TEDIPOR: Thesaurus of Dialectal Portuguese

Sandra Pereira & Raïssa Gillier

Abstract

The Thesaurus of Dialectal Portuguese (TEDIPOR) is a dialectal tool under construction. It is a fact that the changing and decreasing of the rural world and of its ways of life lead to the banning and to the extinction of a huge quantity of dialectal lexical variation. Therefore, TEDIPOR is a testimony of a disappearing lifestyle that will be preserved in a rich lexical database. It also aims to make available to the scientific community and to the society in general an important amount of dialectal, ethnographic and cultural information that is often difficult to access and handle. The sources to be integrated in the database include the glossaries of academic monographs (between 1940s and 1970s), atlases, dialectal inquiries and other papers containing dialectal information. In order to demonstrate the usefulness of this tool, all the designations concerning the concepts *bebedeira* (drunkenness), *bêbedo* (drunk) and *embebedar* (to get drunk) were gathered and are analysed under a lexical approach. The results demonstrate that many of the denominations found in TEDIPOR are not attested by Portuguese dictionaries, revealing that these materials are an important source for lexicographic research. Furthermore, the geographical distribution of the concepts *bebedeira* (drunkenness), *bêbedo* (drunk) and *taberna* (tavern) is also presented from a cartographic perspective. The maps show that it is possible to identify dialectal areas for some of the designations. Both the lexical and geographical analyses illustrate the potential of TEDIPOR, especially for Dialectology and Geolinguistics.

1. A dialectal tool

This paper focuses on the presentation of TEDIPOR,[1] *Tesouro Dialectal Português* (Thesaurus of Dialectal Portuguese), a dialectal database under construction. It will be integrated in a larger macro dialectal project that includes similar tools that are being developed to study Brazilian Portuguese and Galician: the *Tesouro do léxico patrimonial Galego e Português*. This partnership will allow the enrichment of the database designed for Portugal (cf. Álvarez *et al*., 2009).

TEDIPOR intends to assemble, in a single database, a great wealth of dialectal information from dialectal varieties spoken in Portugal. Due to social and technological changes in the last decades that are contributing to the disappearing of the traditional rural world, an uncountable number of dialectal words is gradually disappearing. Therefore, the first objective of TEPIDOR is to preserve an amount of lexical items, their meanings and geographic location that belong to a vanishing lifestyle. Secondly, it aims to gather and link together disperse and very heterogeneous (dialectal monographs, linguistic atlases, databases, etc.) information, that previous to the existence of the TEPIDOR database was difficult or even impossible to access.

Besides providing access to these materials, the database aims to preserve the contents of academic theses that for the most part have never been published, making them available to researchers (see, for instance, Álvarez Pérez, 2010, 2011). The third goal is to offer to the future users of the database the possibility of accessing and using the geographical location of data to build maps that manifest its regional distribution (cartography). This will be a valuable contribution to dialectal and geolinguistic studies, because it will make available a great amount of information that, up until now, was complicated to access and handle.

The paper is organized as follows: in section 2, we describe TEDIPOR's sources (2.1.) and the database's main structure (2.2.); section 3 is dedicated to the analysis of some data,

centring on two different aspects - the lexical approach (3.1.) and the cartographic approach (3.2.); finally, the conclusions are presented in section 4.

2. Presenting TEDIPOR

As TEDIPOR includes different kinds of information, we believe it will become a useful tool for researchers. In this section we will briefly present the database concentrating on its sources and its structure.

2.1. *TEDIPOR: sources*

The sources to be included in TEDIPOR are very wide-ranged. The database incorporates the whole contents of 64 dialectal monographs' glossaries, presented to the Universities of Lisbon and Coimbra between the 1940s and the 1970s.

The monographs are not similar in the way the information is presented. For this reason, a strict methodology is followed in order to make the materials homogeneous. Generally, the glossaries have lexical forms and their meanings, frequently illustrated by examples, refrains, songs, and so on. Occasionally there are opinions and sociological comments about the usage of some concepts in a specific municipality. Often, the monographs present pictures or drawings which show places, cults and customs, typical buildings or tools. All the ethnographic data is included in the database and will be available to the user.

Other sources will also be integrated in the TEDIPOR database. In particular, we are referring to data and information from linguistic atlases (namely ALEPG – *Atlas Linguístico-Etnográfico de Portugal e da Galiza*[2]), dialectal inquiries (*ILB – Inquérito Linguístico Boléo*), monographs and papers containing dialectal information (e.g. *Revista Lusitana* and others).

2.2. *TEDIPOR: a dialectal database*

The TEDIPOR database is designed to be a tool adjustable to the distinct users' needs, allowing different search criteria. It will also display an application for automatic cartography, as mentioned before. In this manner, the database structure has the necessary fields to collect and link together all the information from the sources to facilitate the search and representation of the materials. A brief description of the database main fields is given in the next paragraphs.

The original lexical forms (*variants*) are preserved in the same way they are reported in the original glossaries, in order to maintain all the phonetic, orthographic and morphological variations contained in the sources, for example *granizo, granisso, graniziu, granusiu, graisiu*, variants of the lemma GRANIZO (hail).

These variants will be lemmatized, so that the user can search for the lemma or for any of the existing variants. Lemmas will correspond to the standard form, grouping the different variants that might exist: for example, the lemma CHAMINÉ (chimney) / variants *chaminé, chaminei, chuminé, cheminé, cheminei…*

Due to the nature of the vocabulary in this project, for some *variants* there is no standard form in the language lexicon. In this case, lemmas will be 'built' from the collected variants and according to certain established criteria.

