

The Syntactic Information in the 'Diccionario Crítico Etimológico Castellano E Hispánico' by J. Corominas and J. A. Pascual as Expressed in the Database 'Syntax.Dbf'*

Montse Batllori

Universitat de Girona. Departament de Filologia i Filosofia.
Plaça Ferrater i Mora, 1. 17071 Girona (Catalonia) Spain
batllori@skywalker.udg.es

Isabel Pujol

Universitat de Girona. Departament de Filologia i Filosofia.
Plaça Ferrater i Mora, 1. 17071 Girona (Catalonia) Spain
payet@skywalker.udg.es

Carlos Sánchez

Universitat Autònoma de Barcelona. Departament de Filologia Espanyola.
08193 Bellaterra (Barcelona) Catalonia (Spain)
carlos@filinf.uab.es

Received: October 25th 1996
Accepted: February 28th 1997

Abstract

This paper focuses on the structure of a syntactic relational database which includes all the syntactic information contained in the *Diccionario Crítico Etimológico Castellano e Hispánico*, by J. Corominas and J. A. Pascual. Owing to the fact that the dictionary was written in a way which escapes uniformity, it has been necessary to establish a series of fields in order to compile and accommodate the different kinds of knowledge related to the syntactic history of the lexical items. These fields are organized in five groups according to the subject of study: information related to the lemmata; grammatical aspects; syntactic information; other related knowledge; and relevant observations and comments provided by the researchers. Besides reproducing the data given in the dictionary, this type of database comprises several interpretative fields. Therefore, *Syntax.dbf* constitutes a basic tool for any researcher working on Spanish diachronic syntax.

Key words: diachrony, syntax, lexicography, computerization, database.

* Financed by DGICYT PB95-0656 (Institution: UAB), DGICYT PB96-1199-C04-01 (Institution: UAB), DGICYT PB96-1199-C04-03 (Institution: UdG), CIRIT Projecte1995SGR 0544 (Institution: UAB), CIRIT Projecte1995SGR 00486 (Institution: UAB).

Resum. *La informació sintàctica en el 'Diccionario Crítico Etimológico Castellano e Hispánico' de J. Corominas i J. A. Pascual a la base de dades 'Syntax.dbf'*

L'objectiu d'aquest article és presentar l'estructura de la base de dades relacional que inclou tota la informació sintàctica continguda en el *Diccionario Crítico Etimológico Castellano e Hispánico* de J. Corominas i J. A. Pascual. Tot i que aquest diccionari conté un ampli ventall d'informacions històriques de cadascun dels lemes, aquestes no es mostren de forma estructurada, per la qual cosa ha estat necessari estudiar i classificar tots aquells elements relacionats amb aspectes sintàctics. És a partir d'aquest estudi previ que s'han elaborat els diferents camps de la base de dades, els quals s'agrupen en cinc blocs temàtics: informació leemàtica; gramatical; sintàctica; altres aspectes relacionats; i observacions o comentaris rellevants fets per l'investigador. Aquesta base de dades no només reproduïx els continguts del diccionari, sinó que inclou diferents camps interpretatius. Per aquesta raó, *Syntax.dbf* representa una eina de treball fonamental per a tots aquells investigadors interessats en la sintaxi diacrònica de l'espanyol.

Paraules clau: diacronia, sintaxi, lexicografia, informatització, bases de dades.

Table of Contents

1. Introduction	5. Related information
2. A descriptive account of the database	6. Additional information
3. Grammatical information	7. Conclusion
4. Syntactic information	References

1. Introduction

As is well known, diachronic studies must rely on written data which must be extracted from texts. The main problem you must face is the great number of texts and data available, which implies having to make use of computerized systems in order to simplify the work of linguists when relating such an amount of data (cf. Marcos Marín (1994) and (1996), and Clavería and Sánchez (1996)). As for this, Wanner (1991: 179) states the following:

Extensive data collections are pivotal for the purpose, but their effective control is very laborious. [...] An obvious solution is to take advantage of electronic storage for reading and recall of data. The qualitative improvement from such operational assistance derives from the automated recall capacity extending over a representative corpus of texts. The combination of such an extensive data base with its increased accessibility due to appropriate coding and with flexible data recall can be utilized as a (very weak) analog to the unavailable native speaker in his/her role as an intuitive judge of the relative grammaticality/acceptability of the strings under investigation.

