Dialectologia 3 (2009), 23-51.

DIALECT GRAMMAR OF SPANISH FROM THE PERSPECTIVE OF THE *AUDIBLE CORPUS OF SPOKEN RURAL SPANISH* (OR *CORPUS ORAL Y SONORO DEL ESPAÑOL RURAL, COSER*)

Inés Fernández-Ordóñez
Universidad Autónoma de Madrid
ines.fernandez-ordonnez@uam.es

Abstract

The *Audible Corpus of Spoken Rural Spanish* (or COSER after its Spanish abbreviation) is a corpus of oral interviews which aims to study dialect grammar in the Iberian Peninsula. In this paper COSER characteristics and methodology are described and compared to atlases regards the research of dialect grammar. Thanks to COSER, a number of Spanish dialect syntax aspects which were partially known or fully ignored have been researched: several pronominal paradigms, subjunctive displacement, and mass neuter agreement. Moreover, the geographical distribution of these aspects has been sometimes considerably broadened and traditional explanations have been replaced by new ones based on a better knowledge of the data. The study of dialect grammar has revealed as an important source for a better understanding of many cross-linguistic principles, like the Agreement Hierarchy, and opens up new ways to test their validity.

Key words

Corpora of oral interviews vs atlases, dialect grammar, pronominal paradigms, subjunctive displacement, mass neuter, agreement hierarchy.

Until recently, the study of dialectal variation of Spanish in the Iberian peninsula has been based on various regional atlases and scarce dialectal monographs which devoted particular attention to Castilian (in contrast to the more numerous ones focused on the Asturian-Leonese and Aragonese linguistic domains). Both in atlases and monographs, dialectologists pay more attention to phonetic and lexical variation than to grammatical variation and data have usually been collected by means of a questionnaire. The *Audible Corpus of Spoken Rural Spanish* (referred hereafter as the Spanish abbreviation *COSER*) is a

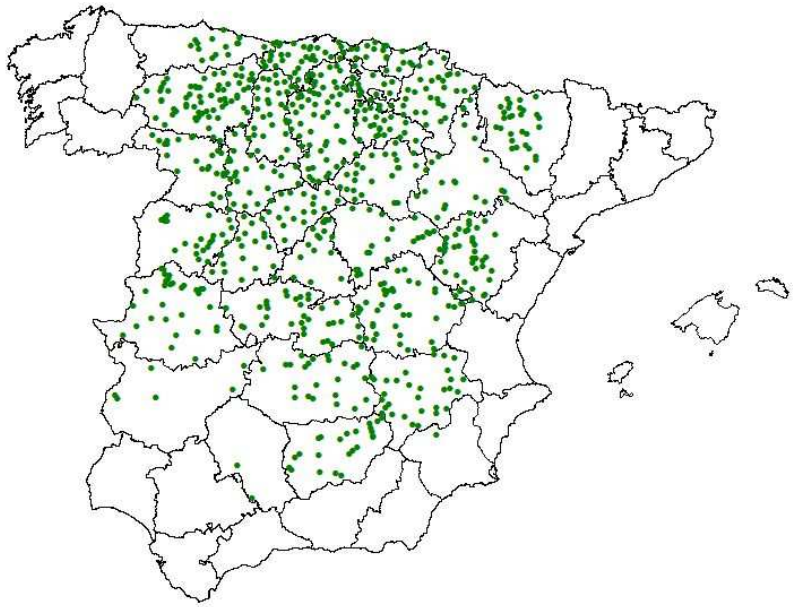
©Universitat de Barcelona

corpus made up by recordings of rural speech which started to be compiled in 1990 to be complemented to those traditional sources and has been expanding since then.

1. Description and methodology

COSER is a corpus restricted to the speech of informants who were deemed interesting for traditional dialectology: rural speakers, elder if possible, of low education and natives of the place where they were interviewed. Actually, *COSER* has the same type of informants as linguistic atlases and many dialectal monographs, although its methodology and objectives are different. So far, (i. e. the year 2009), 1,408 informants have been recorded, among whom 44% were men and 55,9%, women. The average age of the informants was 72.9 years, being slightly higher in men (73,8) than in women (72).

Interviews have been carried out so far in 754 rural enclaves of the Iberian Pensinsula, mainly in the centre and the north. As shown in map 1, the point density is comparable to that of regional atlases, or even denser.



Map I. Geographic distribution of *COSER* enclaves (2009).

©Universitat de Barcelona

COSER consists presently of 940 hours of recording, but this number increases every year thanks to new survey campaigns. The final objective is to obtain recordings of Spanish spoken in rural areas of the whole Iberian Peninsula.

The methodology used in *COSER* has consisted in sociolinguistic interviews, aimed by part of the interviewers at some subjects of traditional country life. The fact that the interview is focussed on such specific subjects does not prevent that, after some time and having gained the informant's confidence, interest is aimed at other subjects, such as education, personal hopes and experiences, life or family, depending on the level of easiness and spontaneity shown by the informant. The decision of focusing the interview on specific subjects related to rural life "of former times" has much to do with the fact that, in order to accept to be interviewed, potential informants must prove to have some knowledge about a way of life in decline. This knowledge is a product of their own personal experience and age and gives them informative "authority" in front of the urban interviewer. Informants accept the interview as they realize that we are interested in a testimony on a way of life in decline about which very few have hardly any memory at all and which they know they are expert on. We think that the informants' spontaneous cooperation would be much more difficult if they would be required at first to be interviewed on personal views or experiences, linguistic matters or other aspects beyond rural life. The fact that the interviewing team has insisted on their specific interest in the strictly local tradition, in contrast to that of other rural enclaves, as well as in the exclusive informant's condition as recipient of such tradition, has been on many occasions a decisive factor for accepting the interview.

Informants are always randomly contacted, with no previous actions, among the local inhabitants fulfilling the above mentioned requirements. Due to the experience, not much gratifying, of some interviews on account of the informants' low communication ability (people not much willing to speak, who answered with very short sentences or just in monosyllables) led us to add subsequently the condition of loquacity ("that the informants like talking") to the informants' selection protocol. As it will be obviously well-known to anyone who has ever carried out fieldwork, success is never assured, and an interview starting under the same conditions may be optimum or dreadful. Thus, not all interviews are equally suitable or informative, depending on the informants' willingness, the interviewers'

©Universitat de Barcelona

skills as well as the interaction between them; however, no testimony should be disregarded for that reason.

This methodology can not avoid the problem of accommodation between the informant and the interviewer, or the challenging representativeness of the informant ramdomly chosen. Nevertheless, we think that the quantity of the data allows to circumvent these potential problems, since the data always show geographical coherence and make it possible to discard those informants who could be considered anomalous with their area.

Regarding the number of informants of each enclave, in general one single person has preferably been thoroughly interviewed in *COSER*, either a man or a woman. Nevertheless, recording conditions have sometimes not allowed to avoid interruptions from other individuals (generally members of the family or acquaintances who, drawn by such an extraordinary event as the interview, cannot resist the temptation to take part in the interview by giving their own testimony.) Thus, although up to 1,408 informants have been recorded in *COSER*, most of the times only one informant per enclave has actually been thoroughly surveyed as desired (almost the half).

The average duration of the recordings is one hour and fifteen minutes (75 minutes) per enclave, although it may range from just half an hour up to more than two hours and a half. The quality of the data recorded is not directly proportional to the duration, since there are excellent and very informative recordings of just half an hour, whose results are comparable to those obtained in a longer session.

2. Which is the use and contribution of *COSER*?

2.1. Linguistic distances measurable in COSER

COSER is a corpus aimed to measure the differences which may be found in the speech of sociocultural groups with a lower education in rural areas. It is therefore a complement to both linguistic atlases and to the different corpora of cultivated and urban speech which have been compiled or are planned to be so in the Spanish-speaking world. The uniformity in the methodology used makes it useful to measure both the linguistic distance which separates

©Universitat de Barcelona

different areas (physical distance) and the linguistic distance which separates this social group from others, like for instance, that of speakers with a higher sociocultural level or that of younger speakers (social distance). Although the proportion of men and women interviewed is not identical (55,9 % women vs. 44 % men), the number of speakers of each gender is statistically representative and also allows to investigate linguistic differences associated with gender.

