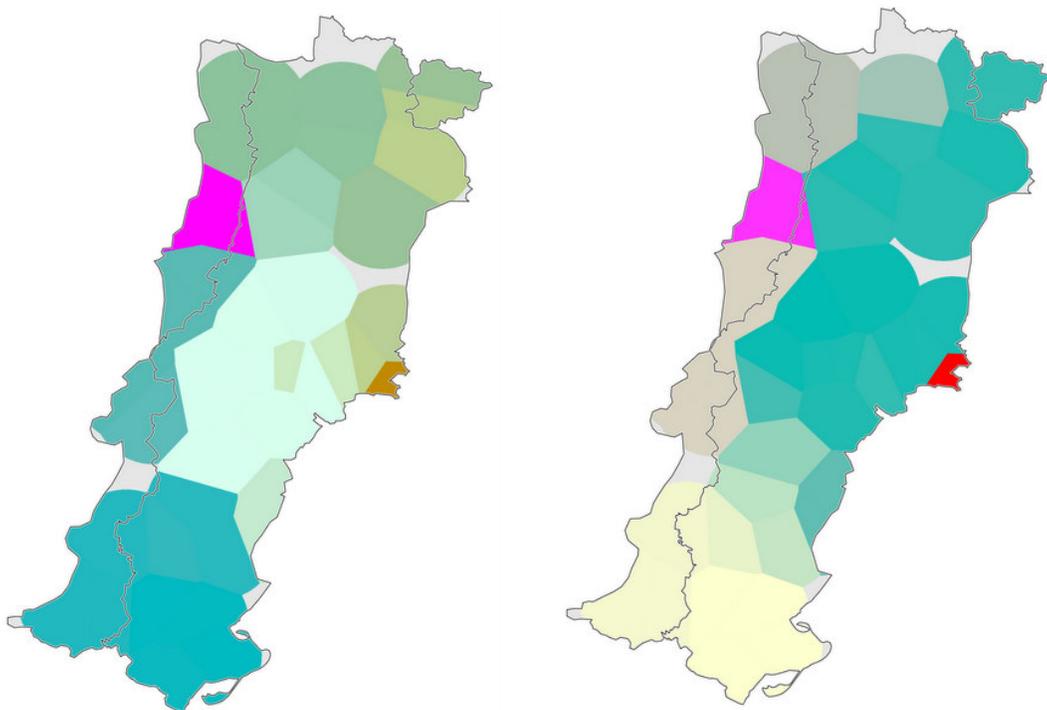


Fig. 4 MDS plot based on the pronunciations of the oldest speakers (F4). The circles indicate Catalan communities, and the squares the Aragonese communities. The two communities in Andorra are marked by a triangle. Urban communities are represented by large symbols, while small symbols represent rural communities. STD indicates the location of the standard language. The numbers correspond to those in Table 1 and illustrate that urban communities (numbers 3, 25 and 39) are closer to the standard than their rural counterparts (numbers 4, 26 and 40). The MDS plot visualizes 86% of the variance.



*Fig. 5 MDS plot on the basis of the pronunciations of the youngest speakers (F1). The circles indicate Catalan communities, and the squares the Aragonese communities. The two communities in Andorra are marked by triangles. Urban communities are represented by large symbols, while small symbols represent rural communities. STD indicates the location of the standard language. The MDS plot visualizes 87% of the variance. Comparing this to Fig. 4, we note that many more varieties are closer to the standard; in particular, the Catalan communities lie closer to the standard than those of Aragon.*



*Fig. 6a and 6b Probabilistic maps on the basis of the pronunciations of the oldest speakers (F4; Fig. 6a, left) and the youngest speakers (F1; Fig. 6b, right). We added 20% of noise.*

So far we have seen that most of north-western dialects are undergoing a process of leveling that entails a reduction of their most marked differentiating features. This situation confirms the directionality of the linguistic change pointed out by Viaplana (1999): it is now no longer possible to divide these varieties into three different areas (orientalized, transition and conservative) depending on their degree of convergence towards the eastern dialects, because the three areas have regrouped and coalesced into only two: a conservative area (to the west) and an orientalized area which now includes the so-called “transition area”.

Strictly speaking, however, Fig. 4, 5, 6a and 6b depict only a situation of language leveling *among* several north-western varieties, which are still located *at some distance* from the standard. To clarify whether this process also entails an approximation *to the standard*, we take advantage of a third useful technique to visualize aggregate distances: the reference point maps. Fig. 7a and 7b display the linguistic distance between the north-western varieties and the standard, which has been selected as the reference point. The darker a locality is coloured, the less distant it is with respect to the standard reference point. The evolution shown by Fig. 7a (corresponding to the oldest speakers) and 7b (corresponding to the youngest speakers) is revealing: the process of leveling does not stem from the *integration* of features of several north-western varieties into a new compromise dialect, but is due to a process of vertical advergence to the standard which has taken place *during the 20th century*. We can see that the progressive spreading of the standard amongst the population and its contact with the north-western dialectal varieties has provoked a gradual reduction of the differences amongst these dialects, resulting in what Hinskens *et al.* (2005: 11) call *structural dialect loss*. Speakers, therefore, tend to replace some specific features of their dialects with the standard forms and use them in their interactions with the other

speakers of the same dialect. This is not the same as a *dialect shift*, as speakers do not completely give up using their dialect in favour of the standard variety, but rather tend to gradually incorporate more and more features of the standard in their speech.

On the other hand, internal convergence always entails divergence at the borders (see, for example, Sapir 1921: 213), as Fig. 7a and 7b show: the advanced process of advergence towards the standard undergone by most north-western varieties contrasts with the stability of the dialects in Aragon. 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 previously there was a clear dialect continuum. There are only two extremely conservative varieties (*ribagorçà*, to the north, and *tortosí*, to the south) where the impact of the border effect has been less important.

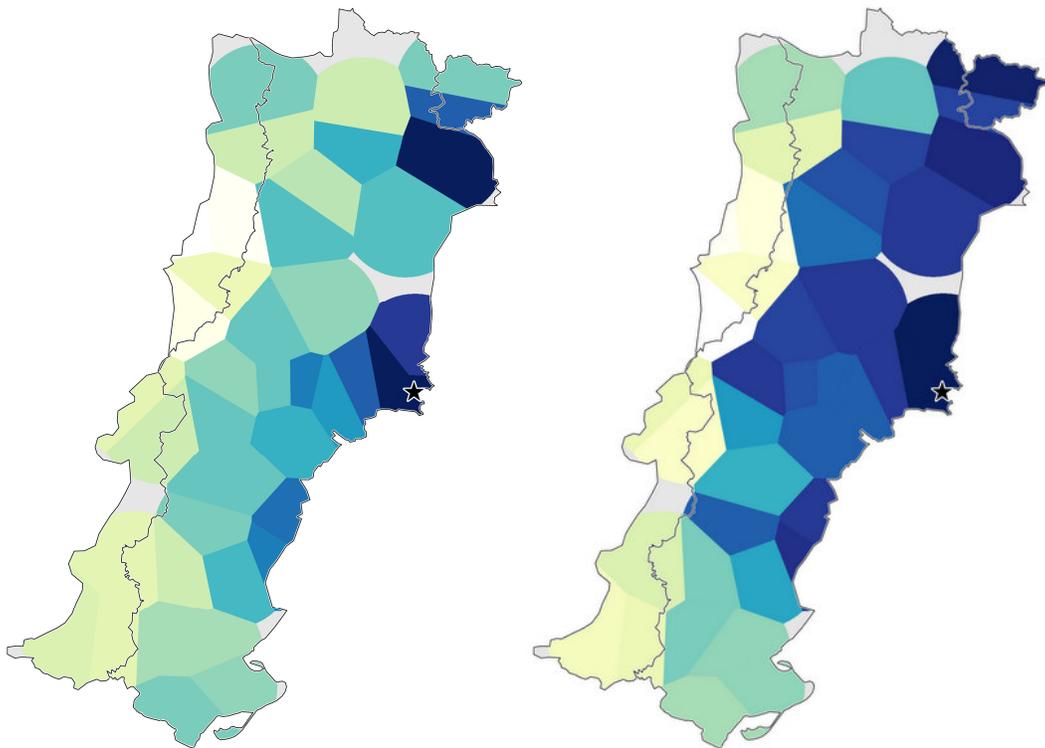


Fig. 7a and 7b Reference point maps with respect to standard Catalan for the older speakers (F4; Fig. 7a, left) and the younger speakers (F1; Fig. 7b; right). The map on the right shows that the eastern and central dialects have become more standard-like.

This concludes our argument that the present-day situation in north-western Catalan is characterized by a two-way division into the eastern, Catalan varieties together with Andorra on the one hand and the western, Aragonese varieties on the other. We discuss border effects which other researchers have detected before closing this section.