Therefore, any syntactic research based on written data must cope with computerized corpora. One way of doing it is by using syntactically and morphologically annotated texts (cf. Pintzuk (1996) for Old English, Kroch and Taylor (1996) for Middle English and Rojo (1992, 1993 and 1994) for Modern Spanish). The other, which is an indirect source, is by making use of the data that occur in lexicographic works. The latter, though, has not been exploited properly since there has been no work comparable to the annotated corpora. Despite of the fact that there have been computerized implementations of dictionaries, they have only led to the creation of CD-ROMS (cf. RAE (1995) and Moliner (1996)). In our belief, the lack of works comparable to annotated corpora is due to the fact that the dictionaries that have been computerized do not include enough examples and, when they do, there is no reference to the author and the literary or non-literary works from which the examples have been extracted. However, the history of lexicography provides us with several dictionaries which contain an annotated corpus (cf. Visser (1963) for the history of English, Rey (1993) for the history of French, and Cuervo (1886-1994), Corominas and Pascual (1980-1991), RAE (1960-) and Müller (1987-) for the history of Spanish). From the researcher's point of view, these works can be regarded as equivalent to the corpus extracted from annotated texts.

In accordance with this, the most important project of the *Seminario de Filología e Informática* (UAB) is the computerization of the *Diccionario Crítico Etimológico Castellano e Hispánico (DCECH)* by Corominas and Pascual (cf. Clavería, Blecua, Sánchez and Torruella (1995)). The main aim of this project is to obtain a series of relational databases that could contain the linguistic information given in the dictionary in order to provide an Hispanic etymological and grammatical data bank in computerized support which will include an independent program that allows queries and which will enable its use in CD-ROM and also will permit consultation by means of Internet.

Syntax.dbf is one of the relational databases mentioned above which compiles, interprets and classifies the syntactic aspects that are commented on in the dictionary.

2. A descriptive account of the database

This section and the corresponding subsections present a brief description of the fields which constitute the database. On the other hand, there will be given a detailed account by means of contrasting some examples of the words that appear in the dictionary with the actual translation into every field.

The database is structured in 20 fields that are organized in five thematic groups (see table 1 below): information related to the lemmata; grammatical aspects; syntactic information; other related knowledge; and all the relevant observations and comments provided by the researchers.

There are two types of fields to be taken into account: *transcription* fields (syntactic contents given by the *DCECH*) and *interpretation* fields (relevant aspects deduced from the text by the researcher or other authors). In the following sections you will find a detailed explanation of the content of these fields.

Table 1.

Field	Name	Type	Width
1	CONSTRGRAM	Character	50
2	LEMA	Character	25
3	SUBVOCE	Character	25
4	CLASEGRAM	Character	35
5	FUNCIOGRAM	Character	25
6	SENTIDGRAM	Character	25
7	PROBLEMAS1	Character	254
8	PROBLEMAS2	Character	254
9	PROBLEMAS3	Character	254
10	PROBLEMAS4	Character	254
11	TIPOPROBL	Character	100
12	DATAACION	Character	25
13	DOCUMENTOS	Logic	1
14	REFERENCIA	Character	200
15	LEMASREL	Character	254
16	LENGUAREL1	Character	254
17	LENGUAREL2	Character	254
18	DIALECTREL	Character	254
19	REFERENREL	Character	150
20	OBSERVACIO	Character	254
	Total		2719

2.1. Information related to the lemmata

In any dictionary the lexical information is distributed by means of entries (cf. Martínez de Sousa (1995), s.v. *entrada*) which are lexical items. These entries are the lemmata of the different databases in the computerization of the *DCECH*. Moreover, in the explanation given to every entry you can find other lexical items which have also been considered lemmata in the computerization. *Syntax.dbf* compiles grammatical lemmata and all the problems concerning them. Therefore it must relate syntactic lemmata with the entry of the dictionary and with the different entries of the general database of the *DCECH*, so that the lemmata in *Syntax.dbf* link this database with other relational databases. In this way, this database must have some fields which enable the connection with other relational databases. These fields are the ones which have been defined as containing the information related to the lemmata and they are the following: CONSTRGRAM, LEMA and SUBVOCE.