The fact that the media are the sources of most Spanish oral corpora lends some singularity to the *COSER*, since the interviewed speakers for *COSER* are rarely recorded in this field. The comparison between the data obtained in *COSER* and in other corpora of spoken Spanish enables thus to point out clear sociocultural differences.[1] In this regard, *COSER* has proved especially useful since it provides the study of non-standard grammatical solutions, which are usually systematically avoided in written language and in the speech of sociocultural groups of higher education. For that reason, Chambers (1995) has proposed, as a sociolinguistic universal, the qualitative character (presence/absence) of grammatical variables in the social scale, in contrast to the quantitative character of phonetic variables.

Standard languages seem to have a lower tolerance towards grammatical variation. Thus, this type of variables are frequently subject to a sociolinguistic filtration which may alter the linguistic principles that explain their original function. This is the case, for instance, of the uses –considered as anomalous- of the unstressed pronouns known as *leísmo, laísmo y loísmo*.[2] Thanks to the sociolinguistic interviews of Klein-Andreu (1979, 1981, 2000) and *COSER* (see Fernández-Ordóñez 1994, 1999), we can know nowadays that what grammarians considered as deviated uses of the regular pronominal use are in fact partial manifestations of alternative pronominal paradigms in which pronouns are selected according to linguistic principles different to those applied in Standard Spanish. Some of these paradigms, like the Castilian referential paradigm, are only fully present in the speech of sociocultural groups of lower status. As the social status becomes higher, most of the

1 The conclusions obtained of this contrast among linguistic groups are methodologically suitable if we take into consideration the identity of the conversational type in which data were obatined both in *COSER* and in other corpora: the interview (type of conversation subject to the question-answer exchange) is always the framework which generates the data recorded in *COSER* and, frequently, in other oral corpora of Spanish as well.

2 As we well see below (cf. 2.2.1), *leísmo* is the use of the dative pronoun *le* instead of the accusative pronouns *lo* and *la* as direct objects. *Laísmo* is the use of the accusative pronouns *la* and *las* instead of the dative pronouns *le* and *les* as indirect objects, and *loísmo* is the use of the accusative pronouns *lo* and *los* instead of the dative pronouns *le* and *les* as indirect objects.

©Universitat de Barcelona

characteristic uses of these paradigms (*leísmo* meant for inanimate objects, *laísmo* and *loísmo*) are discarded. This sociolinguistic distribution has traditionally confused its correct interpretation, since most scholars have drawn their hypotheses on this matter exclusively on the partial data offered by the written and cultivated language (in which *leísmo* for a masculine person is accepted whereas the other *-ismos* are normally rejected). *COSER* data have allowed thus to understand grammatical variables whose linguistic rules became confused as they hardly entered into the standard language or did not enter at all.

COSER offers besides another aspect of remarkable interest: the possibility of studying linguistic changes in real time. Since *COSER*'s informants belong to the same social group as the informants of linguistic atlases, it is possible to compare the speech of several successive generations of speakers of this group. In spite of the differences between the atlas questionnaire methodology and the socolinguistic interview used for *COSER*, the contrast between the data obtained from linguistic atlases of peninsular Spanish and from the *COSER* interviews allows indeed to research any changes in real time: the time elapsed from the beginning of the 20th century to the beginning of the 21st century (chronological distance). It has thus been possible to notice the decline of some aspects of rural grammar, like the *me se, te se* sequences (Heap 2006), while proving the relative stability of others, like the use of the conditional instead of the imperfect subjunctive (Pato 2004).

2.2. Grammatical variation is better explained through COSER

It is important to point out that the *COSER* interviews have proved especially useful to record dialectal phenomena related to grammar, which is traditionally an aspect hardly represented in dialectal monographs and in questionnaires of linguistic atlases. Indeed, the development of the interview enables to research the use of any grammatical phenomenon in a real context of use: instead of the isolated, out of context and unnatural sentences typical of a questionnaire, the interview collects sentences uttered in a real speech, in which it is possible to investigate contrastive values, affective motivations and pragmatic inferences related to a specific structure. Thus, for instance, data from *COSER* enable to understand better a structure which existed in the old Spanish and is only found nowadays in some specific rural varieties, and has clearly a focal value: the use of the article followed by the

©Universitat de Barcelona

possessive adjective (*el mi hijo*) (*the my son*), which in these varieties is used alternately with the regular emphatic possessive structure in Spanish (*el hijo mío*) (*the son of mine*). The focal character of the structure explains that both are preferably applied with possessives of the first and second persons, relating to the speaker and listener, and with objects highlighting the relationship between possessor and possessed, aspects which may be difficult to record in sentences isolated from speech such as in atlas questionnaires or those sporadically quoted in dialectal monographs.[3]

2.2.1. Pronominal paradigms instead of pronominal deviations (*leísmo, laísmo* and *loísmo*)

As mentioned above, the contributions of the interview methodology have been essential in an aspect traditionally poorly understood of Spanish syntax: *leísmo, laísmo* and *loísmo*. *Leísmo* is the use of the dative pronoun *le* instead of the accusative pronoun *lo*: an extension of the dative at the expense of the accusative form. Both *laísmo* and *loísmo* are the use of the accusative pronouns *la* and *lo* instead of the dative pronoun *le:* an extension of the accusative at the expense of the dative forms. Although the relation among these uses had been long perceived, scholars did not succeed in explaining them as the product of a coherent linguistic principle and, in order to account for their genesis, they had to resort to the combination of two contradictory tendencies.[4] On the one hand, the tendency to distinguish personal direct objects (using *le* and personal *leísmo*) from non-personal (using *lo, la*, without *leísmo*). According to this hypothesis, *leísmo* was explained as a parallel development to prepositional accusative marking in Spanish, thus *leísmo* was supposed to be a way to signal animate objects morphologically.This tendency explained personal *leísmo* but did not make clear why

3 Moreover, the *Linguistic Atlas of the Iberian Peninsula* (or after its Spanish abbreviation *ALPI*), whose material until recently went missing, is the only atlas devoted to Spanish which includes a question enabling to record this use: No. 261, *Sus corderos están en nuestro prado (Their sheep are in our field*). Fortunately, *ALPI* material, of which only one volume was published (cf. Navarro Tomás 1962), can be presently consulted on the Internet (cf. Heap 2002, 2003-). Some regional atlases include the question *Mi mujer va a menudo al médico* (*My wife goes often to the doctor*) (*Navarre, Aragon and La Rioja Linguistic and Etnographic Atlas*, or *ALEANR*, map 1743, *Castile and Leon Linguistic Atlas*, or *ALCyL*, map 165), but not the answer regarding the possessive but regarding the adverbial phrase is mapped. *Cantabrian Linguistic and Etnographic Atlas* (or *ALECant)* did not surprisingly include any question regarding this use, in spite of being well-known in the region as proved by the linguistic characterization introductory notes of each enclave.

4 The most renowned exponents of this interpretation are Cuervo (1895), Fernández Ramírez (^{2}1987) and Lapesa (1968). Similar hypotheses are set out by García (1975) and Flores Cervantes (1997, 2002). A critical review of these interpretations may be found in Fernández-Ordóñez (1993 and 2001, with arguments taking into account the Romance context).

