On the areal dimension of non-standard syntax: Evidence from a Portuguese dialect corpus

Ernestina Carrilho and Sandra Pereira
University of Lisbon

Introduction

This paper investigates aspects of the areal dimension of dialect syntax. More concretely, the investigation focuses on the identification of syntactic areas within the Portuguese territory and their connection with traditional dialect-geographic areas of European Portuguese (henceforth EP). After some brief methodological considerations, the main part of the paper presents the geographic distribution of a selection of non-standard syntactic constructions in EP; the final part further illustrates how syntactic areas compare to dialect areas defined on the basis of other type of linguistic data (namely, phonetic data). Dialect syntax is considered here from a geolinguistic perspective aimed at showing how non-standard syntax may be geographically represented, and relating such syntactic areas to more general dialect areas. As a broad conclusion, the paper highlights the contribution of dialect syntax for geolinguistic purposes.

Methodological background

Although the field of syntax can hardly be regarded as a prolific one in dialectology, the last two decades have given rise to important developments in the study of dialect syntax. Empirical, methodological, and conceptual advances in this recent field encourage enhanced approaches to the spatial dimension of linguistic variation, which may take advantage of new insights into the areal study of syntactic constructions.

The approach developed in this paper relies on the empirical exploitation of a dialect corpus, the Syntax-oriented Corpus of Portuguese Dialects, CORDI-

---

1 This work was supported by Fundação para a Ciência e a Tecnologia (FCT), grant SFRH/BD/27648/2006 (S. Pereira) and project PTDC/LIN/71559/2006. We would like to thank the editors of this volume for their comments and suggestions, and the organizers and participants of the Workshop Dialect and Regiolect Syntax, held within Methods in Dialectology 14, for helpful discussion.

2 Detailed information on recent and current research in dialect syntax has been made available by the Edisyn project at http://www.dialectsyntax.org.
AL-SIN. Over the past decade, studies on different aspects of EP dialect syntax have been carried out in tandem with the compilation of this corpus. The geo-linguistic expression of syntactic variation is the consideration of this paper (see also Carrilho & Pereira 2011; Carrilho & Lobo 2012).

CORDIAL-SIN is an annotated corpus compiling spontaneous or almost spontaneous speech drawn from tape-recorded dialect interviews that were conducted for different linguistic atlases. The corpus amounts to more than 600,000 words, collected from 42 locations distributed over the continental and insular Portuguese territory. These data are representative of traditional dialects as spoken by informants who are aged, non-educated, rural, and born and raised in the place of interview.

The areal distribution of EP non-standard constructions

Different from other non-standard constructions which spread over all dialects (even if absent from standard EP), the constructions under inspection here occur mainly in restricted areas in the Portuguese territory. The following sections present the geographic distribution of a selection of non-standard constructions involving: (i) inflected gerunds; (ii) periphrastic aspectual forms with estar ‘to be’ followed by a gerund; (iii) third person plural verbal agreement with the pronominal form a gente (literally ‘the people’, third person singular); (iv) the verb ter ‘to have’ in impersonal existential constructions; and (v) the absence of a definite article before prenominal possessives.

Inflected gerund

Some EP varieties display the presence of person inflection in gerund forms (1a, 2a), while the gerund is categorically invariant in standard EP (1b, 2b):


4 Most of these studies are available at http://www.clul.ul.pt/en/resources/212-cordial-sin-syntax-oriented-corpus-of-portuguese-dialects, where more detailed information on CORDIAL-SIN may also be found (namely the identification of the CORDIAL-SIN locations represented by a code in the examples in this study, such as LVR05 in (1)).

5 These dialect interviews were gathered by the Dialectology Group of the Centro de Linguística da Universidade de Lisboa, within the scope of the atlases ALEAç, ALEPG, and ALLP.
1. a. Em eles *tendem* a outra, já não querem aquela.
   in they have GER.3PL the other.one already not want.3PL that.one
   ‘when they have the other one, they don’t want that one anymore’
   (LVR05)
   b. *Tendo* eles a outra, …
   have GER they the other.one

2. a. E *tendem* uma árvore, não há pássaro nenhum que poise no chão.
   and have GER.3PL a tree not has bird any that land on the ground
   ‘and if a bird has a tree, it does not land on the ground’ (STJ30)
   b. E *tendo* uma árvore, …
   and have GER a tree

Monographic studies of EP dialects have often noticed instances of inflected gerunds, which, according to Lobo (2001, 2008), occur mainly in adverbial clauses, as in (1) and (2). The areal distribution of these inflected forms in CORDIAL-SIN has been investigated in Lobo (2008): As represented in Figure 1, the non-standard variant mostly appears in a homogeneous area in the southern part of Portugal and in one location in the insular territory (in Azores).^6^

---

^6^ Analysis of this non-standard variant faces particular difficulties, as noticed by Lobo (2008). Since the inflected forms for first person singular and third person singular coincide with the non-inflected gerund form (for instance, *tendo* in (1b)), the absence of data may in some cases mean the absence of unambiguous occurrences of the inflected gerund (which may be the case in the blank spots found in the southern area, for which inflected gerunds have been observed in monographic studies).
Periphrastic aspectual construction *estar* + gerund

As a variant of the standard periphrastic aspectual construction illustrated by (3a) and (4a), examples which involve inflected forms of *estar* ‘to be’ followed by the aspectual preposition *a* and the infinitival form of a main verb, some EP varieties display a periphrastic aspectual construction in which inflected *estar* combines with a gerund form of the main verb, (3b) and (4b).