The following subsections will be devoted to technical descriptions of the content of the fields.

2.1.1. CONSTRGRAM (Grammatical Construction). It is composed by the items or grammatical structures (either morphological or syntactic) which appear in the dictionary, either as entries or as their content.

2.1.2. *LEMA (Lemma)*. The form as it occurs in the field called LEMA in the general database (LEMA.DBF) which has a lexical relationship with the syntactic lemmata that appear in CONSTRGRAM.

2.1.3. *SUBVOCE (Sub Voce)*. Entry of the DCECH where the syntactic information compiled in the field CONSTRGRAM is put together. To set an example, see the entry that corresponds to the lemma *dar* in the dictionary as shown in table 2 below.

Table 2.

En lugar de la 1ª pers. del Pres. de Ind. *doy*, es común en lo antiguo, como es sabido, *do* [...] se podría pensar en una aglutinación del adverbio *i* 'allí', lat. HIC (o IBI), empleado como sustituto del pronombre de dativo, partiendo de la idea de 'doy a aquel lugar', p. ej. 'a aquel monasterio' [Comp. *buscar hi* 'buscar para el convento' [...], *offresco hy mjo cuerpo* [...]] [...]; nótese que en los casos de *voy*, *soy* y *estoy* el significado del verbo correspondiente se presta asimismo a la aglutinación de un adverbio de lugar, lo mismo que a una propagación desde *doy*.

Instead of the 1st person of the simple present indicative *doy*, in past times, as is well known, *do* was the usual form [...] it could be argued that this corresponds to an amalgam of the adverb *i* 'allí', Latin HIC (or IBI), used as a substitute of the dative pronoun, bearing in mind that 'doy a aquel lugar' —I give to that place—, e.g. 'a aquel monasterio' —to that monastery— [Compare *buscar hi* 'buscar para el convento' —look for something for the convent— [...], *offresco hy mjo cuerpo* —I offer there my body— [...]] [...]; Notice that in *voy*, *soy* and *estoy* the meaning of the corresponding verb also favours amalgamation of the place adverb as well as a propagation from *doy*.

Table 3. Displays the way this information is organized in the database.

	Record 1	Record 2	Record 3	Record 4	Record 5	Record 6
CONSTRGRAM	doy	voy	soy	estoy	offresco hy	buscar hi
LEMA	dar	ir	ser	estar	ofrecer	buscar
SUBVOCE	dar	dar	dar	dar	dar	dar

3. Grammatical Information

This group contains three interpretative fields (CLASEGRAM, FUNCIOGRAM and SENTIDGRAM) where you find the grammatical aspects of the syntactic lemma. Notice that in the dictionary you cannot always find grammatical information. Whenever this information is found, it is interpreted as neutrally as possible and inserted in the corresponding fields. If there is no comment, the information to be inserted is provided by means of comparison with the other entries. As this database is thought to be a tool for research within different paradigms, the information contained is intended to be as neuter and general as possible. This criterion agrees with the general tendency when working with

computerized syntactic data, as stated by Wanner (1991) above and Panckhurst (1994: 39) in the quotation below:

[...], I propose that a DBQS [database querying system] in language sciences be designed according to linguistic principles and that its nature be directly determined by linguistic data. The divergent ways in which various linguistic schools conceptualize research should not constitute an obstacle for the DBQS. Linguists need to find a shared way to conduct intelligent querying, without denying the differences of individual approaches. Therefore, a DBQS should not only be determined by linguistic data but should also constitute a consensual query tool, thus providing a meeting point between varying linguistic descriptions.