©Universitat de Barcelona

this extension of dative morphology affected basically masculine objects, and not feminine, or why it might be concurrent with non-personal *leísmo* (referred to inanimate objects) *laísmo* and *loísmo*. It thus seemed necessary to resort to another tendency, which aimed to assign pronouns exclusively according to the gender of their antecedents, regardless of their syntactic role in the sentence. In this view, *leísmo, laísmo* and *loísmo* tried to eliminate case distinctions in favour of gender distinctions, with *le* as the masculine pronoun, *la* as the feminine pronoun and *lo* as the neuter pronoun. This tendency could explain masculine *leísmo* (personal and non-personal) as well as *laísmo*, although it did not make clear why the use of *leísmo* could not be established for all kinds of masculine objects, since it was always more frequently used when the antecedent was a personal object rather than a non-personal object. It neither explained why the use of *leísmo* was not completely established in the plural, in which case it contended with *loísmo*.

In accordance with these traditional remarks, the authors of linguistic atlases designed questionnaires which aimed to record personal masculine *leísmo*, *loísmo* and *laísmo* above all, i.e., which intended to record the basic manifestations of the first and second tendencies, respectively. At the same time, other uses, like non-personal *leísmo*, or the usage of the pronoun *lo* referred to mass entities (masculine and femenine) or mass neuter, which, as proved below, is indissolubly linked with it, were not researched.[5] The analysis of the data

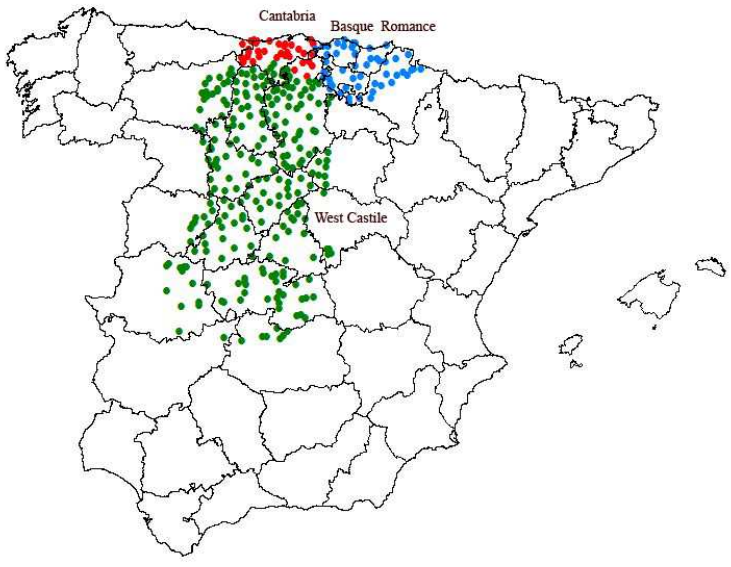
---

5 Thus, *ALPI* devotes five questions to personal *leísmo* (350 *A Miguel le cogieron preso (Michael was held prisoner*, 351 *Le llevaron a la cárcel (He was sent to prison)*, 352 *Al padre le vieron llorando (The father was seen crying)*, 353 *A los niños les socorrieron los vecinos (The children were helped by neighbours)*, 355 *Al enfermo hay que cuidarle (The sick person must be looked after)*): apart from the high number of questions devoted to record the same phenomenon, the standard character of masculine personal *leísmo* is shown by the fact that the questions of the questionnaire are expressed according to a *leísmo* solution. In contrast, those devoted to *loísmo* (356 *Al niño le pusieron un vestido (The child was dressed in a dress)*, 357 *Tráete los candiles para echarles aceite (Bring the oil lamps in order to add some oil to them*) and to *laísmo* (359 *A la madre no le dieron la limosna (the mother was not given any alms)*, 360 *Aquella desgracia le costó a ella la vida (That misfortune cost her her life)*, 361 *A las hermanas les enviaron unas cartas (Some letters were sent to the sisters)*, 362 *A la yegua le cansa el trabajo (The mare gets tired working)*), are expressed with the regular solutions of the pronominal case. No questions related to masculine non-personal *leísmo* were planned. Nevertheless, questions 312 and 313, intended to record the conjugation of the verb *vaciar* (*to empty*), might also allow to research non-personal *leísmo* (312 *¿Dónde vacían el cántaro?(Where is the jug emptied?)*, 313 *No lo vacíes en la calle (Do not empty it in the street)*). *ALEANR* devotes less entries of its questionnaire to such uses and besides, most of them are exact to some of those included in the *ALPI* questionnaire (it reproduces thus those numbered 350-351, 353, 356, 359, 362 corresponding to maps 1708-1711). There are no questions which enable to record non-personal *leísmo*, although there is one question which enables to record femenine personal *leísmo* (*A la madre la vio en la calle (The mother was seen in the street)*, map 1713). Only *ALECant* and *ALCyL* include new questions aimed at non-personal *leísmo* (with animate antecedents, *Al lobo lo vimos (We saw the wolf)*, maps 1194 and 118, respectively, and inanimate, *El libro lo olvidé en casa (I forgot the book at home)*, *ALECant* 1195, *El paquete lo olvidé (I forgot the parcel)*, *ALCyL* 116). These two regional atlases also

©Universitat de Barcelona

from rural speech recordings has enabled to establish the exact geographic delimitation of the areas where each one of these uses is found, while it has also proved that the apparent lack of coherence in their frequency is actually due to the existence of several pronominal paradigms, alternative to the regular paradigm of Spanish. The data from these paradigms were mixed in earlier studies altering thus the interpretations (Fernández-Ordóñez 1994, 1999, 2001). Apart from transition solutions, there are three basic paradigms: one is the paradigm used in the Romance spoken in contact with Basque (Table I), another one is the Cantabrian paradigm (Table II), and the third one is located in West Castile, sometimes called referential system (Table III). The Castilian paradigm resulted from the evolution of the Cantabrian by eliminating the case category. *Laísmo* and *loísmo* are only found in the Castilian paradigm, whereas all three paradigms show personal and masculine *leísmo.* This fact clarifies the reasons why *leísmo* proved to be the most frequent phenomenon in traditional remarks (besides being the only one not to be rejected in cultivated and written language).

The geographic distribution of these paradigms is shown in map II.



Map II. Basque Romance, Cantabrian, and West Castilian Paradigms.

reproduce questions 350, 352-353, 356, 359 and 362 of *ALPI* (*ALECant*, 1243, 1245-1247, 1192, 1197; *ALCyL*, 111-114, 117, 120) and 1713 of *ALEANR*. None of the atlases enables to notice the absence of *leísmo* when the antecedent is a masculine mass object (like *pan (bread), vino (wine), trigo (wheat)*, etc.) or the use of *lo* to refer to femenine mass objects (*agua (water), miel (honey), manteca (butter)*, etc), not even *ALECant*, in spite of the fact that Cantabria is a region where the existence of the mass neuter was well-described.

©Universitat de Barcelona

Basque paradigm (marked in blue in map II) shows the materialization of the hypothesis of *leísmo* as an extension of the dative to personal objects, both masculine and femenine. On the other hand, Cantabrian (marked in red) and Castilian (marked in green) paradigms are distinguished by basing pronominal selection in the semantinc categorization of the antecedent as uncountable or countable, a linguistic category which had not been previously taken into consideration and which accounts for the fact that *leísmo* was universal with masculine personal antecedents (always countable and referred by *le*) but was not generalized with non-personal ones, since they might be countable (referred by *le*) or uncountable (by *lo*). Castilian paradigm, in turn, is distinguished from Cantabrian paradigm by the elimination of the case category, generalising thus *la(s)* and *lo* as dative pronouns. To make things more complicated, in masculine plural, Castilian system shows at least two different solutions according to the preferred pronoun: *les*, used in the North (North West of Burgos, Palencia and Valladolid); *los*, used in the South (East of Salamanca and Cáceres, Ávila, West of Toledo and Madrid). The territories where the Castilian system is used are thus those of the Centre and West of Castile, from the South of the Cantabrian mountain range to La Mancha.

For a better clarification, I have marked in bold italics the partial aspects in which these three paradigms differ from the paradigm of Standard Spanish-(Table IV).