Some of the lemmas may be complex (ÁGUA DE PEDRA (stone water), BOLSA DE CAFÉ (coffee bag) and must be listed in the data base both under the multi-word expression itself and under the corresponding sub-lemmas (e.g. ÁGUA (water) and BOLSA (bag)).

Many of these academic monographs include phonetic transcriptions of the *variants*. The alphabets and conventions used in these transcriptions are highly variable, making it necessary to develop a system of correspondences in order to display this information unambiguously. Thus, we will work with a universal character font (IPA).

Usually, dialectal monographs collect information about word classes, categories and other relevant grammatical indication (*grammatical information*) with respect to the *variant*. This information is disposed in a short form (e.g. *s.m.* for masculine nouns, *v.t* for transitive verbs, *s.f.* for feminine nouns, *pl* for plural, *art. def.* for definite article, and so on) and it may be useful to observe morphological dialectal variation concerning the word's gender, for example.

Regarding the *definition* of the lexical items, all the information found in each source is transcribed literally. The entries in a glossary present the *variants'* meaning by a lexicographic definition or by a synonym; some offer the description given by the informant, usually extracted from the inquiry's recording.

In the monographs that have no indexed glossary, it is the context in which the *variant* appears that is transcribed. The same methodology is used when a word has not an independent entrance, making part of another word's definition.

Examples or *text fragments* that illustrate a *variant* are also disposed in the database, because this type of information may be helpful to know the adequacy of the definition or to pinpoint the syntactic environment of the defined lexical form.

Many monographs had an ethnographic purpose besides the linguistic finality of the academic work. Hence, definitions are frequently enriched with *idiomatic expressions* and *refrains/proverbs*. It is important to collect this type of information because it is relevant not only from a lexicographic perspective but also from a cultural one.

In the same way, *pictures* or *drawings*, usually related to specific semantic fields (agricultural tools, plants, animals, etc.), are abundant in the majority of the monographs and show, in many cases, a disappearing rural world. These materials will be incorporated into the database, linked to the correspondent lexical form (*variant*) and the related subject area.
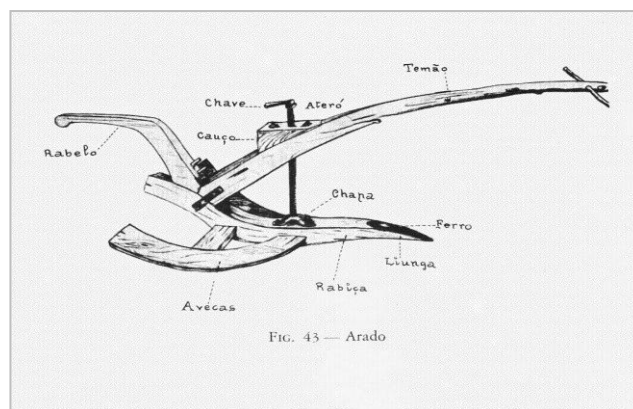

Figure 1. Example of a drawing illustrating the different parts of the plough.

All *variants* will be semantically classified. The assignment of a *semantic classifier* allows the search to be restricted to a particular semantic field (e.g. 'agricultural tools') from a

total of 75 semantic classes of the database. Furthermore this classification allows accessing the full lexicon belonging to a specific semantic domain (e.g. the semantic field 'weather' gathers concepts such as *chuva* (rain)*, neve* (snow)*, relâmpago* (lightning), etc.). As TEDIPOR will be incorporated into a major database, it will be possible to get the diverse denominations of the same concept in the three languages included in *Tesouro do léxico patrimonial Galego e Português.*

   Geo synonyms are lexical forms that associate different lemmas with the same meaning: QUEIXO, BARBA, MENTO, QUEIXADA, QUEIXAL, PAPO, BARBELA, BARBADELA, PERA, QUEIXELO, BARBAROTE, BICO (chin) allowing to obtain all denominations of the same concept.

   Each of the monographs is identified by a code that reveals the full details of the source: author, title, type of work, date, geographic location of data, and so on.

   All *variants* incorporated into the database are geographically identified (*geographical identifier*), so that the user can have access to the geographical location of a searched lexical form. Moreover, it is this geographical identifier that gives rise to the distributional representation of variants or lemmas in a map (cartography).

3. TEDIPOR data

Some of the common semantic fields available concern traditional agriculture (such as cereal crops production, the vineyard and the wine press, the olive oil and the oil press, the flax cultivation, etc.), agricultural tools and machinery, the mill and bread fabrication, among others. Other semantic fields describe physical and psychological characteristics of human beings, diseases or social relationships, to mention a few. To demonstrate the usefulness and richness of TEPIDOR, three concepts associated to the semantic field of wine and the vineyard were chosen: *bebedeira* (drunkenness), *bêbedo* (drunk) and *embebedar* (to get drunk). The different designations for these concepts evidence a social view of this kind of behaviour.

   On the other hand, the following data analysis intends to be a demonstration of two different approaches (lexical and cartographic) that can be studied with TEDIPOR's materials.