3.1. CLASEGRAM (Grammatical Class), FUNCIOGRAM (Grammatical Function) and SENTIDGRAM (Grammatical Sense)

The field designated as CLASEGRAM includes the reference to the grammatical category (i.e. the comments which are abbreviated —those corresponding to the list of grammatical expressions given in the DRAE (1992)—). FUNCIOGRAM remarks the grammatical function of the syntactic lemma. SENTIDGRAM expresses the grammatical sense of the syntactic lemma. To set an example, compare the comments given in the dictionary (s.v. *donde*), as can be seen in table 4, with the contents of table 5 below.

Table 4.

Junto a *ond(e)* o *dond(e)* existió hasta el S. XIV *o* [...] y su sucedáneo *do* [...] el castellano preliterario distinguiría consecuentemente entre *o*, con idea de reposo o de movimiento por donde o hacia donde, y *onde* reservado para la idea de procedencia, como en latín. [...]. La variante *dond(e)* [...] acabó por contaminarse del doble valor de sus sinónimos *do* y *onde*. [La confusión descrita {la creación de *de dónde* para evitar la ambigüedad y el hecho de que entonces *onde* podía parecer una mera variante fonética y vulgar del más corriente *donde*} sería causa, mientras no se restableció el equilibrio sistemático, de que nacieran esporádicamente otros usos sintácticos no justificados históricamente. Así *donde* 'por donde' en *La Ilustre Fregona*]. Entre las innovaciones más modernas me limitaré a citar el empleo de *donde* como preposición en el sentido de 'en casa de', 'junto a' [...].

Coexisting with *ond(e)* or *dond(e)*, until the XIVth century there existed *o* [...] and its substitute *do* [...] consequently the preliterarian Castilian distinguished between *o*, meaning inactive state or movement through or towards, and *onde* which was reserved for the idea of origin, as in Latin [...]. The form *dond(e)* [...] eventually acquired the meaning of its synonyms *do* and *donde*. [The confusion which has been described {the creation of *de dónde* to avoid ambiguity and the fact that then *onde* could be regarded as a mere vulgar form or phonetic variant of the commonest *donde*} would be the origin of other sporadic syntactic forms which cannot be explained historically, while no systematic balance was reestablished. Thus *donde* 'por donde —through—' in *La Ilustre Fregona*]. Among the most recent innovations I will only mention the use of *donde* with preposition in the sense of 'en casa de —at somebody's house—', 'junto a —next to—'.

Table 5.

	Record 7	Record 8	Record 9	Record 10	Record 11
CONSTRGRAM	donde	donde	donde	de dónde	onde
LEMA	donde	donde	donde	donde	donde
SUBVOCE	donde	donde	donde	donde	donde
CLASEGRAM	rel. adv.	rel. adv.	rel. adv.	prep. + rel. adv.	rel. adv.
FUNCIOGRAM	link	prep. ¹	link	link	link
SENTIDGRAM	place	place	placelocative	source	place
	locative	locative	'por donde'	locative	'junto a'

1. Although links and prepositions have usually been treated as relational items of the same nature, the terminology used in this table and particularly in this case the terms 'prep.' is employed to express the exact information given by J. Corominas in the *DCECH* (otherwise, this information would be lost).

4. Syntactic Information

This group includes all the aspects related to the syntactic information given in the dictionary (the explanation of the syntactic phenomenon, chronology, documentation, and related bibliographical references), which enables the creation of a typology concerning the syntactic problems from a diachronic perspective.

4.1. PROBLEMAS1, PROBLEMAS2, PROBLEMAS3 and PROBLEMAS4 (Syntactic Phenomenon)

As stated above, one of the basic aims of SYNTAX.DBF is the compilation of the syntactic comments contained in the *DCECH*. Owing to the extension of such information in some of the lemmata, the field designated as PROBLEMAS has been organized in four fields (PROBLEMAS1, PROBLEMAS2, PROBLEMAS3, PROBLEMAS4) which display a unitary appearance. This division is due to the limitations of the software. As an example see tables 6 and 7. In the former, you are presented with one of the cases in which the four fields mentioned above are filled. Notice that the quotation corresponds to grammatical lemma *onde* in *Syntax.dbf* (cf. the dictionary s.v. *donde*;). In the latter, which is the grammatical lemma *dudar* in this database, the first one is enough to accommodate all the information given (cf. s.v. *dudar*):

Table 6.