	ANIMATE		INANIMATE			
	Masculine / Feminine		Masculine		Feminine	
ACCUSATIVE	Singular	Plural	Singular	Plural	Singular	Plural
	le	***les***	Ø / lo	Ø / los	Ø / la	Ø / las
DATIVE	***le***	***les***	le	les	le	les

Table I. Basque Romance Paradigm.

©Universitat de Barcelona

	COUNT				MASS	
	Singular		Plural			
ACCUSATIVE	Masculine	Feminine	Masculine	Feminine	Masculine	Feminine
	le	la	los	las	lo	***lo***
DATIVE	***le***	le	les	les	le	le

Table II. Cantabrian paradigm.

	COUNT				MASS	
	Singular		Plural			
ACCUSATIVE	Masculine	Feminine	Masculine	Feminine	Masculine	Feminine
	le	la	los (South) / ***les*** (North)	las	lo	***lo***
DATIVE	le	***la***	***los*** (South) / les (North)	***las***	***lo***	***lo***

Table III. Western Castilian paradigm.

	Singular		Plural	
ACCUSATIVE	Masculine	Feminine	Masculine	Feminine
	lo	la	los	las
DATIVE	le	le	les	les

Table IV. Standard Spanish paradigm.

©Universitat de Barcelona

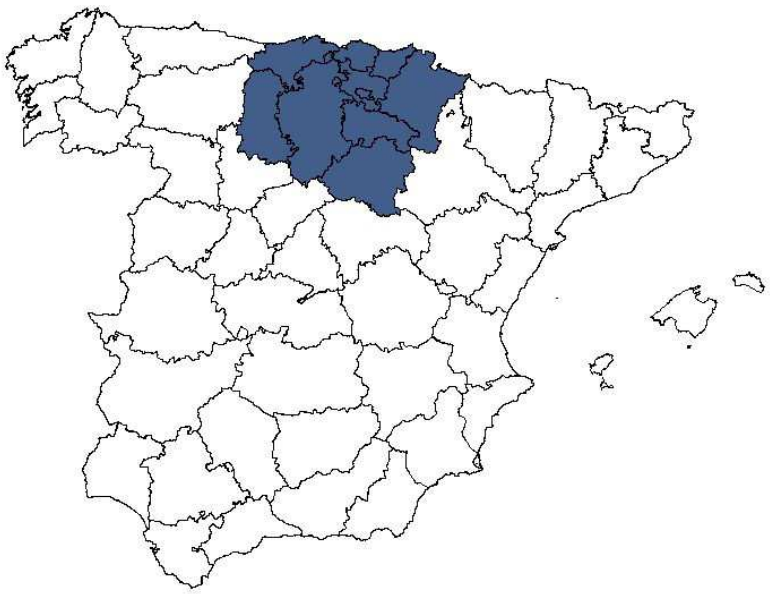
The partial consideration of the uses not coincident with the general paradigm –without integrating them duly in the pronominal paradigms and linguistic principles that ruled them- accounts for the traditional lack of understanding in this regard by grammarians and dialectologists of Spanish language, who moreover only paid attention to the most frequent uses which deviated from the general paradigm without considering any others which were a minority from a global point of view. That happened to uses like *leísmo* referred to femenine personal objects (*A María le vi ayer (I saw Mary yesterday)*) or null pronouns (*Los libros te Ø he dado (I have given you the books)*), exclusive to Basque Romance, or like *lo* referred to femenine mass objects (*La lana lo venden (wool is sold)*), typical of the Cantabrian-Castilian area.

2.2.2. Modal changes, minority variables and data quantifying

COSER therefore enables more correct interpretations of the linguistic principles in force in oral varieties, as we have just shown. This advantage is undoubtedly linked with the possibility of quantifying data: given a specific linguistic variable, the interview enables to quantify the variants in a specific enclave as well as distinguishing contexts of occurrence, whereas in the atlases this quantifying is not usually possible since one single answer is normally given for each enclave and because very few questions related to one specific variable are included. As a result, minority variants of one variable seldom appear in atlases.

This conclusion is drawn for instance by the study of a grammatical use found in the central and Northern area of the Iberian Pensinsula, i.e., the use of the simple conditional (-*ría*) instead of the imperfect subjunctive (-*ra* /-*se*), a use extended to all type of syntactic contexts accepting the imperfect subjunctive in Spanish (Pato 2004) (see map III to locate the area within the Iberian Peninsula).

©Universitat de Barcelona



Map III. Area of displacement of subjunctive within the Iberian Peninsula.

This use had been recorded in atlases, although quite insufficiently, as they omitted the fact that the imperfect subjunctive is not only replaced by the conditional *–ría* (majority variant), but also by the imperfect indicative *–ba* (minority variant). Examples (1) and (2), from *COSER*, show both variants of this use in the same informant from Santervás de la Vega (Palencia):

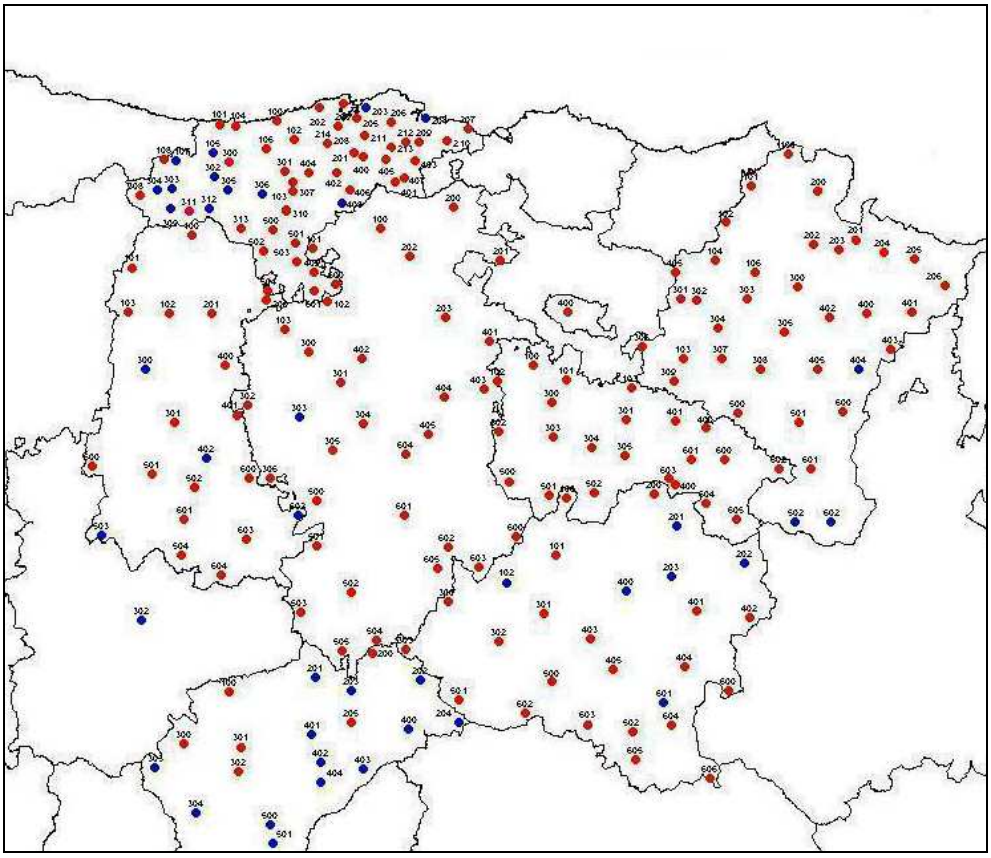
(1) Las costillas y todas esas cosas se metían en ollas para que se *conservarían.*
(Cutlets and all those things were put in pots so that they *would be preserved*).

(2) Se las colgaba en la cocina o en una habitación, o como fuera... que las diera un poco el sol, para que *estaban* más buenas [las morcillas].
(They were hung in the kitchen or in a room or anywhere... provided that they got some sun, so that they *were* more tasty [sausages]).