3.1. *Lexical approach*

The main goal is to list the different (dialectal) designations of the concepts *bebedeira, bêbedo* and *embebedar*. To achieve this purpose we looked at the monographs and ALEPG and grouped all terms/expressions defined as drunkenness, drunk and to get drunk. The second aim is to check if dictionaries register the different denominations that can be found in TEDIPOR. In this way, three reference European Portuguese dictionaries were chosen:

- the *Dicionário da Língua Portuguesa Contemporânea* (cf. Casteleiro, 2001), dictionary organized by the Academia das Ciências de Lisboa, published in 2001;
- the *Dicionário Priberam*, a well-known Portuguese dictionary, only available online[3];
- the *Dicionário Porto Editora*, another familiar dictionary, available in a printed and also in an electronic version.[4]

  Here we present the results of our investigation of three concepts both in TEPIDOR and in the mentioned dictionaries. The results presented in the following tables 1, 2 and 3 are listed by

decreasing number of occurrences. In order to facilitate the visualization of data, the words in white cells relate to the denominations registered in dictionaries with the same meaning; grey cells refer to the terms described by dictionaries with a different sense; and the designations not registered by dictionaries are preceded by an asterisk (*).

The analysis focuses on the semantic motivation for some of the designations, since it is not possible to analyze each of them. Furthermore, we intend to demonstrate the richness of the (new) vocabulary contained in TEDIPOR that is not registered in dictionaries and its potential relevance for Portuguese lexicography.

Starting with the concept *bebedeira*, Table 1 shows the designations found in TEDIPOR.

Table 1. Designations of *bebedeira* (drunkenness).

Designation	Frequency	Designation	Frequency	Designation	Frequency
bebedeira	144	raposa	1	briosa	1
borracheira	47	bezana	1	chumbeira	1
piela	21	embriaguez	1	gaita	1
cadela	12	estar com a pinga	1	trincadeira	1
perua	9	pinga	1	marmelo	1
carraspana	9	estar com a égua	1	brasa	1
tachada	4	perna	1	quentura	1
grossura	4	touca	1	tesa	1
cabra	4	rosca	1	repolho	1
trombose	4	nassa	1	careta	1
bêbeda	3	mona	1	ramada	1
gateira	3	senisga	1	*caçar a piela	1
pifão	3	camoeca	1	*canita	1
torta	3	moafa	1	*barco de chicharro	1
cardina	3	terebentina	1	*gachinha	1
turca	3	carrete	1	*camadão	1
rapioca	3	alegrete	1	*levar cadilhos	1
gatosa	3	mamona	1	*ter carga a mais	1
manta	3	tola	1	*cacharola	1
*jorca	3	bezerra	1	*zurra	1
piteira	2	pele	1	*rijona	1
chiba	2	chispa	1	*ruça	1
gata	2	torga	1	*pilhona	1
torcida	2	turina	1	*pelonha	1
carripana	2	serpentina	1	*taruça	1
verga	2	gaiola	1	*lapiana	1
serapilheira	2	toura	1	*maricuca	1
erisipela	2	senhora	1	*marrasqueira	1
carga	2	carraspeira	1	*piléria	1
*catrina	2	birra	1	*friosa	1
alegrão	1	má	1	*rucega	1
cabeleira	1	bicuda	1		

If we look at this table, we can observe that 63 per cent of the designations found in TEDIPOR are not registered with the meaning drunkenness in any of the three dictionaries, 40 per cent are listed in the dictionaries with another meaning and 23 per cent do not appear at all.

The most recurrent word in the corpus is *bebedeira*, probably the most generic word to name 'drunkenness'.[5] Other general designations are *borracheira*, *piela*, *embriaguez*, *carraspana* and *bezana*, although the last two are classified as colloquial, according to the dictionaries.

A common resource to designate 'drunkenness' is animal's names, like the following: *cadela* (female dog), *perua* (turkey hen), *cabra* (goat), *chiba* (young goat), *raposa* (fox), *senisga* (female piglet), *gata* (female cat) and *gateira* (cat door), formed from the nominal base *gata* + the suffix *–eira*. Other designations found in the corpus are *bezerra* (heifer), *turina* (frisian), *toura* (heifer), *mona* (female monkey), *(estar com a) égua* (mare) and **canita* (female dog). A reference to **canita* should be made, because this word and *cadela* reference the same animal (female dog). Nevertheless **canita* is a diminutive formed from the latin etymon *can + ito(a)*,[6] which is the reason why it wasn't grouped under *cadela*. An interesting fact is that all these animal designations take the female referent to preserve the feminine gender of the terms that designate drunkenness. Moreover, it is interesting to note that except for *perua* all the referred animals belong to the group of mammalian.

Another recurrent semantic field concerns the names of diseases. The semantic motivation for these terms is the relationship between (some of) the symptoms of illness and those felt while drunk or even those felt after (hangover). In this way, it is easy to understand the metaphor of *gatosa* (incontinence) or *carraspeira* (sore throat). Designations such as *trombose* (thrombosis) and *erisipela* (erysipelas) may be the result of the expansion of the primary motivation. The social vision of the consumption of alcohol as a misconduct leading to illness can also be pointed as the motivation for terms connected to diseases. Despite the fact that *terebentina* (turpentine) does not fit in this semantic group, there is a certain relation to the concept of health: it is a negative image of alcohol, since the consumption of turpentine is inadequate and toxic to the human being.