PROBLEMAS1 junto a <ond(e)> o <dond(e)> existió <o>, y <do>. El castellano preliterario distinguiría entre <o>, con idea de reposo o de movimiento por donde o hacia donde, y <onde> reservado para la idea de procedencia, como en latín. El doble sentido de <do>, acab

PROBLEMAS2 ó por traer consigo el empleo indistinto de <onde>, aun con idea de reposo o de lugar hacia donde o por donde. Esto no ocurriría en seguida. Los ejs. de <onde> con este valor son raros al principio y sospechosos, pero esto ya ocurría en el S. XIV (unos p

Table 6. (*Continuation*)

PROBLEMAS3 pocos ejemplos de <on> o <don> pueden ser debidos a los copistas del S. XIV, al leer <ō> en vez de <o>. Desde la creación de <de dónde>, <onde> podía parecer una mera variante fonética y vulgar del más corriente <dónde> y estaba condenado a una pronta de

PROBLEMAS4 desaparición en el idioma culto; la forma actual <onde> (y sus contracciones <ande>, <ponde> y <pande>), frecuentes en el habla del vulgo de todas partes, es difícil asegurar lo que tengan de arcaísmos o alteraciones fonéticas según el tipo <ejar> por <dejar>.

PROBLEMAS1 coexisting with <ond(e)> or <dond(e)> there existed <o>, and <do>. The preliterarian Castilian distinguished between <o>, meaning inactive state or movement through or towards, and <onde> which was reserved for the idea of origin, as in Latin. The double meaning of <do>, eventually

PROBLEMAS2 led to the indistinctive use of <onde>, even with the meaning of inactive state or place towards or through. This would not take place immediately. The examples of <onde> with this meaning are strange at the beginning and suspicious, but this already happened in the XIVth century (a

PROBLEMAS3 few examples of <on> or <don> could be attributed to the people who copied the manuscripts in the XIVth century, when reading <ō> instead of <o>. From the creation of <de dónde>, <onde> could seem a mere vulgar form or phonetic variant of the commonest <dónde> and was condemned to early

PROBLEMAS4 disappearance in the formal language; the present form <onde> (and its contractions <ande>, <ponde> and <pande>), frequent in popular speech from everywhere, it is difficult to state to what degree they are arcaisms or phonetic alterations following the pattern of <ejar> for <dejar>

Table 7.

PROBLEMAS 1 construcciones y acepciones especiales.

PROBLEMAS 1 special constructions and meanings.

4.2. TIPOPROBL (Type of Phenomenon)

It is an interpretative field which helps typological classification of the syntactic problems of each lemma according to the list given in González Pérez and Rodríguez Fernández (1989: 11-14). As has been observed in the previous example with reference to *donde* and its syntactic information, you can deduce that the type of problems related to this item can be classified according to two different thematic aspects. On the one hand, it refers to place adverbs and their status and, on the other hand, to links (as stated in this list). The following table exemplifies the sort of references introduced in order to obtain the above mentioned classification. Notice that it corresponds to the comments concerning record 11 whose grammatical lemma is *onde* (cf. s.v. *donde*).

Table 8.

TIPOPROBL	adverbios de lugar - los elementos de relación (S34114-S37).
TIPOPROBL	place adverbs - the items of relation or linking (S34114-S37).

4.3. DATACION (Chronology)

In this field you can find the comments of which PROBLEMAS or DOCUMENTOS consist of and which are related to the date when the different phenomena took place. The date is systematized in accordance with the criteria followed in DOCU.DBF, which is one of the relational databases that contain the information given in the *DCECH*. The chronological history of the item or structure is extracted in order to supply a means to make consultation easier for scholars interested in these aspects.

For example, starting with the following syntactic information as it occurs in the *DCECH* (s.v. *donde*):

Table 9.