By comparing the maps resulting from regional atlases (*ALCyL*, *ALEANR*, *ALECant*) with the map made out from the *COSER* material, it is possible to confirm, on the one hand,

©Universitat de Barcelona

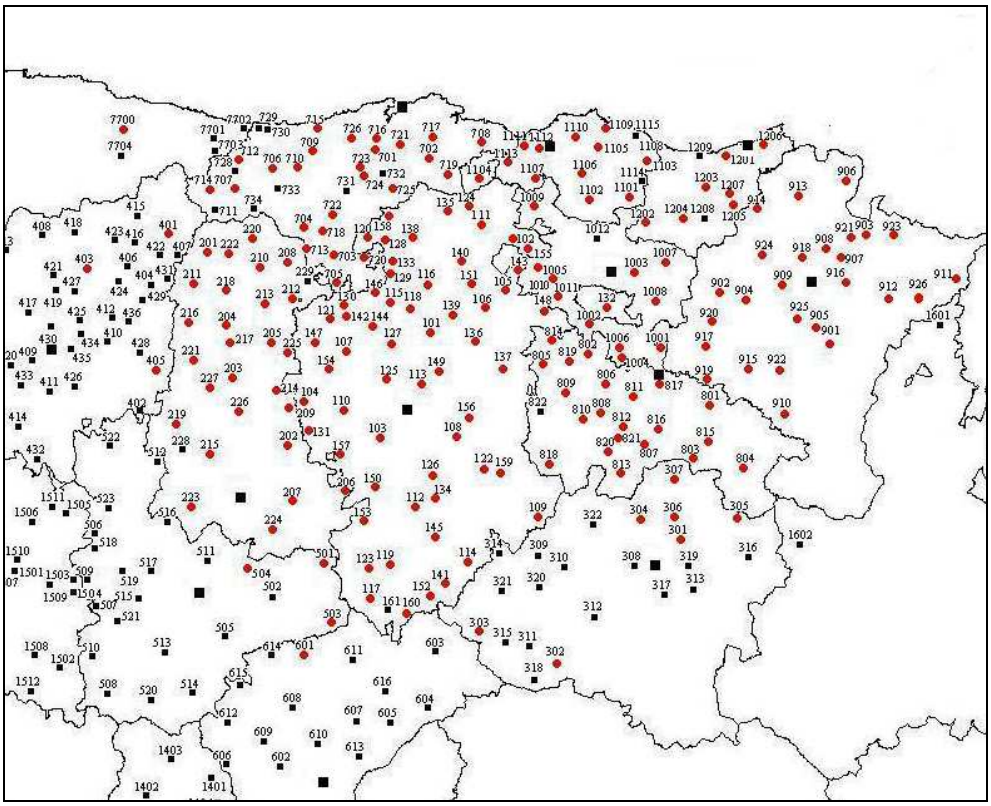
the general coincidence in the linguistic area where this phenomenon takes place (maps IV and V):[6]



Map IV. Use of *-ría* (marked in red) instead of *-ra* / *-se* (marked in blue) according to regional atlases (Pato 2004).

6 Nevertheless, there is a difference between the data provided by *ALCyL* and *COSER* as for the area affected by the phenomenon: the atlas suggests its existence in the centre and South of the province of Soria while it is not recorded in León and Valladolid. In this latter case, the absence is explained by the scarce intensity. On the other hand, the discrepancy of data is not explained in Soria, considering that *COSER* and *ALCyL* led contemporary surveys in this province.

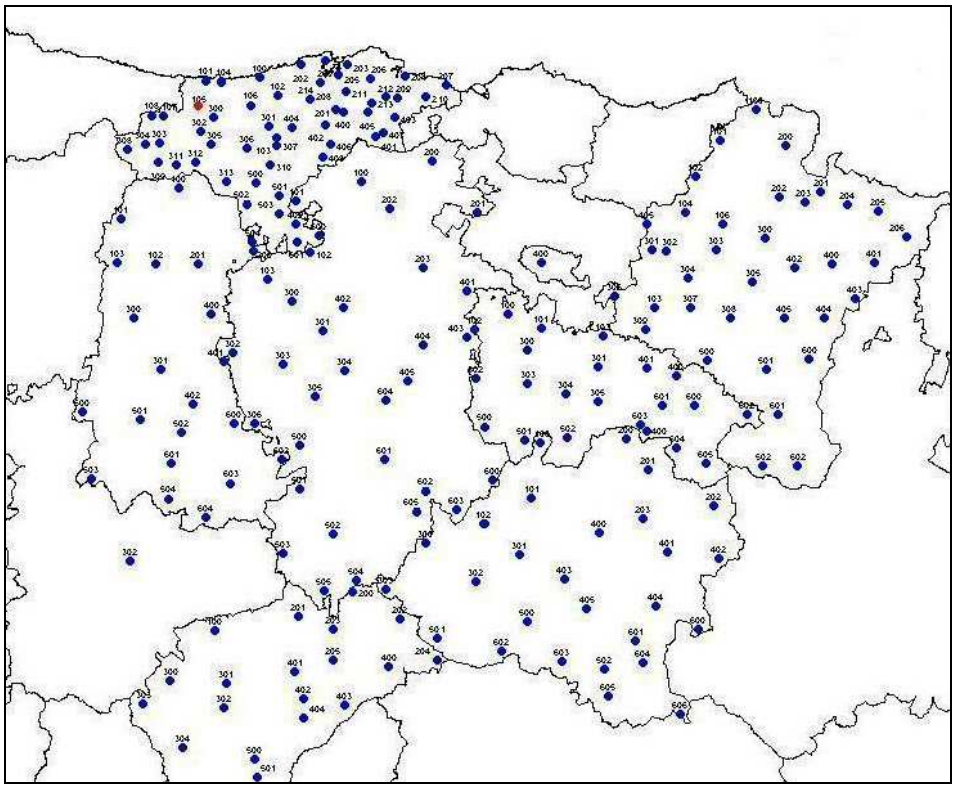
©Universitat de Barcelona



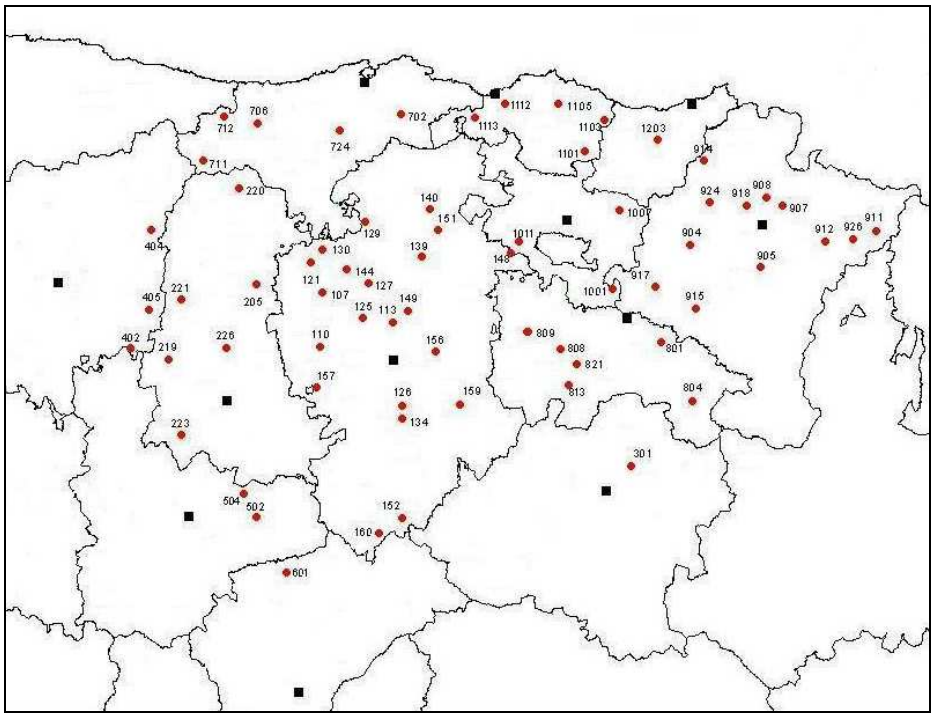
Map V. Use of *-ría* (marked in red) instead of *-ra* / *-se*, according to *COSER* (Pato 2004).