Many terms are related to the vineyard: *ramada* (trellis), *trincadeira* (kind of grape), **catrina* (kind of grape) and **gachinha* (bunch). The denomination **gachinha* is a diminutive form of **gacha* (**gacha + -inho(a)*), a transformation into the feminine gender of *gacho* (bunch). In what concerns **catrina*, it is the reduction of the form *catarina*, which is a kind of grape. Another word concerning the production of wine is *serpentina* (alembic tube).

Some words refer to the human body in allusion to the effect caused by alcohol. In this way, *perna* (leg) is a metaphor of the characteristic staggering of drunkards. It is not surprising that the remaining forms globally refer to the head - *tola* (head), *cabeleira* (hair), *torga* (big head) - as several effects relate to this body part (blurred vision, difficulty of speaking, heavy head, etc.). **Cacharola* is perhaps a variant of *cachola* (head) registered by dictionaries as 'popular'.

Although the designation *pele* (skin) belongs to the human body semantic field, it seems more appropriated to put it together with the following terms: *touca* (cap), *manta* (blanket), *serapilheira* (burlap), **camadão* (thick layer). All these terms can be grouped under the label 'coverings', prevailing the feeling of weight caused by drunkenness. Probably the first terms to appear referred to the head (e.g. *touca*) and then other words more wide-ranged came along. The same weight sensation can be attributed to the term *chumbeira*, probably derived from *chumbo* (lead).

Being off balance and losing equilibrium as consequence of drunkenness are the ideas behind the designations *torta* (twisted) and *rosca* (spiral). Other terms such as *carga* (load), **ter carga a mais* (too much load), *carripana* (heap), **barco de chicharro* (boat of blue jack mackerel), *carrete* (small car), **levar cadilhos* (take fringes) connect to the first ones referred, being associated to the notion of load and weight that causes an unstable walking.

A few terms like *rapioca* (partying), *alegrão* (great joy), *alegrete* (joy), and *gaita* (pipe) can be gathered under the semantic field concerning festivities, generically. What mostly triggers this group is the idea of joy and euphoria commonly felt by someone who drunk too much.

Birra (tantrum) is a denomination that takes into account the person's attitude and behaviour under alcohol influence. The physical symptoms caused by alcohol generate denominations as *brasa* (ember) and *quentura* (heat), especially related to the sensation of heat.

Designations such as *repolho* (round cabbage) and *marmelo* (quince) are obviously connected as both refer to horticultural products. Nevertheless the motivation for these terms is not easily perceptible because none of the referents is used to produce alcohol. A possible explanation might be the expansion of the global semantic field concerning the vineyard to other vegetables and fruits.

Moving to another concept, in Table 2 we present the designations for *bêbedo* (drunk).

Table 2. Designations of *bêbedo* (drunk).

Designation	Frequency	Designation	Frequency
bêbedo	153	tombado	1
borracho	49	almareado	1
borrachão	29	fragoso	1
embriagado	10	turino	1
relaxado	7	embargado	1
descarado	5	*coiraço	1
grosso	5	*mosquito de cartola	1
alcoólico	4	*estar cego e mudo	1
tarraço	3	*ir a cavalo na égua	1
drogado	2	*jorcado	1
bebedanas	2	*beberote	1
borracheiro	2	*clochar	1
*pielo	2	*afenafe	1
cuba	2	*monte de esterco	1
*estar encharcado na jorra	2	*monte de bagaço	1
caneco	1	*chingato	1
embebedado	1	*chambuço	1
égua	1	*pioleira	1
ébrio	1	*chimbeco	1
torto	1	*chimeco	1
carregado	1	*alambriado	1
pingado	1	*ataloado	1
bebedolas	1	*chumba	1
pegado à pinga	1	*não estar católico	1
coirão	1	*estar meio feijoteiro	1
pesado	1	*resinado	1
toino	1		

The table shows that only 30,2 per cent of the forms are registered in the dictionaries; the designations whose definitions in the dictionaries are not related to the name/adjective drunk are 26,4 per cent. So 43,4 per cent of the terms found in TEDIPOR to designate 'drunk' are not part of the dictionaries.

The variants which mostly occur are *bêbedo*, *borracho*, *borrachão* e *embriagado* (meaning drunk). Other words that exclusively designate 'drunk' are: *alcoólico*, *bebedanas*,

embebedado, *ébrio*, *bebedolas* and *pegado à pinga*. Some forms not attested with this meaning in Portuguese dictionaries, such as *beberote and *borracheiro*, are very similar to the former referred and their meaning can be easily achieved.

Many terms are usually applied for psychological characteristics, namely *descarado* (shameless), *relaxado* (slovenly), *toino* (idiot) and *almareado* (queasy). These designations evidence the attitude and behaviour of someone drunk.

As far as physical characteristics are concerned, the expression *estar cego e mudo (to be blind and dumb) and the words *grosso* (thick), *torto* (awry), *tombado (knocked down), *pesado* (heavy), *carregado* (loaded) and *chumba (the feminine of lead) express the idea of losing skills and the idea of weight.

Animal names are a common semantic field to designate a drunken person, as has been shown for drunkenness designations. *Égua* (mare) and *turino* (frisian) are some examples. The metaphor *mosquito de cartola refers to the little flies that go around the wine that is in *cartola*, the Azorian designation for this specific container (barrel). Other variants, like *tarraço* (churn), *caneco* (can) and *cuba* (vat) also belong to the semantic field of containers: *tarraço* and *caneco* are cups to drink by and *cuba* is a barrel where wine is poured.

The word *coirão* has two meanings as it is the name of a quality of grape and it also means drunk. *Coiraço seems to be a word derived from the former, although dictionaries do not register it.