La variante *dond(e)* se crearía primero como mero refuerzo enfático de *onde*, [...], después quedaría como única expresión inequívoca de la idea de procedencia, y en definitiva acabó por contaminarse del doble valor de sus sinónimos *do* y *onde* [...]. Pero el doble valor sigue vigente no sólo en el S. XV [...] sino aun en el XVI [...] y el XVII [...] [e incluso, por tradición literaria, en verso, se ha seguido admitiendo esta posibilidad hasta h. 1800].

The form *dond(e)* was possibly created first as a mere emphatic reinforcement of *onde*, [...], afterwards it would remain as the unique unequivocal expression for the idea of origin, and eventually acquired the double meaning of its synonyms *do* and *onde* [...]. But the double meaning was still in use not only until the XVth century [...] but also in the XVIth yet [...] and the XVIIth [...] [and even, because of literary tradition, in verse, this possibility has been admitted until around 1800].

You can translate the references to date of use of the lemma as it follows:

Table 10.

DATACION 1800HS

As for the way the date is expressed in the database, it must be said that it is subjected to a systematization established for the sake of clearness (see table 11).

Table 11.

anterior al	'previous to'	<i>fechaA</i>
antes de	'before'	<i>fechaA</i>
después de	'after'	<i>fechaD</i>
desde (fecha)	'from (date)'	<i>fechaDD</i>
desde finales de (fecha) o (fecha)	'from the late' '(date) or (date)'	<i>fechaDDF</i> 1925/1936
finales de siglo	'turn of the century'	1298_1299
finales siglo XIII principios XIV	'end of XIIIth beginning of XIVth century'	1299-1301
hacia	'about'	<i>fechaH</i>
hasta	'until'	<i>fechaHS</i>
mediados siglo XIII	'middle of XIIIth century'	1249_1251
primer cuarto	'first quarter'	1200_1225
primer tercio	'first third'	1230_1231
primera mitad	'first half'	1200_1250
princ. Siglo XIII	'beginning of XIIIth c.'	1200_1201
segunda mitad	'second half'	1250_1299
segundo cuarto	'second quarter'	1225_1250
tercer cuarto	'third quarter'	1250_1275
siglo XIII	'XIIIth century'	1200_1299
siglos XVI y XVII	'XVIth and XVIIth century'	1500_1699
siglo de oro	'Golden Age'	1500_1650
+ (fecha)	'+ (date)'	<i>fechaAM</i>

To set another example, take the comments related to *deber de* (s.v. *deber*) that appear in the dictionary:

Table 12.

[...] la construcción *deber de*, en el Siglo de Oro denota indiferentemente obligación, o bien inducción o conjetura; la distinción entre *deber de* con el último y *deber* con el primero de estos valores fue introducida por los gramáticos del S. XVIII.

The construction *deber de*, in the Golden Age meant either obligation, or induction or conjecture indifferently; the distinction between *deber de* with the latter and *deber* with the former of these meanings was introduced by the XVIIIth century grammarians.

This is expressed as follows:

Table 13.

DATAACION 1500-1650; 1700_1799

4.4. DOCUMENTOS (Documents)

Logical field which indicates whether the dictionary includes 'documents' in which all references to authors and bibliography where the syntactic lemma occurs are expressed. These documents are to be found in DOCU.DBF.

4.5. REFERENCIA (References)

This field includes the bibliographical references about the different syntactic aspects given by the dictionary. In (10) and (11) display the contents of this field for *dudar* (s.v. *dudar*) and *doy* (s.v. *dar*), respectively.

Table 14.

REFERENCIA	Cuervo, <Dicc.> II, 1333-36;
------------	------------------------------

Table 15.

REFERENCIA	Cuervo, <Dicc.> II, 726-59; Staaff, 321
------------	---

5. Related Information

This group is intended to compile the knowledge concerning the relation of the lemma with either other lemmata of the language and dialects or even of other languages. It also contains bibliographical references concerning the related lemmata.

5.1. LEMASREL (Related Lemmata)

This field consists of all those lemmata which are related with the item or syntactic structures that appear in the same entry of the dictionary. The lemmata are all part of the general relational database called LEMA.DBF. See the example of *onde* (s.v. *donde*), corresponding to record 11:

Table 16.