Some authors have already paid attention to two other border effects which have arisen in the Catalan-speaking area. Bibiloni (2002: 5), for instance, refers to the lexical divergence of the varieties located on either side of the Spanish-French border. From his perspective, the expansion of France after 1659 to a few counties that had previously been part of Catalonia changed the Catalan dialects spoken in these counties, as they systematically borrowed new words from French. In contrast, all dialects located to the south of the state border have borrowed new words from Spanish, which has naturally resulted in a remarkable increase of the lexical distance between the Catalan varieties spoken in the southern corner of France and those in Spain.

A second border effect has attracted more attention, namely the process of linguistic divergence taking place between the Autonomous Communities of Catalonia and the Valencian area. During the last decades, and especially since 1998, an alternative standard of Catalan based on the most general features of the Valencian central dialects is being supported by the regional government. As a consequence, some authors argue, northern Valencian dialects are slowly converging towards this alternative standard. Some sociolinguists, such as Virgili (1992: 559) and Pradilla (2008a, b), have warned that, if this border effect increases in the future, some groups might use this development to argue that Valencian and Catalan are separate languages, despite the fact that there is currently consensus among linguists about their being different dialectal varieties of Catalan.

To some extent these facts corroborate Auer and Hinskens (1996: 17), who point out that “vertical convergence towards different standard varieties is likely to be the primary source of horizontal divergence from similar dialects across the border”. However, the situation also differs somewhat in the Catalan area we investigate, since vertical advergence occurs only in Catalonia and Andorra, whereas dialects remain relatively stable in Aragon. The first group of dialects is thus converging vertically towards the standard and diverging horizontally at the same time from another stable group of dialects.<sup>9</sup>

To summarize the most important result of this section:, we provide evidence that linguistic divergence has increased during the last 80 years between Catalonia and Aragon.

#### *4.2 Orientalization or standardization?*

We argue in this subsection that the developments we note above are due both to the advergence of the Catalan dialects toward each other as well as to recent efforts in standardization in Catalonia and Andorra.

At the beginning of this paper we mentioned the work of Viaplana (1999), who stated that most north-western varieties were undergoing a process of de-dialectalization, of progressive “language leveling *toward the eastern varieties*” (Viaplana 1999: 115 [our italics]). Viaplana showed that the linguistic innovations were moving westward “as a shock wave” from the “orientalized area” (closer to the eastern dialects and, therefore, to the capital, Barcelona) to the more distant “conservative area”. Hence, the diffusion pattern of linguistic change depicted by Viaplana suggests that the advergence of the north-western varieties must be due to factors such as mobility and interpersonal contact involving the eastern (more prestigious) dialects.

These were the main factors in the process of linguistic change which preceded the standardization of Catalan. Viaplana thus agrees with other linguists that the influence of the standard languages on the dialects must have been slight at least until the 20th century:

The standard-language concept is relatively young and mass literacy is a twentieth-century attainment. Therefore, cross-dialectal levelling must be the older, and historically, the main and probably only type of convergence. As 'late' as 1914, Terracher's investigations of the dialects of the Angoulême area brought to light that 'l'agent destructeur de la morphologie des patois n'est pas le français, mais les parlers limitrophes' (Pop 1950: 106 [cited by Hinskens *et al.* 2005: 27]).

Regarding Catalan, some authors have already proposed that features now considered characteristic of one dialect arose as its speakers used the eastern prestigious varieties as models. Recasens (1996), for instance, thinks that this is the reason for the appearance of [ɛ] as the most typical final post-stressed vowel in *lleidatà*:

This change is an attempt to approximate the low variant /a/ to the prestigious variant [ə] of the eastern varieties; the new pronunciation might have spread from Lleida and/or areas and county capitals of the north-western dialect which are close to the border with the eastern area [...]. The absence of the vowel [ə] in the non-stressed vowel system of the north-western dialects might explain the change from [a]/[ɑ] to [ɛ], as [ɛ] is located between the original low variant and the prestigious [ə] (Recasens 1996: 96-97).

Although the standardization of Catalan is relatively recent (Fabra's *Gramàtica Catalana* dates only to 1918 and his *Diccionari general de la llengua catalana* was not published until 1932), and despite the fact that Franco's regime suspended this process for more than 40 years, it is plausible to think that almost one century of diffusion of

the Catalan grammar must have influenced its geolects to some extent, especially since Catalan became official again when the current language policy was implemented during the democratic transition in the early 1980's.<sup>10</sup>

The works of Carrera-Sabaté (2002, 2003 and 2006) and Carrera-Sabaté & Freixenet-Esteve (2003) confirm that geolects are indeed adopting standard features. On the one hand, these authors argue that the variation in 1st person singular and the 3rd person singular of the Conditional, the Present Indicative and the Imperfect Indicative is being neutralized under the influence of the written language. They state that the traditional pronunciations of [ɛ] in the first person singular ('jo [kaŋtáβɛ]', '*I sang*') and [e] for the third person singular ('ell/ella [kaŋtáβe]', '*He/she sang*') are being neutralized in favor of [ɛ] by some speakers of the north-western capital Lleida, particularly in formal contexts. The neutralization accords with the written language ('jo cantava', 'ell cantava'), as final non-stressed /a/ is regularly pronounced [ɛ] in this dialect. Although they admit that this change is only incipient in informal contexts, they state that "the process of change depicted here has allowed us to detect again the progressive imposition in the area of a prestigious model for oral language which is directly related to the written language" (Carrera-Sabaté & Freixenet-Esteve 2003: 9). In another paper one of the co-authors reaches the conclusion that the language models taught at school might be favoring the birth of an oral standard Catalan to the detriment of the regional dialects:

This process of change has allowed me to detect the speakers' tendency to match the phonic forms of speech to written models. [...] Regarding written language as prestigious to the detriment of oral language is an increasingly popular attitude. It is becoming widespread across generations and affects the youngest

speakers, who, with the help of their schooling in Catalan, are losing the dialectal richness they are entitled to as linguistic heirs of a speech community (Carrera-Sabaté 2006: 17).

In another paper (Carrera-Sabaté 2002), she suggests an explanation for the substitution of [a] for [e] in initial non-stressed vowels in words like *escola* ‘school’ or *embenar* ‘to bandage’. These words would have traditionally been pronounced with initial [a] in north-western Catalan: [askólɛ], [ambená]. However, they are written with initial <e> in standard Catalan (*escola*, *embenar*) so that many children now pronounce [eskólɛ], [embená]. The results of her research led the author to state that “ce processus de changement n’est pas influencé par les dialectes orientaux, puisque la nouvelle voyelle [e] est plus éloignée de la voyelle qu’utilisent les locuteurs des dialectes orientaux en position prétonique [ə]” (Carrera-Sabaté 2002: 12).

The situation we have just described forces us to interpret the process of linguistic change undergone by the north-western varieties in a more complex way. Now, it is no longer possible to explain it in terms of *orientalization* or *standardization*, because it stems from the combination of these two processes of advergence. Hinskens *et al.* (2005: 27) raise several crucial questions about it:

Of course, the question arises as to whether the present-day levelling of cross-dialectal variation occurs completely independently of the standard language. What may seem to be purely cross-dialectal levelling may be motivated by the fact that the dialect converged towards is perceived as being (and may in fact be) closer to the standard variety by the speakers of the converging dialect. [...] Most of the situations in which there appears to be an independent cross-dialectal dimension concern speech communities where, alongside a range of dialects, a prestigious standard language is in common use. Is cross-dialectal convergence sociolinguistically independent of the standard language in these communities? Would this

type of dynamics also have occurred if there had been either no standard language or another standard language?

Fig. 8a, 8b, 9a and 9b might help to clarify the role played by the standard, on the one hand, and the eastern varieties, on the other hand, in this process of de-dialectalization of the north-western dialects. Fig. 8a and 8b show the distribution of the 1st person plural accusative clitic 'mos' in the context  $\_#V$ : 'ens esperen' [moz#aspéren] 'they are waiting for us'. Fig. 8a corresponds to the F4 (oldest speakers) and Fig. 8b corresponds to F1 (youngest speakers). The darker the blue, the more frequent the dialectal variant 'mos' is in a place. On the contrary, the lighter the blue, the more frequent the standard variant 'ens' is in a locality. The comparison of Fig. 8a and 8b shows two things: first, the gradual diffusion westward of the prestigious variant 'ens' (shared by the eastern and the standard varieties); and second, that this prestigious variant also appears quite often among the oldest speakers of the eastern counties of the north-western Catalan area. The regression of the dialect variant 'mos' must have started, therefore, *before* the diffusion of the standard promoted by the institutions of Catalonia and Andorra. This is exactly the same situation we find in Fig. 9a and 9b, which show the distribution of the north-western masculine singular article 'lo' in the context  $\_#C$ : 'el cap' [lo#káp] 'the head'. Fig. 9a corresponds again to the oldest speakers (F4) whereas Fig. 9b corresponds to the youngest speakers (F1). Again, the standard variant 'el' is the same for the eastern dialects.