On the other hand, it is also evident that atlases are not able to reflect the minority variant *-ba*, whereas it is regularly recorded by *COSER* in the whole area (maps VI and VII).

©Universitat de Barcelona



Map VI. Use of *-ba* (in red) instead of *-ra* / *-se* (in blue), according to regional atlases (Pato 2004).



Map VII. Use of *-ba* (in red) instead of *-ra* / *-se*, according to *COSER* (Pato 2004).

©Universitat de Barcelona

Although both variants exist, as we have seen, their proportion of use is not equivalent, which accounts for the fact that the minority variant was hardly recorded by atlases: when the imperfect subjunctive is displaced by these forms of indicative, *-ría* was prevalent in 96% of the cases, whereas *–ba* appeared just in 4% of the cases.

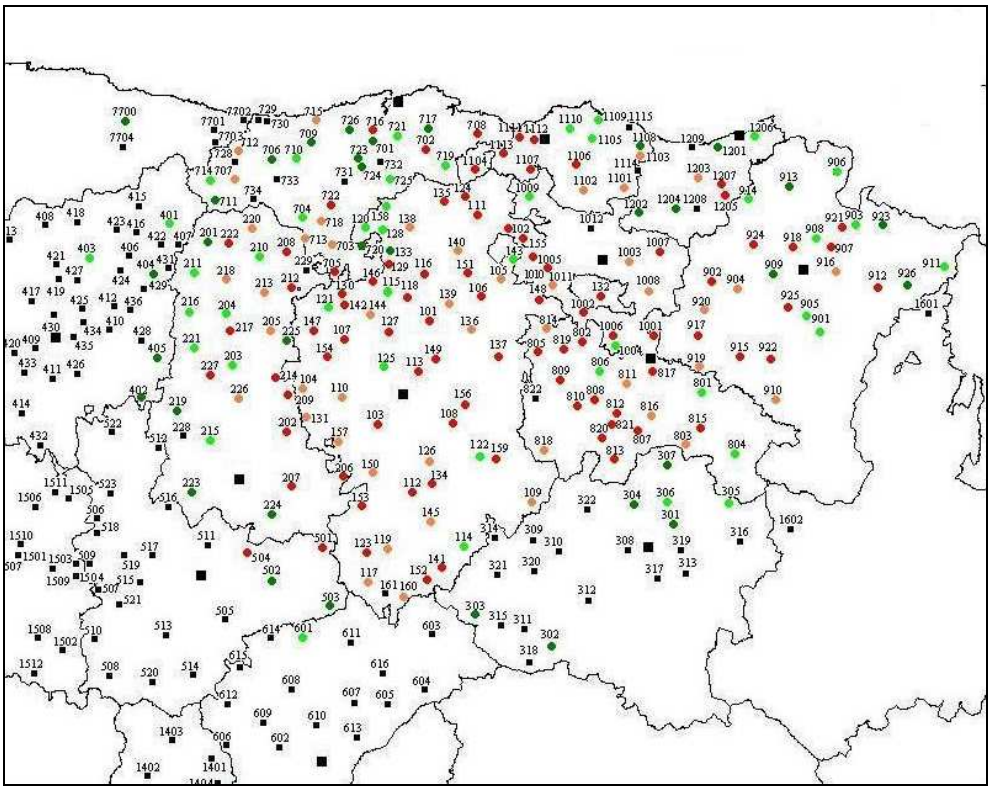
Data quantifying is not impossible from data obtained by atlases, but it is statistically more reliable if data come from a corpus like *COSER*. First of all, because the phenomenon is sometimes recorded in contexts which were unexpected when atlas questionnaires were designed. This was indeed the case, as we have seen above, for *leísmo, laísmo* and *loísmo*. This problem also happens in the recording of the use of *-ría* / *-ba* instead of *-ra* / *-se*, since atlases had planned to record this use preferably in the protasis of conditional sentences and in desiderative sentences using *ojalá (I wish, I hope)*,[7] while in fact the phenomenon appears in noun, adjective, final, concessive, causal clauses, etc: i.e., in any subordinate clause where the imperfect subjunctive is likely to be found in Spanish (as already noticed by Ridruejo (1975), Silva Corvalán (1985) or Martínez Martín (1983) in studies limited to the areas of La Rioja and Burgos). In the case of both pronominal and verbal uses, the atlas questionnaire records as partial deviations of the general use what is actually an alternative use controlled by different linguistic principles and which takes place in a significantly wider range of contexts.

Secondly, the number of records regarding the phenomenon obtained in any interview is always necessarily higher than that provided by an atlas questionnaire, even if all syntactic contexts likely to show this phenomenon had hypothetically been included. It is this significant number of records what enables to detect the presence of minority variants, which are in fact concealed in atlases. Therefore, in statistical terms, data quantifying from a corpus like *COSER* enables to draw conclusions far closer to reality as regards linguistic uses. For instance, this quantifying enables thus to clarify the above maps (maps IV, V, VI, and VII),

[7] Four relevant questions were included in *ALPI* (386 *Si tuviera dinero lo compraría (If I had money, I would buy it)*, 387 *Si estudiase aprendería (If I studied, I would learn)*, 388 *Si pudiera la mataría (I would kill her if I could)*, 390 *Ojalá lloviese (If only it would rain)*), of which the first and last ones were reproduced in *ALEANR* (maps 1704, 1706), in *ALECant* (maps 1216, 1220) and in *ALCyL* (148, 152). *ALEANR* enriched the syntactic contexts by adding an entry which included a noun clause (1705 *Le dijo que trajera un pan (He told him to bring some bread)*), which *ALECant* and *ALCyL* also inherited (maps 1218 and 150, respectively). *ALECant* added in turn a concessive clause to the list (1217 *Aunque pudiera no lo haría (I would not do it, even if I could)*), reproduced in *ALCyL* (map 149). Finally, only the *ALCyL* questionnaire includes a final clause (151 *Esto te lo dije para que fueras bueno (I told you this so that you were a good boy)*).

©Universitat de Barcelona

by showing which is the focal area of the use of *-ría* / *-ba* instead of *-ra* / *-se*, and which are the transitional areas: see map VIII, where the focal area is clearly delimited in the North and East of Burgos, and in the bordering regions of Cantabria, Biscay, Álava and La Rioja Alta.



Map VIII: Use of *-ría* /*-ba* instead of *-ra* / *-se*, according to *COSER*, including quantifying of the phenomenon (Pato 2004) 75-100% • / 50-75% • / 25-50% •

Another general characteristic of this modal displacement which the *COSER* data have enabled to identify is that it is characteristic of simple tenses, reaching in the area an average frequency of 61.9%, whereas it is hardly found in compound tenses (21.6%).

The quantity of data also makes it possible to apply statistical tests like logistic regression, enabling to assess the simultaneous influence of several variables on the phenomenon manifestation. In the case of the use of *-ría* / *-ba* instead of *-ra* /*-se* in the Castilian varieties, it has been proved that the most widespread opinion in this regard according to which, the protasis of the conditional sentences was considered as the origin of

©Universitat de Barcelona

this phenomenon, was not actually correct. Instead, the prevalence of *-ría* and *-ba* over the subjunctive forms *-ra* / *-se* was first found in noun clauses, extending next to the adjective and dependent adverbial clauses and finally, to the conditional and final clauses, as well as the rest of syntactic contexts (Pato 2003, 2004).