The words *afenafe and *clochar are not present in dictionaries because there are loanwords: *afenafe is a construction formed from the English expression 'half and half' (meaning half drunk) and *clochar is based on the French term *clochard*, meaning tramp or homeless person.

For some words it is very difficult to trace its semantic motivation. Often because they are new words or words that are not recognized by a standard speaker. In some cases, we follow the suggestions of Kröll (1952-56) and (1963) but many of the designations are not included in his study.

Let's now look at table 3 that lists the designations for getting drunk.

Table 3. Designations of *embebedar* (to get drunk).

Designation	Frequency	Designation	Frequency
embebedar	6	*beber chá de Setembro	1
apanhar um pifão	2	*borrichar	1
andar na copofonia	1	*empielar	1
apanhar uma torta	1	*estar com ela	1
apanhar uma carraspana	1	*estar ligado à terra	1
estar com o petróleo	1	*estar chingarro	1
estar com os copos	1	*estar chingato	1
estar quente	1	*encatrinar-se	1
emborrachar-se	1	*puxar fogo	1
inchar-se	1	*puxar porção	1
*apanhar um repolho	1	*trabalhar no arame	1
*arranjar uma sopeira	1		

As can be seen, most of the expressions (56,52 per cent) are not registered in the dictionaries. Around 4,35 per cent of the variants are described in the dictionaries but not with the meaning of getting drunk. Less than 40 per cent of the forms can be found with coinciding definitions.

Concerning the verbal form *embebedar* (to get drunk), the results demonstrate that multi-word expressions, henceforth MWE, are much more frequent than one simple verbal form.

As far as single forms are considered, *embebedar* and *emborrachar-se* are recurrent terms that exclusively designate 'to get drunk'.

Although the verb **borrichar* is not attested by dictionaries, probably it is a variant of *emborrachar* (cf. Kröll, 1952: 24). The form **empielar* seems to be derived from the name *piela* (meaning *drunkenness*), a common Portuguese designation of drunkenness, as has been demonstrated by Table 1. The verb *inchar-se* means 'to get big or thick' (note that thick can mean drunk) but the dictionaries do not relate this verb to drunkenness.

In spite of not being listed as a dictionary entry, some of these MWE were considered attested: *apanhar um pifão* (to get plastered), *andar na copofonia* (to be at musical glasses), *apanhar uma torta* (to get a twisted), *apanhar uma carraspana* (to get a drunkenness), *estar com o petróleo* (to be with the oil), *estar com os copos* (to be with the glasses), *estar quente* (to be hot). The acceptation of these expressions is justified by the fact that they are composed of a copula verb - usually *apanhar* (to get), *andar* and *estar* (meaning 'to be' in these type of constructions) - and a name that is attested in dictionaries to designate drunkenness.

On the other hand, the expressions **apanhar um repolho* (to get a round cabbage), **estar com ela* (to be with her), **estar ligado à terra* (to be connected to the soil), **estar chingarro* (to be drunk) and **estar chingato* (to be drunk) were not considered to be attested because the nouns in these contributions do not designate drunkenness according to the dictionaries.

There are other expressions with transitive verbs, like **arranjar uma sopeira* (to get a maid), **beber chá de Setembro* (to drink September tea), **puxar fogo* (to pull fire), **puxar porção* (to pull portion) and **trabalhar no arame* (work at the wire). Once again, the dictionaries do not include these expressions.

Examining the three concepts, we conclude that the major part of the designations is not available in the dictionaries. With respect to the relationships between the three concepts it is possible to establish some connections.

For instance, *piela* is a very common designation of drunkenness described by all dictionaries but the person who gets a *piela* may be called of **pielo* and the act of drinking can be **empielar*. These variants are present at TEDIPOR but not at the Portuguese dictionaries. The same for *resina*, which means drunkenness, but **resinado*, the person who gets a *resina*, is not described as such.

Looking at words that do not occur in dictionaries, for instance, we have **catrina* and **encatrinar-se*. The first form is a noun which also means drunkenness, and the second one is the verb that describes the act of drinking. Both words are related to on this semantic field. If we look at **repolho* (round cabbage) and **apanhar um repolho* (to get a round cabbage) the same situation happens: **repolho* is a noun meaning drunkenness and the MWE includes that noun for meaning to get drunk.

Another parallel can be drawn between the forms not present in the dictionaries. For instance, someone who picks up a **chumbeira* (drunkenness) is called **chumba* (drunk). And also, someone is **jorcado* (drunk) when gets a **jorca* (drunkenness).

So, although the dictionaries do not refer to many of the terms found in TEDIPOR, they do exist in the Portuguese language maintaining semantic and morphological relationships with each other.

The brief description of the results of the three Portuguese concepts designating drunkenness, drunk and to get drunk demonstrates the abundance of 'new' dialectal data that is not reported in dictionaries. For each concept, TEDIPOR provides (i) terms presented in dictionaries (e.g. *bebedeira, bêbedo, embebedar*), (ii) words presented in dictionaries without this meaning (e.g. *trombose, cuba, inchar-se*) and (iii) new variants that do not appear in dictionaries at all (e.g. **taruça, *chambuço, *borrichar*).

3.2. *Cartographic approach*

In this section, we intend to show the geographical distribution of three concepts: *bebedeira* (drunkenness), *bêbedo* (drunk) and *taberna* (tavern). This last concept replaces the verb *embebedar* (to get drunk) which is absent in the ALEPG database and was chosen in order to maintain the same semantic area.