The replacement of the traditional dialect variants (Veny 1982: 95) by the standard/eastern variants among the oldest speakers provides evidence that the orientalization process has been *historically* important. However, is it possible to

maintain that this cross-dialectal convergence is independent of the diffusion of the standard language which has taken place during the 20th century?

Fig. 10a and 10b suggest that we might dismiss this hypothesis. These figures show the distribution of the 3d person reflexive clitic in the context *w#\_* as it was traditionally pronounced in north-western Catalan: ‘arregleu-vos’ [areɣléws] (and, in a few places, [areɣléwtos]) ‘get yourself ready (to go out, for instance)’. If the prestigious variant adopted was always the eastern one, we would expect an evolution from [areɣléws] (the north-western pronunciation) to [areɣléwze], where the speakers would adopt the eastern clitic ‘-se’ and would adapt it phonetically to their dialect (from the eastern pronunciation [əɾəɣléwzə] to the north-western pronunciation [areɣléwze]). Fig. 10a and 10b, however, show that the adopted variant is ‘-vos’, the standard form which also competes with the eastern traditional variant ‘-se’ in the eastern dialects. This means that in the last decades the process of orientalizacion has been accompanied by a process of standardization which is also reducing the differences between the oral and the written language (Carrera-Sabaté 2006) and which might also be changing the perception that the eastern dialects are the most prestigious ones in favor of the standard. Thus, in at least some instances standardization is the more important dynamic, clearly more influential than “orientalizacion”.

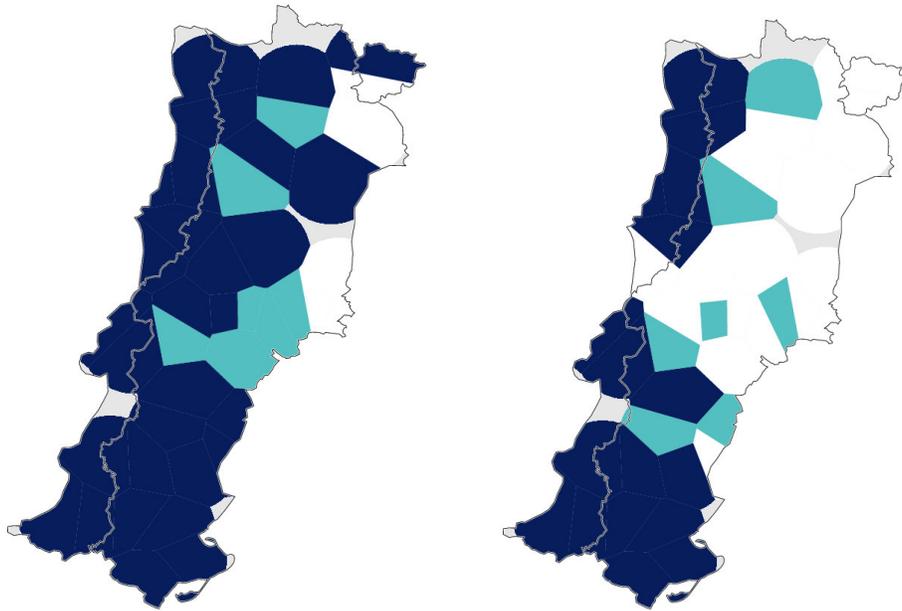


Fig. 8a and 8b Distribution maps of the 1st person plural accusative clitic 'ens' (context:  $\_#V$ , 'ens esperen', 'they are waiting for us') according to the pronunciations of the oldest speakers (F4: Fig. 8a, left) and the youngest speakers (F1: Fig. 8b; right). The dark blue indicates sites only having the dialectal variant 'mos' (in standard/eastern varieties 'ens').

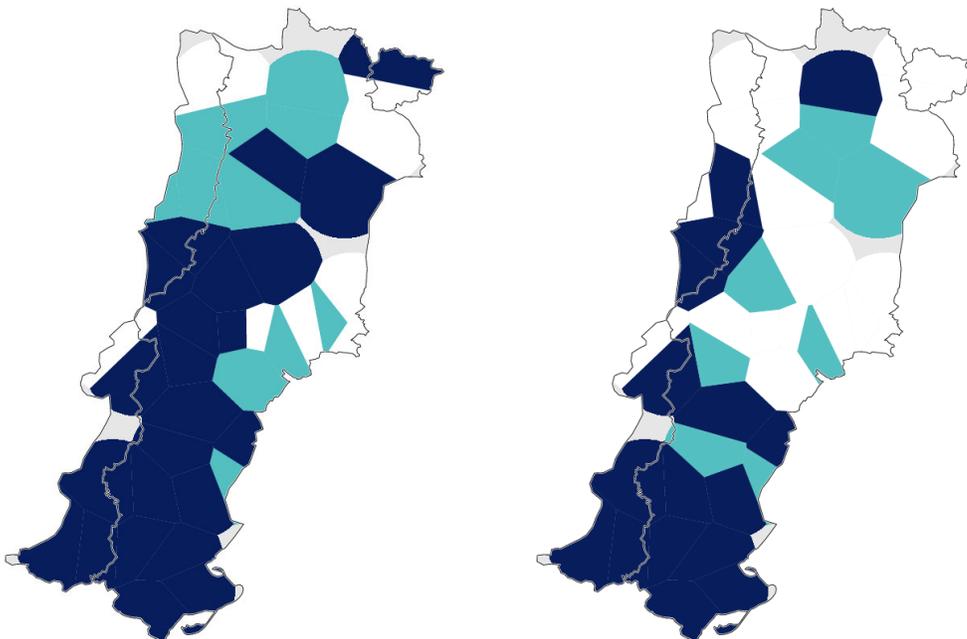


Fig. 9a and 9b Distribution maps of the masculine singular article 'lo' (context:  $\_#C$ , 'lo cap', 'the head') according to the pronunciations of the oldest speakers (F4: Fig. 9a, left) and the youngest speakers (F1: Fig. 9b; right). The dark blue indicates sites only having the dialectal variant 'lo' (in standard/eastern varieties 'el').

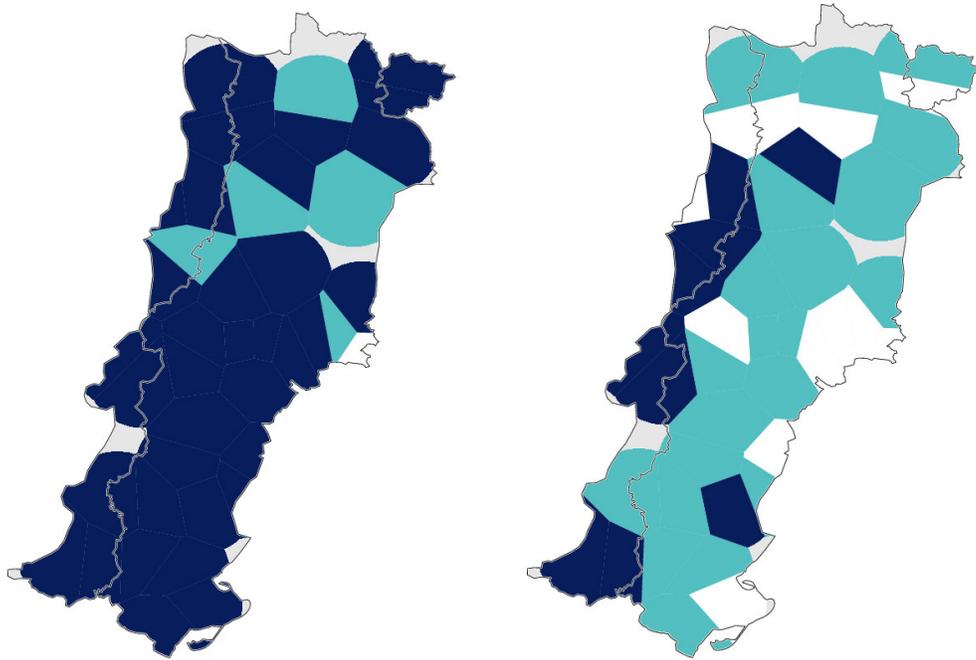


Fig. 10a and 10b Distribution maps of the second person singular reflexive clitic 'us' (context: *w#\_*, 'arreglu's', 'get yourself ready') according to the pronunciations of the oldest speakers (F4: Fig. 10a, left) and the youngest speakers (F1: Fig. 10b; right). The dark blue indicates sites only having the dialectal variants 's' and '-tos' (in standard '-vos': 'arreglu-vos').