Frequence of subjunctive displacement according to the type of clauses

Noun clauses (72.1%) > Adjective and adverbial (modal, locative and temporal) clauses (61.7%) > Conditional / Final (57.5%)

Maybe the order *noun clauses* > *adjective clauses* > *conditional* > *rest* was overlooked because in noun and adjective clauses it is possible to find an alternation of modes in Spanish without always seeing clear differences in the interpretation of the modal contents. In noun clauses, the modal contents may be both expressed by the imperfect subjunctive and the conditional. The selection of indicative vs. subjunctive is normally associated with the +/- assertive value of the statement in the embedded clause (see 3b vs. 3a,c):

(3a) María sabía que Jaime vendría (*viniera) a visitarla
(Maria knew that Jaime would come (*came) to visit her)
[+assertion].

(3b) María esperaba / no creía que Jaime vendría / viniera a visitarla (Maria hoped / did not think that Jaime would come / came to visit her)
[+/- assertion].

(3c) María deseaba que Jaime viniera (*vendría) a visitarla
(Maria wished that Jaime came (*would come) to visit her)
[- assertion].

In (3a) it is assumed that Jaime will come, so indicative is required. In turn, in (3c) it is uncertain whether Jaime will come, so subjunctive is compulsory. But in (3b) it is also unsure whether Jaime will come or not, and both indicative and subjunctive are possible.

©Universitat de Barcelona

Something similar occurs with adjective and dependent adverbial clauses. As it is well-known, an alternation of moods in relative and adverbial clauses is normally explained in Spanish by the more or less specific character of the antecedent: if it is specific, existing, the indicative is used and if it is unspecific, i.e. if it is not stated that it exists, the subjunctive is then required. As shown in the following sentences, the conditional and the imperfect subjunctive may also co-appear in contexts of uncertain interpretation (see 4b vs. 4a,c):

(4a) El hombre, que sabría (*supiera) aquel misterio hacía tiempo, había desaparecido
(The man, who would know (*knew) that mistery long time ago, had disappeared)
[+ specific].

(4b) El hombre que sabría / supiera aquel misterio había desaparecido (The man who would know / knew that mistery had disappeared)
[+/- specific].

(4c) No hubo nadie que supiera (*sabría) aquel misterio
(There was nobody who knew (*would know) that mistery)
[-specific].

In the first example (4a), the adjective clause is explicative and thus requires the specific character of the antecedent: therefore, it is not possible to use the subjunctive. In the third example (4c), the antecedent is unspecified: *nobody* implies the non-existence of the referent: the indicative is in this case ungrammatical. The second clause (4b), however, is open to a +/- specific interpretation of the antecedent and thus enables to use both moods alternately.

In accordance with the above, in the Northern Castilian area, the prevalence of the indicative over the subjunctive is more frequent as the antecedent is more specific (definite and explicit), as it may be assumed from the scale below:

©Universitat de Barcelona

Frequence of displacement of the subjunctive according to the type of antecedent in adjective clauses

Explicit definite antecedent (*the N who / which*, 70.2%) > Explicit indefinite antecedent (*a N who / which*, 60.3%) > Non-explicit definite antecedent (*the one who / which*, 58.8%) > Non-explicit indefinite antecedent (*one who / which*, 56.4%) > *nobody who / which* (0%).

Whereas these alternative uses of mood noun and adjective clauses seem general in Spanish, the presence of the conditional instead of the subjunctive in the protasis of conditional clauses (*Si tuviera / tendría dinero, lo compraría (If I had / would have money, I would buy it)*) has always been considered a restricted use to some dialects and, in the Iberian Peninsula, it is and has been a stereotype of Basques' speech. This impression might have led to set the origin of the phenomenon in the conditional clauses. However, in Northern Castilian, conditional protases do not actually constitute the original context of the phenomen, but the part which seems to have traditionally proved more "visible" to grammarians and dialectologists.

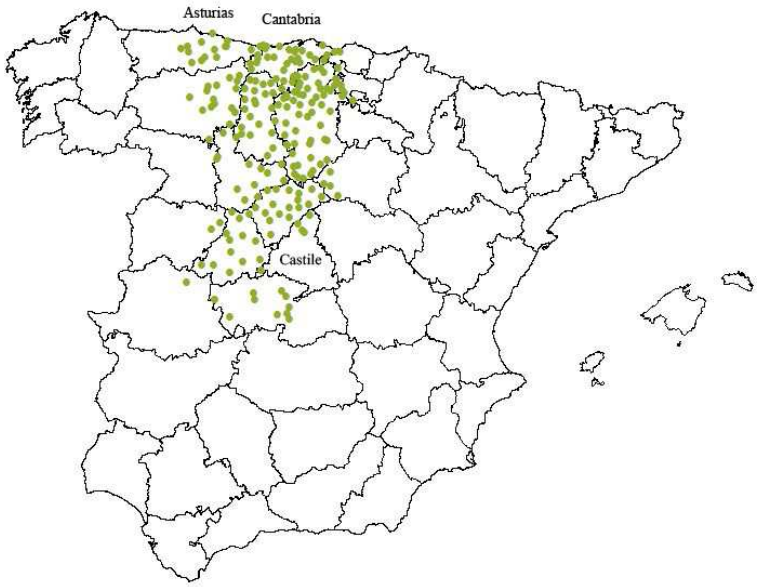
The loss of the subjunctive according to the scale *noun clauses* > *adjective clauses* > *conditional clauses* > *rest* may be probably explanatory for other varieties of Spanish, like the Spanish in America, or even to other Romance languages. In typological terms, it is a predictable change, since the indicative is less marked than subjunctive and the subjunctive may be independently lost in different varieties.

2.2.3. Mass neuter agreement, non-existent variables and typological implications

In the cases just reviewed, *COSER* recordings made it possible to study and understand better dialect phenomena which were hitherto partially known. But *COSER* interest is enhanced by the fact that it has recorded dialect phenomena completely ignored by grammarians and dialectologists up to now. The best example is mass neuter agreement. This agreement was traditionally known in Central and Eastern Asturias and Cantabria, but went fully unnoticed in Castile. Thanks to *COSER* recordings, the geographical area with mass

©Universitat de Barcelona

neuter agreement has been considerably enlarged to the South (Fernández-Ordóñez 2007, 2006-2007), as map IX illustrates.



Map IX. Mass neuter agreement area.

What is mass neuter agreement? In these dialects neuter pronoun agreement is to be seen not only with non-lexical referents (which is the norm in Spanish), but with lexical antecedents when the masculine or feminine noun (singular or plural) has a mass interpretation. This neuter agreement is not limited to pronouns but it extends to adjectives: post-nominal attributive adjectives, predicative adjectives and depictive adjectives. Nevertheless, the mass neuter agreement is never exhibited by pre-nominal elements, such as articles or adjectives, or rarely by the noun itself, as Table V illustrates.

Determiners (Article and Demonstrative)	**Noun**	**Adjectives**	**Personal and Demonstrative Pronouns**
MASC/FEM	MASC/FEM	MASC/FEM/NEUT	MASC/FEM/NEUT

Table V. Gender distinction according to the word class.

©Universitat de Barcelona

(5) a. El buen vinu blanc-o se toma frí-o. Pruéba-lo
the.M good.M wine.M white-N is drunk cold-N taste-it.N
"Good white wine is drunk cold. Taste it"

b. La buen-a leche fresc-o se toma templad-o. Pruéba-lo
the.F good-F milk.F fresh-N is drunk warmed-N taste-it.N
"Good fresh milk is drunk warmed. Taste it"

c. La-s medicina-s es car-o. Lo compramos en la farmacia
the.F-Pl medicine.F-PL is expensive-N it.N we buy in the pharmacy
"Medicines are expensive. We buy them in the pharmacy"

In (5) we see how mass neuter agreement occurs with nouns that receive a mass interpretation, both masculine and feminine and singular and plural. We also see how adjectives can exhibit neuter agreement both in predicative and attributive positions. But this full expression of mass neuter agreement is limited to Asturias. In Cantabria and Castile it is virtually non-existent with atrributive adjectives, and restricted to predicative adjectives and pronouns. Moreover, the statistical and geographic distribution of the agreement in Asturias, Cantabria and Castile proves that it arose in the pronouns and that it gradually extended by steps: firstly to depictive adjectives (secondary predicates), secondly to predicative adjectives, and finally to attributive adjectives, as table VI illustrates. In the table the + sign refers to 30% or more of mass neuter agreement and the – sign to less than 30%:

Mass neuter agreement with feminine nouns	**Determiner**	**Noun**	**Post-nominal Attributive Adjective**	**Predicative Adjective (*ser*)**	**Predicative Adjective (*estar*)**	**Adjective as Secondary Predicate**	**Personal Object Pronoun**
Asturias	–	–	+ 29%	+ 40,5%	+ 58,7%	+ 65%	+ 86,4%
Cantabria	–	–	– 10%	+ 35%	+ 55%	+ 59,3%	+ 81,5%
Castile	–	–	–	– 18,5%	+ 51,2%	+ 53,3%	+ 76,5%

Table VI. Gender distinction according to the syntactic position.