Since the number of designations and their occurrences are very high, we concentrate exclusively on data from ALEPG. Furthermore, it is not possible to present all the results for each concept. For this reason only designations that had more than four occurrences were mapped. The symbols in the maps represent an area of occurrence of a certain designation instead of a specific location.

The following map in figure 2 represents the distribution of the concept *bebedeira* and its designations in the Portuguese territory.
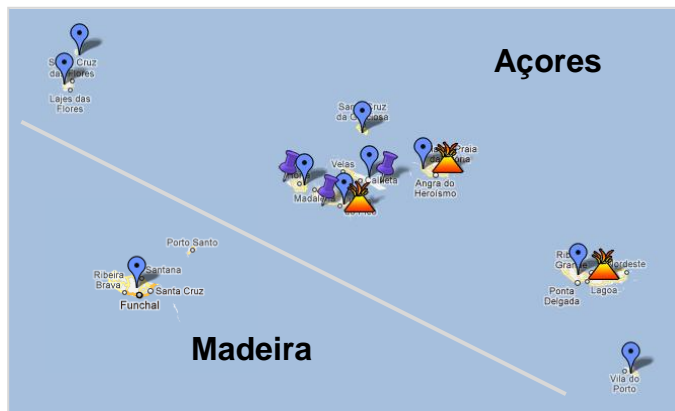
Figure 2. Geographical distribution of the concept *bebedeira.*

Bebedeira is the most widespread designation, occurring continuously and uniformly along the continental and insular territory, as the map above demonstrates.

In what concerns *borracheira*, it is found in the north, co-occurring with *bebedeira*. Although some isolated border points can be seen in the south, it is clear that this word is prevalent in the northern territory. Additionally, it does not occur in the islands.

The distribution of *piela* is not as well-defined as the two terms referred before. It seems to be two different areas: one in the northwest, more concentrated, from the interior to the coast; and another one in the centre that is less cohesive than the former, from the coast to the border with Spain.

The designation *cadela* appears in the central region, being absent from the extreme points. It can also be found in some Azorean islands.

Regarding the term *trombose*, its distribution is restricted to the Azorian islands.

Moving to the concept *bêbedo*, its distribution is mapped in figure 3.
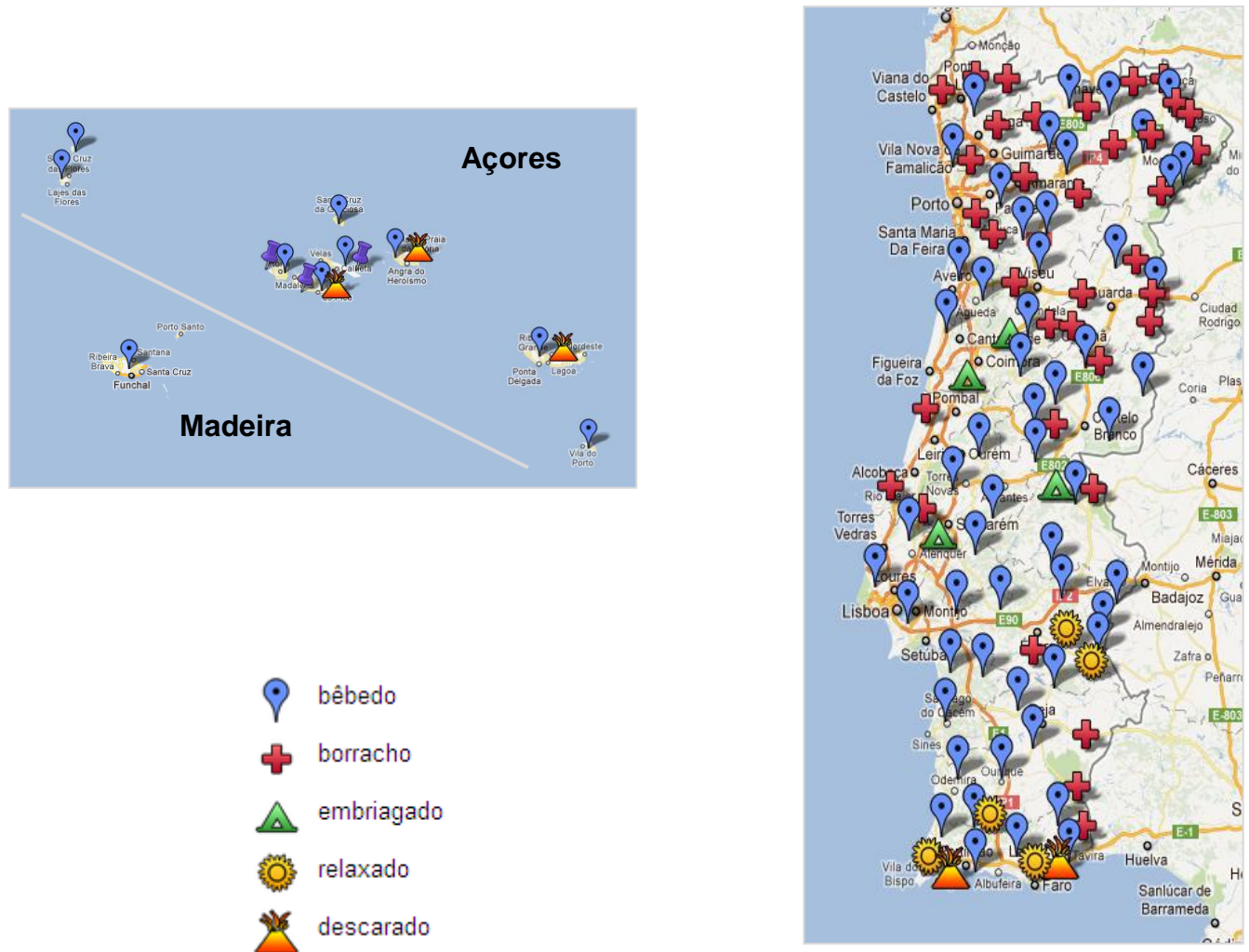
Figure 3. Geographical distribution of the concept *bêbedo*.