#### 4.3 The diffusion of linguistic change: a combination of patterns

As it is widely accepted in dialectology that “large cities usually show a higher rate of innovation than surrounding areas (to which innovations then radiate)” (Samuels 1972: 93), we tested this hypothesis with our data. In addition, we also paid attention to the patterns of spatial diffusion of the linguistic change that emerge from the results. Specifically, we tried to confirm the basic assumption of Taeldeman (2005: 265) that the *hierarchical diffusion pattern* and the *contagious diffusion pattern* interact closely, even while the first one plays the more important role. In other words, this means that linguistic innovations usually start in large cities and are spread to rural areas via smaller satellite towns more often than they are propagated in waves from one area to another (see also Trudgill 1974). As a consequence, we expect the older urban informants to speak a more standard variety than the rural ones. Carrera-Sabaté (2006:

23-24) confirmed this when she studied the maintenance of the traditional north-western realization [a] in words with an absolute initial pre-stressed vowel <e> (see Section 4.2). She adduced evidence that informants aged 61 to 80 (born between 1914 and 1933) were more conservative in Alguaire (a north-western rural village) than in Lleida (the capital of the county, which we also have in our corpus). These differences in the degree of standardization completely disappeared among the speakers aged 21-30, and appeared again among younger teenagers (speakers whose idiolects are not completely stable yet, see Section 2) and especially among children aged 3-5 (speakers with very unstable idiolects). We had expected the hierarchical diffusion pattern to be more evident *among the older speakers*, as the dialects of the younger informants are already so similar that it is not possible to distinguish between rural and urban dialects from an aggregate perspective anymore.

As we expected, the results show that these two patterns interact closely. On the one hand, Fig. 7a, 7b, 8a, 8b, 9a, 9b, 10a and 10b clearly show a contagious diffusion pattern: the standard features are propagating in waves from the innovative East (closest to the standard) to the more conservative West.

Regardless of this gradual westward spreading of the standard forms, Fig. 4 and 5 also suggest that the hierarchical diffusion pattern (HP) has played a role in the diffusion of the standard forms. The MDS plot in Fig. 4 illustrates that the older speakers in the largest cities of each county (e.g., La Seu d'Urgell, Mollerussa, and Tàrrrega marked by numbers 3, 25 and 39, respectively in Fig. 4) speak in a manner that is systematically closer to the standard than the speakers in the corresponding rural settlement (e.g., Coll de Nargó, Sidamon, and Preixana, marked by numbers 4, 26 and 40, respectively in Fig. 4). In general the larger symbols are closer to the standard than their smaller counterparts. This means that the more urban areas usually show a higher

rate of innovation than their surrounding (more conservative) peripheral rural areas. This can also be observed in Fig. 7a and 7b, where urban centres tend to be darker coloured than the rural villages in the same counties (at least amongst older, F4 speakers). It is therefore possible to infer that cities tend to propagate innovations towards the rural areas *in the beginning* of linguistic change. Once advergence is nearly complete the differences between the urban and the rural varieties disappear, so that it is no longer possible to detect differences related to this variable among the young (F1) speakers.

The aggregate perspective makes it clear that in the process of linguistic change modifying the north-western dialects, and particularly in Catalonia and Andorra, the hierarchical diffusion pattern must have played a crucial role. This pattern clearly interacts with the contagious diffusion pattern, as the first urban varieties to converge towards the eastern dialects (in the beginning) and towards the standard variety (nowadays) are those located in the eastern counties of the north-western Catalan area. This is one of the reasons why the most conservative dialects are the *tortosí* and the *ribagorçà* (located, together with the Aragonese dialects, far away from the counties in the east).

#### *4.4 Intra-systemic variation*

In this section we will briefly examine the evolution of the dialects in the Aragonese counties, as they are in a sense isolated from the influence of the official standard Catalan language and might show intra-systemic variation, causing the dialects to diverge from the standard and from the rest of the western Catalan dialects. A closer look at the data reveals that analogy, for instance, is already the primary source of

change in the unification of the second and third verbal classes, in at least some localities.

Verbal class		Older (F4)	Younger (F1)
II[-EXT]	PI	[pérðo]	[pérk]
<i>Perdre</i> ‘to lose’	PS	*but: [pérɣa], [pérɣes], [pérɣe], [perɣém], [perɣéw], [pérɣen]	*and: [pérɣa], [pérɣes], [pérɣe], [perɣém], [perɣéw], [pérɣen]
II[+EXT]	PI	[bék]	[bék]
<i>Beure</i> ‘to drink’	PS	*and: [béɣa], [béɣes], [béɣe], [beyém], [beyéw], [béɣen]	*and: [béɣa], [béɣes], [béɣe], [beyém], [beyéw], [béɣen]
III[-EXT]	PI	[sínto], [síns], [sín], [señtím], [señtíw], [sínten]	[señtísko], [señtíses], [señtís], [señtím], [señtíw], [señtísen]
<i>Sentir</i> ‘to hear’	PS	*and: [sínta], [síntes], [sínte], [señtím], [señtíw], [sínten]	*and: [señtíska], [señtíses], [señtíske], [señtíském]/[señtím], [señtískéw]/[señtíw], [señtísken]
III[+EXT]	PI	[serβísko], [serβíses], [serβís], [serβím], [serβíw], [serβísen]	[serβísko], [serβíses], [serβís], [serβím], [serβíw], [serβísen]
<i>Servir</i> ‘to serve’	PS	*and: [serβíska], [serβíses], [serβíske], [serβíském]/[serβím], [serβískéw]/[serβíw], [serβísken]	*and: [serβíska], [serβíses], [serβíske], [serβíském]/[serβím], [serβískéw]/[serβíw], [serβísken]

Table 3. An example of analogy’s role in the spreading of the velar and palatal extensions in two verbal tenses in Fraga and Vilella de Cinca (el Baix Cinca, Aragon). The left column shows the pronunciations of the older speakers (F4), while the right column shows the pronunciations of the younger speakers (F1).

As may be seen in Table 3, the older speakers of Fraga and Vilella de Cinca (Aragon) had already broadened the use of the velar extension (a feature of the II[+EXT] verbal class) to verbs where it was absent some decades ago<sup>11</sup> (those of the II[-EXT] verbal class). These are the forms shown in the left (F4) column (older speech). The homogenization of paradigms has increased even more amongst young speakers, because some of them also add this extension to PI 1 [-PLU]: ‘perdo’ [pérk] ‘I lose’. A leveling of the third verbal class has also taken place in this area; whereas the palatal extensions [íj] and [íj]k formerly appeared only in verbs in III[+EXT], speakers now also add it to the stems of the III[-EXT] verbs. The consequence is that both II[-

EXT] and II[+EXT], on the one hand, and III[-EXT] and III[+EXT], on the other, have merged into only two verbal classes, II and III.

As we have tried to show in this paper, the evolution of the dialects in Aragon differs considerably from the evolution of the rest of the north-western Catalan dialects. In the following we will focus more on explaining these differences.

## **5. Explaining linguistic change**

As Woolhiser (2005: 236) points out:

with the rise of the modern nation state in the nineteenth century, accompanied in the twentieth century by the emergence of modern communications, improved transportation networks, greater geographical and social mobility of populations, and universal education, political borders have become a far more potent factor in dialect divergence and convergence.

In this paper we have seen that the north-western Catalan dialects have undergone a gradual process of language leveling toward the eastern dialects, traditionally considered the most prestigious ones. The factors mentioned by Woolhiser (2005) (the expansion of the media, the improvement of the transportation networks and the subsequent increase of geographical mobility) must have contributed to the diffusion of the eastern features in the territories where the dialect of Barcelona was considered the “dialect of culture”. This hypothesis is confirmed by the fact that eastern features have historically spread westward following two main paths: the current highway A-2, a road which has traditionally linked Barcelona and Lleida going through Cervera, Tàrraga and Mollerussa; and the course of the river Segre, which links la Seu d’Urgell with Lleida going through Balaguer (see Fig. 1). These were the two most important means of communication between the East and the West of Catalonia and between the Pyrenees and the plain of Lleida, respectively. They were, therefore, the main paths along which the speakers of the eastern and the western dialects could interact.