3. a. Estão *a* arder
   are.₃PL a-ASP burn-INF
   ‘They are burning’

   b. Estão *ardendo* (PAL36)
      are.₃PL burn-GER

4. a. Essa pessoa estava *a* varrer
    that person was.₃SG a-ASP sweep-INF
    ‘That person was sweeping’

   b. Essa pessoa estava *varrendo* (LUZ08)
      that person was.₃SG sweep-GER

---

Figure 2: Geographic distribution of aspectual *estar* + gerund in CORDIAL-SIN

7 Besides the EP varieties under inspection, the gerund periphrastic construction is extensively found in Brazilian Portuguese. The construction involving *estar* is taken in this study as representative of the corresponding aspectual construction with other verbs, such as *ficar* ‘to remain’, *continuar* ‘to continue’, far less frequent in CORDIAL-SIN.
Through the CORDIAL-SIN data, the gerund variant appears consistently spread over the southern part of Portugal and in the islands (Figure 2).\(^8\)

The pronominal form *a gente* agreeing with 3PL verbal forms

Variation in verbal agreement with the pronominal subject *a gente* is frequent in EP varieties (for a recent survey, see Pereira (2003)). The pronominal form stems from a third singular phrase (literally ‘the people’) but has plural human reference, including the speaker. In popular varieties, this frequently correlates with occurrences of first person plural verb agreement with the subject *a gente*, while the standard variant displays third person singular agreement (examples (5a) and (5b), respectively). As CORDIAL-SIN data show, Azorean dialects further display a different pattern of agreement, involving third person plural, as illustrated in example (5c) (Pereira 2003).

5. a. *A gente* não dávamos nome nenhum àquilo.
*b* the people\(^{3SG}\) not gave\(^{1PL}\) name any to\(^{that}\) ‘We didn’t give any name to that.’

b. *A gente* não dava nome nenhum àquilo.
*b* the people\(^{3SG}\) not gave\(^{3SG}\) name any to\(^{that}\)

c. *A gente* não davam nome nenhum àquilo. (MIG08)
*b* the people\(^{3SG}\) not gave\(^{3PL}\) name any to\(^{that}\)

This regional variant is restricted to Azorean dialects. Among these, it appears with special frequency on the island of São Miguel, and only sporadically in some other locations, as shown differentially in Figure 3.

---

\(^8\) The presence of this variant at an isolated location on the northern border should be understood in a wider geolinguistic context, taking into consideration its distribution also in the Galician area (see Álvarez & Xove 2002:361).
Impersonal construction with existential verb *ter* ‘to have’

In standard EP, impersonal existential constructions involve a third singular form of the verb *haver* (etymologically ‘to have’, from Latin HABERE), as illustrated in (6a). The CORDIAL-SIN data also display semantically existential instances of the verb that express possession in standard EP, *ter* (‘to have’, from Latin TENERE).\(^9\) Thus, some EP varieties allow this existential use of *ter*, which occurs in impersonal constructions; regardless of the number displayed by the co-occurring constituent, the verbal form invariably displays third person singular agreement, as in example (6b).

6. a. Mas havia muitos moinhos por aqui fora.
   but ‘had’\(^{3SG}\) many water-mills\(^{PL}\) by here out
   ‘But there were many water-mills this way out.’

   b. Mas tinha muitos moinhos por aqui fora. (CLH03)
   but had\(^{3SG}\) many water-mills\(^{PL}\) by here out

In CORDIAL-SIN, the non-standard construction discussed here is mainly found in the insular territory, both in Azores and in Madeira (Figure 4).

---

\(^9\) This existential construction is commonly used in Brazilian Portuguese (see, for example, Avelar 2004).
Absence of article before prenominal possessives

The last examples concern the variable presence of the article before a prenominal possessive, a phenomenon often noticed as a matter of cross-linguistic and intra-linguistic variation (e.g. Mattos e Silva 1989; Rinke 2010). The variant in question corresponds strictly to contexts in which the standard variety displays the article, as in (7a) and (8a). In this case, the geographical spread of the non-standard variant, illustrated in (7b) and (8b) is not homogeneous, as presented more extensively in Carrilho and Pereira (2011).