It is clearly visible that *bêbedo* is the most frequent designation since it is spread all over the country, including the islands.

The denomination *borracho*[7] forms an area extended to the centre of the country, being specially concentrated in the north. Nevertheless, it also occurs in some south border points, probably due to the influence of the Spanish word 'borracho' which also means drunk.

Embriagado appears only in the central region of the continental territory, forming an isolated although not cohesive area. Furthermore, it is also present in Azores and Madeira.

The occurrence of *relaxado* is limited to the southern region, even though it is not distributed in a continuous way.

With respect to the distribution of *descarado*, it is very restricted to the extreme south.

It is easy to establish a parallel between the two presented maps concerning the designations *bebedeira* and *bêbedo*. The distribution of this pair is similar since it is widely

spread and can be found in continental and insular territory. The same association can be made for the denominations *borracheira* and *borracho*: although the area of *borracho* is further extended to south, both concepts occur mainly in the northern region.

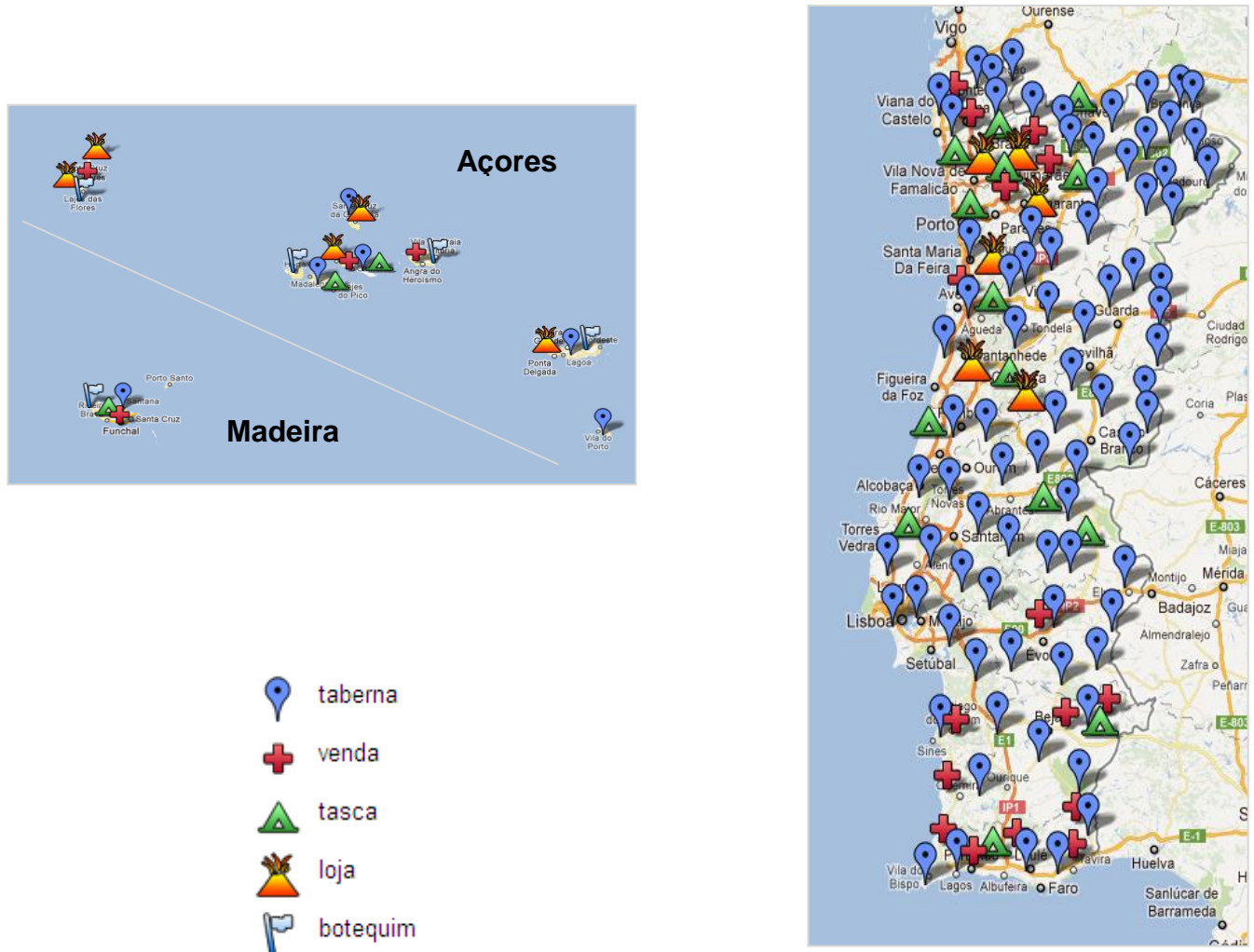The map in figure 4 presents the distribution of *taberna* in the Portuguese territory.