As we mentioned before (see Section 1), the greater prestige of the eastern varieties increased again when Fabra decided to base the modern compositional grammar of Catalan on this dialect. This grammar was introduced at school in Catalonia during the 1920s. Some researchers have pointed out that the prestige of the north-western dialects diminished as a consequence (Ferrando 2000). The diminished prestige decreased further during the transition period which followed Franco's regime, when the Catalan government implemented a language policy in favor of Catalan in the early 1980s. Catalan became (together with Spanish) the official language of Catalonia and the common language of school and the public media. At the same time, Catalan became an obligatory requirement for civil service jobs. We have already seen that some studies (e.g. Carrera-Sabaté 2002, 2003 and 2006 and Carrera-Sabaté & Freixenet-Esteve 2003) concluded that the diffusion of the standard language might be furthering a tendency to adapt speech forms to written models. They had observed changes among the young speakers of north-western Catalan, which they attributed to an advergence process towards the standard. Since the authors especially detected these tendencies in formal contexts, it might be argued that the standard is not really the principal cause of change in the north-western dialects, since speakers who learn it are only enriching their stylistic repertoire (in other words, they broaden their repertoire by adding standard Catalan to their own dialect in a diglossic relationship). This situation might be possible among the speakers of *tortosí* and *ribagorçà*, two varieties which are still considered prestigious to some extent among their speakers. At the end of compulsory schooling these speakers might be able to use the standard variety in formal or written contexts, while keeping to their dialect in other (more informal) contexts. However, it seems clear that this is not the case for rest of the north-western varieties, where the vertical process of advergence towards the standard entails a process of structural dialect loss, as some

attitudes make clear: for instance, several young informants stated that they only used certain north-western features, such as the dialectal masculine singular article “lo” “when they wanted to joke”, because it sounded too “peasant-like”. This is only one anecdotal example of the speakers’ attitude toward their own dialects, but this sort of attitude may be widespread.

In our analysis we have observed that there are two more conservative areas within Catalonia: the *ribagorçà* area (a “dialectal elephant cemetery”, according to Sistac 1997), and the *tortosí* area, without a doubt the least influential of the prestigious Catalan varieties. Sistac (2009), who studied what he calls “the slow death throes of the dialects”, notes that there are several factors responsible for their greater conservatism, such as geographical and political isolation. We must keep in mind that they both are peripheral areas: the first one surrounded by the Pyrenees; the second located at the southern edge of Catalonia. The relations between these areas and the capital Barcelona have traditionally been quite weak. According to Sistac (1997), however, the crucial point is “the differential awareness/pride of these regions”. He finds that there is a clear difference between these dialects and *lleidatà* (as we have seen, the most standardized variety) when it comes to their “sense of belonging to Catalonia”:

The West –especially the countryside– has always been deeply *catalanist*, without a clear awareness of the internal [*linguistic*] differences, but with a clear consciousness of the external [*referring to Spain*] ones. This has led to a situation where the internal differences are considered an anachronism, a nuisance [...]. As a consequence, the north-western features are much more alive in the southern area (where *tortosí* is spoken) than in the West (much more influenced by Barcelona), where *lleidatà* has become an external sign of rurality and marginalization (Sistac 1997: 45).

This explanation accords with Labov’s approach in taking into consideration attitudinal and ideological factors in accounting for linguistic change; according to Labov (1966), a

positive attitude towards a dialect results in less adaptation to the standard language than does a negative one. The results of our study, thus, seem to confirm again this sociolinguistic generalization.

The ideological factor is also crucial to understanding the different evolution of the Aragonese dialects in contrast to those of Catalonia and Andorra. On the one hand, the Catalan-speaking counties of Aragon (the so-called “Franja”) form a long stripe within which many sociolinguistic and cultural differences coexist: therefore, there is no consciousness of internal unity. Speakers regard themselves as Aragonese and know that they speak something different from Spanish, the only official language in Aragon, but most of them do not really know where these dialects are spoken. As a consequence, their acculturation and diglossia increases their local orientation: the dialects spoken in Aragon have not traditionally been called “Catalan” by their speakers, but *fragatí*, *lliterà*, *ribagorçà*...; every village had its own term. Another way of referring to the dialect was to use the disdainful word *xapurreau*, similar to the French *patois*, a term which avoided identifying it with Catalan. This poor reputation of the local varieties increased diglossia, so that Spanish was (and still is) the language used in formal contexts and the “roofing” language of the local vernaculars.

Further, the Autonomous Community of Aragon has not made any effort to provide these speakers with a law to guarantee their linguistic rights. Therefore, they are in a situation of complete legal vulnerability: not only has the government not made Catalan official in the Franja, it has also hidden the name of the language in the Statute of Autonomy of Aragon. In this document, the local vernaculars are considered mere “linguistic modalities”, which should be protected and promoted by the government. In addition to this legal vulnerability, the rest of Aragon is not interested in respecting the Catalan-speaking minority (the Franja, where 90% of the inhabitants speak Catalan,

represents only around 5% of the Aragonese population). They are called, for instance, “Polish” (the same name used by some Spaniards to designate Catalans, as they do not (want to) understand them when they speak). That is why the inhabitants of the Franja have felt the need to profess that they are Aragonese and that they feel Aragonese, because they want to be accepted in their social and political community. Needless to say, this situation has made it more difficult for them to accept that their language is indeed Catalan (“Catalan is spoken in Catalonia but I am Aragonese, so my language cannot be Catalan”).<sup>12</sup>

In recent years this term has gradually been accepted, especially since Catalan was introduced at school as a voluntary subject during the course 1984-1985. However, speakers still point out the differences between their linguistic and their national identity: most of the inhabitants of the Franja do indeed speak Catalan; but most of them also consider themselves Aragonese, not Catalan. It is, therefore, a clear example of how political borders can give rise to sociocultural discontinuities, and how these differences become more important than the multiple elements of shared history between the counties located on the two sides of this border.<sup>13</sup> This border has been crucial to perpetuate the diglossic situation in Aragon, with the help of the Aragonese public institutions:

When the rural bourgeoisie disappeared, the dominance of Spanish lost its support, although the descendants of the old wealthy families kept speaking Castilian in spite of having lost their social position. But what went on, and even increased thanks to massive schooling, was the support of school to the only official language. This was also the attitude of the Church, which, after the Second Vatican Council, would never use the language of the people in the liturgy, but would exclusively use Spanish. And certainly this situation was indebted to the support of the public administration and the public media during the dictatorship, but also after Franco’s regime (Bada 1990: 15).<sup>14</sup>

To sum up, therefore, we can see that nowadays the political border between Catalonia and Aragon has become an ideological boundary with palpable consequences in the evolution of the Catalan dialects located at both sides of the border. With respect to the Aragonese situation, we should be aware, as Woolhiser (2005) reminds us, that:

The degree to which cross-border horizontal (dialect-dialect) or vertical (dialect-standard) convergence or divergence occur within a politically divided dialect area or continuum is thus determined not only by the linguistic relationship between the roofing varieties and the local vernaculars on the two sides of the border and by the physical obstacles presented by different types of border regimes, but also by the degree of success of national institutions in instilling in local populations a higher level of solidarity with co-nationals than with citizens of the neighbouring state (Woolhiser 2005: 262).

The big effect is that education in Catalan is causing standardization and de-dialectalization in Catalonia and Andorra, but not in Aragon, where Catalan is not official and is not (almost) taught at school. This big effect is complemented by the effect that locals in the Franja feel Aragonese and do not feel the necessity of learning Catalan because, to them, it is a variety with less prestige than Spanish.

## **6. Discussion and prospects**

In this paper we have accounted for the process of structural dialect loss undergone by the north-western Catalan dialects during the last 80 years. To reach this goal, we used a contemporary corpus and took advantage of a range of dialectometric methods that allowed us to calculate and analyze the linguistic distance between varieties in apparent time from an aggregate perspective. We provided evidence that this evolution is due to a twofold process of linguistic advergence: towards the eastern varieties, historically; and towards the standard, during the last decades. Two patterns have interacted in the

diffusion of the prestigious features: the contagious diffusion pattern, where the innovations are propagated westwards in waves; and the hierarchical diffusion pattern, as cities have propagated innovations towards the rural areas, becoming “linguistic islands” in the beginning which have been crucial in enhancing language leveling. This situation has resulted in a major border effect between the varieties located on either side of the Catalan-Aragonese border. As a consequence, linguistic distance has increased considerably between varieties which had been almost identical. Finally we have tried to explain the reasons for these changes, paying attention to the ideological and attitudinal factors that have strengthened the border effect.

Of course, there are several aspects that we would like to investigate in further research. The first interesting research question arose when we described linguistic leveling in Catalonia. We hypothesized that the convergence amongst the youngest speakers has not been caused by the emergence of a north-western regiolect, but we cannot be entirely certain that there is no morphological variation left in these youngest speakers. In further studies we will investigate the different speeds of linguistic change, taking into account either morphological or phonological data. As a consequence, we should be able to clarify whether these leveled varieties can be still considered autonomous dialects or just standardized varieties with dialectal pronunciations.

It might also be very interesting to relate all these processes to other sociolinguistic variables, such as gender, and demographic variables (e.g. population size, average population income, average population age and mobility of the population). In addition, we would like to investigate the varieties located near the Aragon-Catalonian border, to more closely examine the border effect.

We mentioned above one example of the internal evolution which some Aragonese varieties are currently experiencing. It would be interesting to analyze the

intra-systemic variation within the Aragonese counties in more detail, in order to see if there are several autonomous changes that might increase the linguistic distance between the Aragonese dialects and the rest of the north-western varieties even further.