LEMASREL	o, (II)
----------	---------

5.2. LENGUAREL1 and LENGUAREL2 (Related Languages)

The field called LENGUASREL has been extended up to two fields because of the lack of space. It contains information concerning those related languages and documents (i.e.: authors and bibliographical references, mainly) which are mentioned in the comments on a particular syntactic problem or aspect. See table 17 for *doy* (s.v. *dar*) and 18 for *deber de* (s.v. *dar*).

Table 17.

LENGUAREL1 en textos gallegoportugueses <dou>, o en forma más arcaica <dau>, del lat. vg. DAO;

LENGUAREL1 in Galician-Portuguese <dou>, or in the more arcaic <dau>, from vulgar Latin DAO.

Table 18.

LENGUAREL1 en portugués antiguo: (<Canc.> de D. Denis, v. 304); en gallego moderno «obligación»: os vellos non <deben de> namorarse, título de farsa de Castelao, 227.6; <dever fazer> se construía <dever a fazer> en gallego medieval (V. las <Ctgs.> de Guilhade o de

LENGUAREL2 D. Denis): devo a leixar aquele casale pobrado con outra tanta poblanza, devo-a leixar... a. 1326, Pontevedra (Sarm. <GaG.> 87v)

LENGUAREL1 in Old Portuguese: (<Canc.> by D. Denis, v. 304); in Modern Galician «obligation»: os vellos non <deben de> namorarse —the elder must not fall in love—, as the title of a masque by Castelao, 227.6; <deber fazer> was usually constructed as <deber a fazer> in Medieval Galician (V. the <Ctgs.> by Guilhade or by

LENGUAREL2 D. Denis): devo a leixar aquele casale pobrado con outra tanta poblanza, devo-a leixar ... —I must leave that place inhabited by so many inhabitants, I must leave ...— a. 1326, Pontevedra (Sarm. <GaG.> 87v.)

5.3. DIALECTREL (Related Dialects)

This field displays the comments concerning the different dialects and their documents. Notice that they are always related to a definite syntactic lemma. In table 19, you can see the comments on the grammatical lemma *doy* (s.v. *dar*), whereas in table 20 you have the information of *dar (las) horas* (s.v. *dar*).

Table 19.

DIALECTREL en textos leoneses <dou>, o en forma más arcaica <dau>, del lat. vg. DAO

DIALECTREL in Leonese texts <dou>, or in the more arcaic <dau>, from vulgar Latin DAO

Table 20.

DIALECTREL en algunas partes, como en Chile, todavía existe conciencia de la interpretación de <(las) horas> como complemento directo, pues se dice <las han dado, las dieron, las darán>

DIALECTREL in some places, as in Chile, there still exists the conception of the sense of <(las) horas> as the direct object, for <las han dado, las dieron, las darán> is commonly said.

5.4. REFERENREL (Related References)

It includes the bibliographical references from which syntactic information about the languages and dialects mentioned in the previous fields can be extracted. For example, Bello, in his grammar, talks about the expression *dar (las) horas*:

Table 21.

REFERENREL Bello, <Gram.>, ed. 1936, epígrafe 777n.

6. Additional Information

This information is compiled in a group by itself which consists of a single field to be known as OBSERVACIO whose function is to enable the introduction of additional marginal comments either extracted from the words of Corominas himself (see table 22: *onde*, s.v. *donde*) or considered relevant by the researcher (see table 23: *denante*, s.v. *delante*).

Table 22.

OBSERVACIO la historia del uso de los dos vocablos concurrentes (<ond(e)> o <dond(c)> y <o> o <do>) y sus variantes, y de la evolución de sus sentidos, es complicada de por sí, con la agravante de la desconfianza que deben inspirarnos muchas ediciones.

OBSERVACIO the history of the use of both concurrent terms (<ond(e)> or <dond(e)> and <o> or <do>) and their variants, and the evolution of the meanings, is intrinsically problematic, with the additional skepticism that many an edition must bring forward to us.

Table 23.

OBSERVACIO No hay suficientes datos para postular la existencia de delante, (I) y (II). Se trata del mismo lema con dos funciones sintácticas diferentes.