Figure 4. Geographical distribution of the concept *taberna.*

The designation *taberna* (tavern) occurs all over the territory, including both archipelagos.

Venda (store) forms two distinct areas: one in the northwest of the country and another one much more expanded in the south. In what concerns the islands, this term appears in Madeira and in three islands of Azores.

The word *tasca* (dive) appears as an extended band along the coast, from the Galician west border to the center. Some diffuse points can be seen in the south. In addition, it is also present in Madeira and in two islands of the central group of Azores.

Regarding the term *loja* (store), the map shows that this is a designation specific of the northwest. Despite of being more extended to coastal centre region, it co-occurs approximately with the northern area of *venda*. It can also be found in the Azorean territory, but it is absent in Madeira. Both *loja* and *venda* are designations related to a place where other products can be

bought, besides the sale and consumption of alcohol. On the contrary, *tasca* refers exclusively to a place for alcohol consumption.

Botequim (pub) is a designation characteristic of the archipelagos – Azores and Madeira – as it is not present in the continental territory.

The mapping of these three concepts reveals another TEDIPOR's possibility: it turns out to be an important tool to Geolinguistics because it allows the creation of maps and thereafter the identification of dialectal areas.

4. Final remarks

In conclusion, we are developing a tool that will be available to the scientific community and to the society in general, advantageous to several linguistic research domains: lexicography (especially regarding the incorporation of dialectal vocabulary in dictionaries and other lexicographic products), sociolinguistics, historical grammar, dialectology, standard language versus dialectal varieties, terminology, onomastics, and etymology, among others.

In this paper we tried to underline the importance of TEDIPOR by demonstrating that the TEPIDOR database contains a rich and detailed vocabulary that is not included in the 3 dictionaries consulted for this study. Furthermore, we have shown how TEPIDOR can be used to map concepts related to drunkenness in order to identify dialectal areas.

Additionally, the applicability of this database is not limited to the philological field, since TEDIPOR will be a valuable research tool to the study of people and technological changes, the dissemination of culture, geography and history.

Notes

[1] TEDIPOR is an ongoing project at Centro de Linguística da Universidade de Lisboa (CLUL) and Centro de Estudos de Linguística Geral e Aplicada (CELGA). It is funded by FCT (PTDC/CLE-LIN/102650/2008). For more information, please see: http://www.clul.ul.pt/en/research-teams/330-tedipor-thesaurus-of-dialectal-portuguese .

[2] For more information about ALEPG, please see: http://www.clul.ul.pt/en/research-teams/205-linguistic-and-ethnographic-atlas-of-portugal-and-galicia-alepg .

[3] http://www.priberam.pt/DLPO/

[4] http://www.infopedia.pt/

[5] We have tried to translate into English the Portuguese designations but sometimes that it is not possible, not only because some of the words are idiomatic but also because some of them are only used for expressing drunkenness.

[6] This is a frequent phenomenon in the south of Portugal for diminutive constructions (ex: *pan* – panito, meaning bread; *man* – manita, meaning hand).

[7] The variants *borrachão* and *borrachola* were gathered under the lemma *borracho*.

References

A. Dictionaries
Dicionário Priberam da Língua Portuguesa. 2012. http://www.priberam.pt/dlpo/.
*Infopédia – **Enciclopédia e Dicionários Porto Editora.** 2012. http://www.infopedia.pt/.
Casteleiro, M. (ed.) 2001. *Dicionário da Língua Portuguesa Contemporânea,* Lisboa, Editorial
    Verbo / Academia das Ciências de Lisboa.

B. Other literature

ALEPG, Linguistic and Ethnographic Atlas of Portugal and Galicia.
http://www.clul.ul.pt/en/research-teams/205-linguistic-and-ethnographic-atlas-of-portugal-and-galicia-alepg.

Álvarez, R., X. A Álvarez, J. Saramago and X. Sousa 2009. 'Presentación de un corpus digital de léxico tradicional: Tesouro do léxico patrimonial galego e portugués.' *Fonetică şi Dialectologie* 28: 5–19.

Álvarez Pérez, X. A. 2010. 'E os da banda d'alá son máis estranxeiros ca os de Madrí? Estudo contrastivo de designacións galegas e portuguesas no campo semántico da gandaría.' *Verba. Anuario galego de Filoloxía* 37: 57–88.

Álvarez Pérez, X. A. 2011. 'O léxico do namoro no Tesouro Dialectal Português (TEDIPOR).' Paper presented at XXVII Encontro Nacional da Associação Portuguesa de Linguística, Lisbon, October 26–28.

Kröll, H. 1952-1956. 'Designações portuguesas para 'embriaguez.' *Revista Portuguesa de Filologia* V, 1952: 27-87; VI, 1953–55: 73–135; VII, 1956: 17–118.

Kröll, H. 1963. 'Aditamentos às «Designações Portuguesas para 'Embriaguez'».' *Revista Portuguesa de Filologia* XII: 27–52.

***Revista Lusitana: Arquivo de Estudos Filológicos e Etnográficos Relativos a Portugal.* 1887–1943**. http://cvc.instituto-camoes.pt/bdc/etnologia/revistalusitana/index.html.