Finally, we mentioned two other examples of border effects which are currently taking place in the Catalan-speaking area (i.e. near the Spanish-French border and near the Catalan-Valencian border). Similar studies should therefore be conducted for these areas to investigate the direction and speed of these phenomena in these cases.

## Notes

---

<sup>1</sup> We would like to thank Mar Massanell for having gathered *in situ* 75% of the data used in this research.

<sup>2</sup> Traditionally, the eastern dialects of Catalonia have been considered the most prestigious ones, as they were the varieties of the capital, Barcelona, when Catalan was promoted again as a language of culture during the last decades of the 19th century. The fact that most of the cultural production took place using this variety led Pompeu Fabra (“the architect of modern Catalan”, see Costa 2009) to base his compositional Catalan grammar on the eastern dialects. With the terms *orientalization* or *easternization*, therefore, we refer to the process by means of which the north-western dialects have gradually become more similar to the eastern (prestigious) ones.

<sup>3</sup> According to the classification made by Viaplana (1984, 1995).

<sup>4</sup> We are using the acute accent (´) to designate stressed vowels.

<sup>5</sup> Note that we did not normalize the other seven morphological categories, as this drawback affected only the verbs.

<sup>6</sup> All visualization techniques are implemented in the online dialect visualization interface “Gabmap” (Nerbonne et al. 2011), created at the *Center for Language and Cognition* of the University of Groningen and available at <http://www.gabmap.nl>.

<sup>7</sup> To prevent clutter, the numbers referring to the varieties have not been added to the graph.

<sup>8</sup> Obviously, there might be some shibboleths which could still indicate the linguistic provenance of the speakers, but we are focusing on aggregate differences in this paper.

<sup>9</sup> We cannot rule out the possibility of a vertical (maybe lexical) convergence of the varieties in Aragon towards standard Spanish, however. Broadly speaking, we may say that the diglossic situation now differs on the two sides of the Catalan-Aragonese political border: while the “roof” variety for the Catalan dialects is standard Catalan in Catalonia and Andorra, it is Standard Spanish in Aragon –with the exception of the people involved in the revitalization of the Catalan language in these counties.

<sup>10</sup> It must be taken into account, however, that the legal status of Catalan in Aragon is quite different from its status in the other two areas studied. According to the *Statute of Autonomy* (2006), Catalan is the “native” and, together with Castilian, the “official” language of Catalonia, and it is the “native, national and official” language of Andorra according to its Constitution (1993). In contrast, Catalan has no official status in Aragon, despite passing the “Law of Languages” in 2009, which theoretically allows the Aragonese government to design and implement a language policy in favor of Catalan and Aragonese, the other minority language spoken in the area.

<sup>11</sup> See Perea (2003, 2006) for an analysis of the distribution of the velar and the palatal extensions in the Catalan verbs at the beginning of the 20th century.

<sup>12</sup> In Andorra the situation has been quite different: although they have a strong identity of being Andorran, they do not need to be accepted by any other higher political or social community. They consider themselves Andorran and consider, *at the same time*, that Catalan is *also* their national language, as it is in many other territories.

<sup>13</sup> For instance, many inhabitants of the central counties of the Franja were born in the hospital of Lleida, they studied in Lleida or they have their jobs in the city. In addition, many inhabitants of the northern

---

county of the Franja go to school at El Pont de Suert in Catalonia. Besides, many Aragonese have pointed out the economic dependence of the Franja with respect to Catalonia. On the other hand, the Catalan public TV channels have also been watched in the Franja since 1985 with great success. The fact that these dialects have not been influenced by the standard seems to accord with the idea that “the possible influence of television has become another ‘language myth’ (cf. Chambers 1998). Leaving aside recent German scholarship, the majority of linguists seem to view any possible influence of television as very weak, possibly providing information about linguistic variation, presenting alternative linguistic models, and affecting attitudes to existing varieties” (Stuart-Smith 2006: 142).

<sup>14</sup> It is interesting to know that the author of this paragraph is José R. Bada, the first Minister of Culture of Aragon after the dictatorship. It is obvious, therefore, that he is not suspicious of Catalan imperialism although he was born in the Franja and his first language is Catalan. He made many efforts to convince the population of Aragon that Catalan is also *an Aragonese language* and that the rest of the Autonomous Community must accept the cultural and linguistic diversity of Aragon.

---

**Funding**

This research was (partly) supported by the project “Descripción e interpretación de la variación dialectal: aspectos fonológicos y morfológicos del catalán” (FFI2010-22181-C03-02), financed by MICINN and FEDER.

---

## References

**Auer, Peter; Hinskens, Frans & Paul Kerswill** (2005): *Dialect Change. Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press.

**Bada, José R.** (1990): *El debat del català a l'Aragó (1983-1987)*. Calaceit: Associació Cultural del Matarranya / Centre d'Estudis del Baix Cinca / Associació de Consells Locals de la Franja.

**Bailey, Guy** (2002): "Real and Apparent Time". In Chambers, Jack; Trudgill, Peter & Natalie Schilling-Estes (eds.): *The Handbook of Language Variation and Change*. Malden, Oxford & Carlton: Blackwell, 312-332.

**Berruto, Gaetano** (2005): "Dialect/standard convergence, mixing, and models of language contact: the case of Italy". In Auer, Peter; Hinskens, Frans & Paul Kerswill: *Dialect Change. Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, 81-95.

**Bibiloni, Gabriel** (2002): "Un estàndard nacional o tres estàndards regionals?" In Bernat Joan (ed.): *Perspectives sociolingüístiques a les Illes Balears*. Eivissa: Res Publica.

**Black, Paul** (1976): "Multidimensional Scaling Applied to Linguistic Relationships". In *Cahiers de l'Institut de Linguistique de Louvain*, 3, n. 5-6.

**Britain, David** (2002): "Space and Spatial Diffusion". In Chambers, Jack; Trudgill, Peter & Natalie Schilling-Estes (2002): *The Handbook of Language Variation and Change*. Malden / Oxford / Victoria: Blackwell, 603-637.

**Carrera-Sabaté, Josefina** (2002): "Approche de la variation phonétique dans la ville de Lleida à partir de l'observation de ses habitants et des émissions télévisées locales". In *Marges Linguistiques* 3, 201-214.

---

**Carrera-Sabaté, Josefina** (2003): “Variació verbal al Segrià: la pronunciació del morfema de 3a persona en diversos àmbits d’ús”. In *Noves SL. Revista de Sociolingüística*.

**Carrera-Sabaté, Josefina** (2006): “Some connections between linguistic change and the written language: The behaviour of speakers aged 3 to 20”. In *Language Variation and Change* 18, 15-34.

**Carrera-Sabaté, Josefina & Elisabet Freixenet-Esteve** (2003): “‘Auliva’ o ‘oliva’? Caracterització de l’alternança al Segrià”. In *Noves SL. Revista de Sociolingüística*.

**Chambers, Jack** (1998): “Myth 15: TV Makes People Sound the Same”. In Bauer, Laurie & Peter Trudgill (eds.): *Language Myths*. Harmondsworth: Penguin Books, 123-131.

**Chambers, Jack** (2002): “Patterns of Variation including Change”. In Chambers, Jack; Trudgill, Peter & Natalie Schilling-Estes (2002): *The Handbook of Language Variation and Change*. Malden / Oxford / Victoria: Blackwell, 349-372.

**Chambers, Jack & Peter Trudgill** (1998): *Dialectology*. Cambridge: Cambridge University Press, 2d edition.

**Chambers, Jack; Trudgill, Peter & Natalie Schilling-Estes** (2002): *The Handbook of Language Variation and Change*. Malden / Oxford / Victoria: Blackwell.

**Clua, Esteve** (1998): *Variació i distància lingüística. Classificació dialectal del valencià a partir de la morfologia flexiva*. Barcelona: Universitat de Barcelona.

**Costa, Joan** (ed.) (2009): *The Architect of Modern Catalan. Pompeu Fabra (1868-1948). Selected writings*. Amsterdam / Philadelphia: John Benjamins

**Fabra, Pompeu** (1918): *Gramàtica catalana*. Barcelona: Institut d’Estudis Catalans.

**Ferrando, Antoni** (2000): “La presència dels dialectes territorials en l’obra de Pompeu Fabra”. In Ginebra, Jordi; Martínez, Raül-David i Miquel-Àngel Pradilla (eds.): *La*

---

*lingüística de Pompeu Fabra*. Alacant: Institut Interuniversitari de Filologia Valenciana / Universitat Rovira i Virgili, 279-342.

**Gerritsen, Marinel** (1999): “Divergence of dialects in a linguistic laboratory near the Belgian–Dutch–German border: Similar dialects under the influence of different standard languages”. In *Language Variation and Change* 11, 43-65.