7. a. E a *minha* tia lia aqueles livros.
   and the$_{FEM\cdot SG}$ my$_{FEM\cdot SG}$ aunt read those books
   ‘And my aunt read those books.’

    b. E *minha* tia lia aqueles livros. (GRJ06)
       and my$_{FEM\cdot SG}$ aunt read those books

8. a. Olha, fala com *o teu* avô
   look talk with the$_{SG}$ your$_{SG}$ grandfather
   ‘Look, talk to your grandfather.’

    b. Olha, fala com *teu* avô […] (CTL18)
       look talk with your$_{SG}$ grandfather

An important areal asymmetry emerges, however, which makes us consider this phenomenon among those non-standard constructions that are specific to certain EP areas. When one takes into account the relative frequency of the article before prenominal possessives (more accurately, before possessives preceding kinship nouns, as in the examples above), an evident geographical pattern
emerges: While the non-standard construction occasionally occurs from the northern to the southern area of continental EP,\textsuperscript{10} it is only in some insular areas that its incidence amounts to more than 65%, namely in the Madeira archipelago. The corresponding locations are represented in Figure 5.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure5}
\caption{Geographic distribution of absence of article before prenominal possessive (> 65\%)}
\end{figure}

**Syntactic areas and EP dialects**

The non-standard constructions considered in the previous section contribute to the definition of specific syntactic areas in EP. For each non-standard variant, a particular geographical distribution has been identified through CORDIAL-SIN. Although corpus data may, in certain cases, offer some evidence for constructions which are less frequent in a language, the constructions under study here could be systematically found throughout this EP dialect corpus. Moreover, the maps illustrate that the occurrence of each non-standard variant is geographically well delimited. Thus, in Figures 1 through 4, the blank areas represent homogeneous absence of the non-standard variant, while manifestations of this variant only appear and concentrate on the marked areas. As mentioned above, Figure 5 differs somewhat in the type of data represented. Nevertheless, even in this case, a clear geographic asymmetry arises when one takes into account the higher preference for the non-standard variant.

The homogeneous distribution of the non-standard variants may thus straightforwardly lead to the identification of isoglosses for the phenomena at

\textsuperscript{10} This geographical pattern (evocative of ‘leopard spots’) has been noticed as fairly recurrent in the case of syntactic variants (Poletto 2010).
On the areal dimension of non-standard syntax

On the areal dimension of non-standard syntax stake. Additionally, although the phenomena examined here display different geographical distributions, it is worth noting that all the syntactic areas delimited by these isoglosses are concentrated over the southern part of the continental territory and over the Portuguese islands, as represented in Figure 6.

![Figure 6: Summary of distribution of EP non-standard constructions in CORDIAL-SIN](image)

Significantly, the geographical distribution of these non-standard constructions corresponds to important dialect boundaries in EP which have been identified on the basis of regional phonetic variants. Thus, if one considers the boundaries within EP continental dialects, as proposed by Cintra (1971), it becomes clear that the phenomena inspected here are limited, in the continental territory, to the area of Central Inland and Southern Portuguese dialects (whose Northern limit roughly follows the course of River Tagus in Portugal). In addition, some of these phenomena, in particular those involving the gerund, are equally found in insular dialects, which reinforces the linguistic similarity between both areas, as previously noticed on the basis of phonetic variants (Cintra 1971).

Although Cintra’s proposal did not cover the insular territory, the identification of insular dialects is often acknowledged and important differences have been noted for dialects from Azores, from Madeira, and within each of these groups (e.g. Cintra 1990; Segura 2006). Crucially, some of the regional syntactic constructions considered here display an insular distribution which varies from generalized in both archipelagos (the non-standard existential construction), or restricted to one of them (the absence of article before possessive in Madeira), to
particular to one insular area whose dialectal identity has been noted independently (the case of third plural agreement with *a gente* in São Miguel).

In sum, it is possible to identify particular syntactic areas in the Portuguese territory, which has been accomplished by inspection of the geographic distribution of a selection of non-standard constructions in EP, as they occur through CORDIAL-SIN. Taking into consideration the particular areas thus identified, it must be observed that these syntactic variables strengthen important EP dialect limits which had been established on the basis of phonetic differentiation: (i) Central Inland and Southern dialects behave fairly differently from Central Littoral dialects and Northern dialects in the continental territory; (ii) Central and Southern dialects behave fairly similarly to insular dialects; (iii) insular dialects share some dialectal similarities; and (iv) inside the main dialect areas, some syntactic variants help to identify local varieties.

**Conclusion**

The work presented here emphasizes the relevance of dialect corpora as tools for the geolinguistic study of dialect syntax. The identification of syntactic areas in EP and the type of connection that these areas establish with EP dialects contribute to a twofold general conclusion. First, syntactic variation may be confined to cohesive dialect areas. Second, syntactic areas provide additional geolinguistic support to the tasks of identifying and delimiting dialects.

**References**


Carrilho, Ernestina and Sandra Pereira. 2011. Sobre a distribuição geográfica de construções sintácticas não-padrão em português europeu. *Textos Seleccion-
On the areal dimension of non-standard syntax