There are not enough data to postulate the existence of *delante* (I) and (II) 'in front of/before' It is the same lemma with two different syntactic functions.

7. Conclusion

This paper has offered a description of the structure and contents of *Syntax.dbf*. This database is a basic tool for diachronic research due to the lack of Old Spanish annotated texts, because it also has the advantage of being both one of the relational databases of the *Diccionario Crítico Etimológico Castellano e Hispánico* and a corpus of the history of Spanish and related languages and dialects.

References

- Clavería, G.; Blecua, J.M.; Sánchez, C.; Torruella, J. (1995). *Bases para la informatización del DCECH de J. Corominas y J. A. Pascual*. Unpublished Manuscript. UAB, Bellaterra.
- Clavería, G.; Sánchez Lancis, C. (1996). «La aplicación de las bases de datos al estudio histórico del español». *Hispania* (in press).
- Corominas, J.; Pascual, J.A. (1980-1991). *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos (BRH, *Diccionarios*, 7).
- Cuervo, R.J. (1886-1994). *Diccionario de construcción y régimen de la lengua castellana*, 8 volumes, continued and editado by the ICC. Bogotá: Instituto Caro y Cuervo.
- González Pérez, R.; Rodríguez Fernández, A.Mª (1989). *Bibliografía de sintaxis española (1960-1984)*. *Verba* (Anexo 31).
- Kroch, A.; Taylor, A. (1996). «The Penn-Helsinki Parsed Corpus of Middle English». Unpublished Manuscript. University of Pennsylvania, presented at the *Fourth Workshop on Corpus Compilation, Tagging and Exploitation*; 28 May 1996, Institut Universitari de Lingüística Aplicada.
- Marcos Marín, F. (1994). *Informática y humanidades*. Madrid: Gredos.
- (1996). *El comentario filológico con apoyo informático*. Madrid: Síntesis (*Teoría de la Literatura y Literatura Comparada*, 17).
- Martínez de Sousa, J. (1995). *Diccionario de lexicografía práctica*. Barcelona: Vox Bibliograf.
- Moliner, Mª (1996). *Diccionario de uso del español*, CD-ROM. Madrid: Gredos.
- Müller, B. (1987-). *Diccionario del español medieval*. Heidelberg: Carl Winter Universitätsverlag.
- Panckhurst, R. (1994). «A Database for Linguists: Intelligent Querying and Increase of Data». *Computers and the Humanities* 28 (1): 39-52.
- Pintzuk, S. (1996). «Annotating the Helsinki Corpus: Design and Implementation. The Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English». Unpublished Manuscript. University of York, presented at the *Fourth Workshop on Corpus Compilation, Tagging and Exploitation*; 28 May 1996, Institut Universitari de Lingüística Aplicada.
- Real Academia Española (1960-). *Diccionario Histórico de la Lengua Española*. Madrid: Real Academia Española.
- (1995). *Diccionario de la lengua española*, 21st edition, electronic support. Madrid: Espasa Calpe.
- Rey, A. [dir.] (1993). *Dictionnaire historique de la langue française*. Paris: Dictionnaires Le Robert.

- Rojo, G. (1992). «El futuro *Diccionario de construcciones verbales del español actual*». In: Martín Vide, C. (1992). *Actas del VIII Congreso de Lenguajes Naturales y Lenguajes Formales*. Barcelona: PPU, p. 41-50.
- (1993). «La base de datos sintácticos del español actual». *Español Actual*, 59: 15-20.
- (1994). «Problemas lingüísticos e informáticos en los diccionarios de construcción y régimen». In: *Actas del congreso de la lengua española (Sevilla, 7 al 10 de Octubre de 1992)*. Madrid: Instituto Cervantes, p. 307-315.
- Visser, F.Th. (1963). *An Historical Syntax of the English Language*, 3rd impression. Leiden, The Netherlands: E. J. Brill, vol. I.
- Wanner, D. (1991). «Historical Syntax and Old Spanish Text Files» In: Harris-Northall, R.; Cravens, Th.D. (eds.). *Linguistic Studies in Medieval Spanish*. Madison: The Hispanic Seminary of Medieval Studies, p. 165-190.