**Goebel, Hans** (1991): “Dialectometry: a short overview of the principles and practice of quantitative classification of linguistic atlas data”. In Köhler, Reinhard & Burghard B. Rieger (eds.): *Contributions to quantitative linguistics*. Dordrecht: Kluwer, 277-315.

**Goebel, Hans** (2000): “Langues standards et dialectes locaux dans la France du Sud-Est et l’Italie septentrionale sous le coup de l’effet-frontière: une approche dialectométrique”. In *International Journal of the Sociology of Language* 145, 181-215.

**Hinskens, Frans; Auer, Peter & Paul Kerswill** (2005): “The study of dialect convergence and divergence: conceptual and methodological considerations”. In Auer, Peter; Hinskens, Frans & Paul Kerswill (2005): *Dialect Change. Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, 1-48.

**Kerswill, Paul** (2002): “Koineization and Accommodation”. In Chambers, Jack; Trudgill, Peter & Natalie Schilling-Estes (2002): *The Handbook of Language Variation and Change*. Malden / Oxford / Victoria: Blackwell, 669-702.

**Kleiweg, Peter, John Nerbonne & Leonie Bosveld** (2004): “Geographic Projection of Cluster Composites”. In Blackwell, Alan F.; Marriott, Kim & Atsushi Shimojima (eds.): *Diagrammatic Representation and Inference. Diagrams 2004*. Berlin: Springer, 392-394.

**Labov, William** (1966): *The social stratification of English in New York City*. Washington D. C.: Center for Applied Linguistics.

---

**Levenshtein, Vladimir** (1965): *Binary codes capable of correcting deletions, insertions and reversals*. Doklady Akademii Nauk SSSR, 163, 845–848.

**Mattheier, Klaus J.** (1996): “Varietätenkonvergenz: Überlegungen zu einem Baustein einer Theorie der Sprachvariation”. In Ammon, Ulrich; Mattheier, Klaus J. & Peter H. Nelde (eds.): *Sociolinguistica 10. Konvergenz und Divergenz von Dialekten in Europa*, 31-52.

**Massanell, Mar** (1995): *Aproximació a la morfologia flexiva actual de la Seu d’Urgell i de Coll de Nargó*. Universitat de Barcelona.

**Moll, Francesc de B.** (1991): *Gramàtica històrica catalana*. València: Universitat de València.

**Nahkola, Kari & Marja Saanilahti** (2004): “Mapping language changes in real time: A panel study on Finnish”. In *Language Variation and Change* 16, 75-92.

**Nerbonne, John** (2010): “Mapping Aggregate Variation”. In Lameli, Alfred; Kehrein, Ronald & Stephanus Rabanus (eds.): *Mapping Language*, Vol. 2 in series *Language and Space*. Berlin: Mouton De Gruyter. Chap. 24, 476-495, maps 2401-2406.

**Nerbonne, John; Kleiweg, Peter; Manni, Franz & Wilbert Heeringa** (2008): “Projecting Dialect Distances to Geography: Bootstrap Clustering vs. Noisy Clustering”. In Preisach, Christine; Schmidt-Thieme, Lars; Burkhardt, Hans & Reinhold Decker (eds.): *Data Analysis, Machine Learning, and Applications. Proceedings of the 31st Annual Meeting of the German Classification Society*. Berlin: Springer, 647-654.

**Nerbonne, John; Colen, Rinke; Gooskens, Charlotte; Kleiweg, Peter & Therese Leinonen** (2011): “Gabmap — A Web Application for Dialectology”. In *Dialectologia. Special Issue 2011*.

---

**Pedersen, Inge Lise** (2005): "Processes of standardisation in Scandinavia". In Auer, Peter; Hinskens, Frans & Paul Kerswill (2005): *Dialect Change. Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, 171-195.

**Perea, Maria Pilar** (2003): "La tendència expansiva del segment velar dels verbs de la segona conjugació en els dialectes catalans". In: *Actas del XIII Congreso Internacional de Lingüística y Filología Románica. Salamanca 2001*. Tübingen: Max Niemeyer Verlag, vol 1, 403-417.

**Perea, Maria Pilar** (2006): "L'expansió dels verbs incoatius en català. Distribució i valor del segment extensió -eix". In *Actes del VIIIè Congrès de Lingüística General*. Barcelona: Universitat de Barcelona, CD-ROM.

**Pradilla, Miquel-Àngel** (2008a): *Sociolingüística de la variació i llengua catalana*. Barcelona: Institut d'Estudis Catalans.

**Pradilla, Miquel-Àngel** (2008b): *La tribu valenciana. Reflexions sobre la desestructuració de la comunitat lingüística*. Benicarló: Onada.

**Prokić, Jelena & John Nerbonne** (2008): "Recognizing Groups among Dialects". In *International Journal of Humanities and Arts Computing, Special Issue on Language Variation* ed. by John Nerbonne, Charlotte Gooskens, Sebastian Kürschner, and Renée van Bezooijen, 153-172.

**Recasens, Daniel** (1996): *Fonètica descriptiva del català (assaig de caracterització de la pronúncia del vocalisme i consonantisme del català al segle XX)*. Barcelona: Institut d'Estudis Catalans, 2d edition.

**Samuels, Marilyn** (1972): *Linguistic Evolution – with Special Reference to English*. Cambridge: Cambridge University Press.

**Sapir, Edward** (1921): *Language. An Introduction to the Study of Speech*. San Diego: Harcourt Brace Jovanovich.

---

**Shaw, David** (1974): “Statistical analysis of dialect boundaries”. In *Computers and the Humanities* 8, 173-177.

**Sistac, Ramon** (1997): “Els dialectes a la fi del mil·lenni”. In *El català d'Àneu*, 37-55.

**Sistac, Ramon** (2009): “La lenta agonia del dialecte”. In M. Carme Junyent (ed.): *Transferències. La manifestació dels processos extralingüístics en les llengües del món*. Vic: Eumo, 107-118.

**Stuart-Smith, Jane** (2006): “The influence of the media”. In Llamas, Carmen; Mullany, Louise & Peter Stockwell (eds.): *The Routledge Companion to Sociolinguistics*. Londres: Routledge, 140-148.

**Taeldeman, Johan** (2005): “The influence of urban centres on the spatial diffusion of dialect phenomena”. In Auer, Peter; Hinskens, Frans i Paul Kerswill (2005): *Dialect Change. Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, 263-283.

**Trudgill, Peter** (1974): “Linguistic Change and Diffusion: Description and Explanation in Sociolinguistic Dialect Geography.” In *Language in Society* 2, 215–246.

**Valls, Esteve** (2008): “La desdialectalització del català nord-occidental: cap a una convergència total amb l'estàndard?” In *De Lingua Aragonensi* 4, 55-89.

**Valls, Esteve** (2009): “Cap a una diversificació metodològica de la dialectometria catalana: primers resultats d'aplicar la distància de Levenshtein al Corpus Oral Dialectal”. In *Interlingüística XX*.

**Veny, Joan** (1982): *Els parlars catalans (Síntesi de dialectologia)*. Palma: Editorial Moll.

**Viplana, Joaquim** (1980): “Algunes consideracions sobre les formes pronominals clítics del barceloní”. In *Anuario de Filología* 6. Barcelona: Universitat de Barcelona, 459-483.

- 
- Viaplana, Joaquim** (1984): “La flexió verbal regular del valencià”. In *Miscel·lània Sanchis Guarner 1. Quaderns de Filologia*. València: Universitat de València, 391-407.
- Viaplana, Joaquim** (1995): “Sobre la irregularitat verbal”. In *Caplletra* 19, 333-348.
- 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.
- Virgili, Josep-Maria** (1992): “Problemes d’estàndard en l’àrea fronterera entre el Principat de Catalunya i el País Valencià”. In *Actes del Segon congrés de la llengua catalana IV. Àrea 3. Lingüística social*. Palma: Universitat de les Illes Balears, 555-560.
- Wieling, Martijn** (2012): *A Quantitative Approach to Social and Geographical Dialect Variation*. PhD Thesis. University of Groningen.
- Wieling, Martijn; Margaretha, Eliza & John Nerbonne** (2012): “Inducing a measure of phonetic similarity from pronunciation variation”. In *Journal of Phonetics* 40, 307-314.
- Woolhiser, Curt** (2005). Political borders and dialect divergence/convergence in Europe. In Auer, Peter; Hinskens, Frans & Paul Kerswill (2005): *Dialect Change. Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, 236-262.

## Appendix A: Detailed item list

	M[-PLU]	F[-PLU]	M[+PLU]	F[+PLU]
<b>_<sub>-</sub>+C</b>	el	la	els	les
<b>_<sub>-</sub>+V</b>	l'	l'	els	les
<b>_<sub>-</sub>+V (if V=stressed i/hi)</b>		l'		
<b>_<sub>-</sub>+V (if V=stressed u/hu)</b>		l'		
<b>_<sub>-</sub>+V (if V=unstressed i/hi)</b>		la		
<b>_<sub>-</sub>+V (if V=unstressed u/hu)</b>		la		
<b>_<sub>-</sub>+C (if C=anthroponym)</b>	el/en	la		
<b>_<sub>-</sub>+V (if C=anthroponym)</b>	l'	l'		

Table A1. Articles

	<b>_<sub>-</sub>+V</b>	<b>_<sub>-</sub>+C</b>	<b>V+_<sub>-</sub></b>	<b>w+_<sub>-</sub></b>	<b>r+_<sub>-</sub></b>	<b>nt+_<sub>-</sub></b>
<b>1[-PLU]</b>	m'	em	'm	-me	-me	-me
<b>2[-PLU]</b>	t'	et	't		-te	-te
<b>Reflexive</b>	s'	es	's		-se	-se
<b>1[+PLU]</b>	ens	ens	'ns	-nos	-nos	-nos
<b>2[+PLU]</b>	us	us		-vos	-vos	-vos
<b>Partitive</b>	n'	en	'n	-ne	-ne	
<b>Anaphoric article</b>	l'	el	'l	-lo	-lo	-lo
<b>M[-PLU]</b>						
<b>Anaphoric article</b>	l'	la	-la	-la	-la	-la
<b>F[-PLU]</b>						
<b>Anaphoric article</b>	els	els	'ls	-los	-los	-los
<b>M[+PLU]</b>						
<b>Anaphoric article</b>	les	les	-les	-les	-les	-les
<b>F[+PLU]</b>						
<b>Anaphoric article</b>	l'					
<b>M[-PLU] (if article=definite attribute)</b>						
<b>Anaphoric article</b>	l'					
<b>F[-PLU] (if article=definite attribute)</b>						
<b>Anaphoric article</b>	els					
<b>M[+PLU] (if article=definite attribute)</b>						
<b>Anaphoric article</b>	les					
<b>F[+PLU] (if article=definite attribute)</b>						
<b>Neuter pronoun</b>	ho	ho	-ho	-ho	-ho	-ho
<b>Neuter pronoun [-PLU] (if pronoun=indefinite attribute)</b>	ho					
<b>Neuter pronoun [+PLU] (if pronoun=indefinite attribute)</b>	ho					
<b>Dative [-PLU]</b>	li	li	-li	-li	-li	-li
<b>Dative [+PLU]</b>						-los
<b>Locative</b>	hi	hi	-hi	-hi	-hi	-hi

Table A2. Clitic pronouns

	<b>M</b>	<b>F</b>	<b>M</b>	<b>F</b>
<b>1</b>	aquest	aquesta	aquests	aquestes
<b>2</b>	aqueix*	aqueixa*	aqueixos*	aqueixes*
<b>3</b>	aquell	aquella	aquells	aquelles

Table A3. Demonstrative pronouns

	M
1	açò
2	això*
3	allò

Table A4. Neuter pronouns

	M
1	ací
2	aquí*
3	allí

Table A5. Locative adverbs

	Present Indicative	Imperfect Indicative	Future	Conditional	Present Subjunctive
1[-PLU]	canto	cantava	cantaré	cantaria	canti
2[-PLU]	cantes	cantaves	cantaràs	cantaries	cantis
3[-PLU]	canta	cantava	cantarà	cantaria	canti
1[+PLU]	cantem	cantàvem	cantarem	cantariem	cantem
2[+PLU]	canteu	cantàveu	cantareu	cantàrieu	canteu
3[+PLU]	canten	cantaven	cantaran	cantarien	cantin

	Imperfect Subjunctive	Imperative	Infinitive	Gerund	Past Participle
1[-PLU]	cantés		cantar	cantant	cantat
2[-PLU]	cantessis	canta			
3[-PLU]	cantés	canti <sup>+</sup>			
1[+PLU]	cantéssim	cantem			
2[+PLU]	cantéssiu	canteu			
3[+PLU]	cantessin	cantin <sup>+</sup>			

Table A6. Verb cantar 'to sing'

	Present Indicative	Imperfect Indicative	Future	Conditional	Present Subjunctive
1[-PLU]	perdo	perdia	perdré	perdría	perdi
2[-PLU]	perds	perdies	perdràs	perdríes	perdis
3[-PLU]	perd	perdia	perdrà	perdría	perdi
1[+PLU]	perdem	perdiem	perdreu	perdríem	perdem
2[+PLU]	perdeu	perdiéu	perdreu	perdríeu	perdeu
3[+PLU]	perden	perdien	perdran	perdríen	perdin

	Imperfect Subjunctive	Imperative	Infinitive	Gerund	Past Participle
1[-PLU]	perdés		perdre	perdent	perdut
2[-PLU]	perdessis	perd			
3[-PLU]	perdés	perdi <sup>+</sup>			
1[+PLU]	perdéssim	perdem			
2[+PLU]	perdéssiu	perdeu			
3[+PLU]	perdessin	perdin <sup>+</sup>			

Table A7. Verb perdre 'to lose'

	Present Indicative	Imperfect Indicative	Future	Conditional	Present Subjunctive
1[-PLU]	bec	bevia	beuré	beuria	begui
2[-PLU]	beus	bevia	beuràs	beuries	beguis
3[-PLU]	beu	bevia	beurà	beuria	begui
1[+PLU]	bevem	beviem	beurem	beuríem	beguem
2[+PLU]	beveu	beviéu	beureu	beuríeu	begueu
3[+PLU]	beuen	bevien	beuran	beurien	beguin

	Imperfect Subjunctive	Imperative	Infinitive	Gerund	Past Participle
1[-PLU]	begués		beure	bevent	begut
2[-PLU]	beguessis	beu			

3[-PLU]	begués	begui <sup>+</sup>
1[+PLU]	beguèssim	beguem
2[+PLU]	beguèssiu	beveu
3[+PLU]	beguessin	beguin <sup>+</sup>

Table A8. Verb beure ‘to drink’

	Present Indicative	Imperfect Indicative	Future	Conditional	Present Subjunctive
1[-PLU]	sento	sentia	sentiré	sentiria	senti
2[-PLU]	sents	senties	sentiràs	sentiries	sentis
3[-PLU]	sent	sentia	sentirà	sentiria	senti
1[+PLU]	sentim	sentíem	sentirem	sentiríem	sentim
2[+PLU]	sentiu	sentíeu	sentireu	sentiríeu	sentiu
3[+PLU]	senten	sentien	sentiran	sentirien	sentin

	Imperfect Subjunctive	Imperative	Infinitive	Gerund	Past Participle
1[-PLU]	sentís		sentir	sentint	sentit
2[-PLU]	sentíssis	sent			
3[-PLU]	sentís	sentí <sup>+</sup>			
1[+PLU]	sentíssim	sentim			
2[+PLU]	sentíssiu	sentiu			
3[+PLU]	sentíssin	sentin <sup>+</sup>			

Table A9. Verb sentir ‘to hear’

	Present Indicative	Imperfect Indicative	Future	Conditional	Present Subjunctive
1[-PLU]	serveixo	servia	serviré	serviria	serveixi
2[-PLU]	serveixes	servies	serviràs	serviries	serveixis
3[-PLU]	serveix	servia	servirà	serviria	serveixi
1[+PLU]	servim	servíem	servirem	serviríem	servim
2[+PLU]	serviu	servíeu	servireu	serviríeu	serviu
3[+PLU]	serveixen	servien	serviran	servirien	serveixin

	Imperfect Subjunctive	Imperative	Infinitive	Gerund	Past Participle
1[-PLU]	servís		servir	servint	servit
2[-PLU]	servíssis	serveix			
3[-PLU]	servís	serveixi <sup>+</sup>			
1[+PLU]	servíssim	servim			
2[+PLU]	servíssiu	serviu			
3[+PLU]	servíssin	serveixin <sup>+</sup>			

Table A10. Verb servir ‘to serve’

	M	F	M	F
1[-PLU]	meu	meva	meus	meves
2[-PLU]	teu	teva	teus	teves
3[-PLU]	seu	seva	seus	seves
1[+PLU]	nostre	nostra	nostres	nostres
2[+PLU]	vostre	vostra	vostres	vostres

Table A11. Possessive pronouns

	M	F
1[-PLU]	jo	
2[-PLU]	tu	
3[-PLU]	ell	ella
1[+PLU]	nosaltres	
2[+PLU]	vosaltres	
3[+PLU]	ells	elles

Table A12. Personal pronouns

\* These words do not occur in all varieties.

---

<sup>+</sup> In these cases there is some missing data, because the traditional way in which north-western speakers would address another person in a respectful way would be using Imp 2[+PLU] instead of Imp 3[-PLU] and 3[+PLU].