©Universitat de Barcelona

So the mass neuter agreement begins with the pronouns, usually placed in the sentences following the one which contains the antecedent. Thereafter the mass neuter agreement progresses backwards reaching the previous sentence which contains the nominal antecedent. Once in this sentence, the extension continues from the predicate (secondary or primary) to the subject. Inside the nominal phrase, the mass neuter agreement cannot reach across to all its components: just those located after the noun, the post-nominal attributive adjectives. The determiners and pre-nominal attributive adjectives remain masculine or feminine, just as the nouns. There is also an interesting difference between the percentage of tokens of predicative adjectives with the copula *ser* (individual-level predicates) and those with the copula *estar* (stage-level predicates). Mass neuter agreements are clearly more frequent when the adjective denotes a stage-level predicate (as a predicative adjective or a secondary predicate).

These facts are interesting not only because they imply a better knowledge of dialect grammar, but also because this pattern of diffusion coincides with the semantic agreement hierarchy stated by Corbett (1991, 2006):

Agreement Hierarchy

attributive > predicate > relative pronoun > personal pronoun

> "For any controller that permits alternative agreements, as we move rightwards along the Agreement Hierarchy, the likelihood of agreement with greater semantic justification will increase monotonically (that is, with no intervening decrease)" (2006:207).

Mass agreement patterns in Ibero-Romance dialects

attributive > predicative > secondary predicate > personal and demonstrative pronoun

The research of mass neuter agreement has thus revealed to have typological implications. Given the closer analysis of linguistic data that dialects allow, given the small

and progressive changes between neighbouring varieties, it is possible to suggest a refinement of the Agreement Hierarchy. According to the behaviour of Ibero-Romance dialects with mass neuter agreement, the position "predicate" in the hierarchy could be divided in primary and secondary predicates, being the latter the most probable to show semantic agreement. But as seen in Table VI, secondary predicates and primary predicates with *estar* show almost the same frequency of mass neuter agreement vs the lower incidence in primary predicates with *ser*. Both secondary predicates and *estar* predicates have in common to be stage-level predicates, whereas *ser* predicates are usually individual-level predicates. Thus the Ibero-Romance data support the hyphotesis that the Agreement Hierarchy could be rather revised to account for this difference:

A proposal of refinement of the Agreement Hierarchy

attributive > individual level predicate > stage level predicate > relative > pronoun

Be as it may, dialect grammar has revealed as an important source for a better understanding of many cross-linguistic principles and opens up new ways to test their validity.

2.2.4. Nevertheless atlases are useful

Corpora like COSER are thus instrumental for a better knowledge of dialect grammar. Nevertheless, although atlases and dialectal monographs show some shortcomings as for the study of grammar, we should not discard their usefulness, even as regards morphosyntax. At the time when some of these works were conceived and made out, neither syntax nor sociolinguistics had reached the theoretical development they have experienced in the last fifty years. Although the methodology of the questionnaire used to reflect the speech in *ALPI* and in subsequent regional atlases is very different to the methodology of the sociolinguistic interview of *COSER*, we must admit that they are both the product of the theoretical state of dialectology at that time. The development of sociolinguistics has shown multiple limitations of atlas methodology; however, it is important to bear also in mind that, since there are no

speech recordings of past times which are equivalent to current recordings (and there is no human means to obtain them), atlas data remain thus a precious testimony –however imperfect it may be- for the study of rural speech (as well as the grammar, as proved by works like Heap's, 2000). Moreover, it has never been emphasized enough that the comparability of data provided by a questionnaire is rarely obtained with the methodology of the interview, in which researchers may try to obtain some specific data, but without ever being certain if their aim will be successfully achieved. For that reason, projects like the *Dynamic Syntactic Atlas of the Dutch Dialects* (DynaSAND) (Barbiers 2006) supplement oral interviews with questionnaires. On the other hand, linguistic atlases offer a type of information which is not provided by corpora like *COSER*. Sociolinguistic interviews have proved especially productive to record phenomena of grammatical character but not as far as lexis is concerned. Since it is a semi-structured conversation, the words of dialectal character recorded in COSER are not always repeated and no conclusions are drawn comparable to those of an atlas as regards vocabulary. Therefore, *COSER* constitutes a complement of the material collected in linguistic atlases as well as in other type of dialectal sources, a complement which opens up enriching prospects for the study of dialectal grammar. Beyond the interest in phonetics and lexis, contemporary dialectologists know that dialectal grammar is a source of precious information (until recently insufficiently valued) not only for the characterization of a particular linguistic domain but also for the typological study of languages (see Kortmann 1999, 2004a, 2004b).

References

ALCyL: ALVAR, Manuel (1999) *Atlas lingüistico de Castilla y León*, 3 vol., Valladolid: Junta de Castilla y León.

ALEANR: ALVAR, Manuel, et al. (1979-83) *Atlas lingüístico y etnográfico de Aragón, Navarra y La Rioja*, with the collaboration of Antonio LLORENTE, Tomás BUESA and Elena ALVAR, 12 vol., Madrid : Institución Fernando el Católico / CSIC.

ALECant: ALVAR, Manuel, et al. (1995) *Atlas lingüístico y etnográfico de Cantabria*, with the collaboration of José Antonio MAYORAL, 2 vol., Madrid: Fundación Marcelo Botín.

ALPI: NAVARRO TOMÁS, Tomás (dir.), et al. (1962) *Atlas lingüístico de la Península Ibérica*, vol. 1, *Fonética*, with the collaboration of Francesc de Borja MOLL, Aurelio M. ESPINOSA [junior], Luís F. Lindley CINTRA, Armando Nobre de GUSMÃO, Aníbal OTERO, Lorenzo RODRÍGUEZ-CASTELLANO and Manuel SANCHIS GUARNER, Madrid: CSIC.

BARBIERS, S., et al (2006) *Dynamic Syntactic Atlas of the Dutch dialects* (DynaSAND). Amsterdam, Meertens Institute. URL: http://www.meertens.knaw.nl/sand/.

CHAMBERS, Jack K. (1995) *Sociolinguistic theory*, Oxford: Blackwell.

CORBETT, Greville G. (1991) *Gender*, Cambridge: Cambridge University Press

CORBETT, Greville G. (2006) *Agreement*, Cambridge: Cambridge University Press.

COSER: FERNÁNDEZ-ORDÓÑEZ, Inés (dir.) (2005-09) *Corpus oral y sonoro del español rural*, Madrid: Universidad Autónoma de Madrid. URL: http://www.uam.es/coser.

CUERVO, Rufino José (1895) "Los casos enclíticos y proclíticos del pronombre de tercera persona en castellano", *Romania* 24, 95-113, 219-263.

FERNÁNDEZ RAMÍREZ, Salvador (^{2}1987) *Gramática española. 3.2. El pronombre*, Madrid: Arco/Libros.

FERNÁNDEZ-ORDÓÑEZ, Inés (1993) "Leísmo, laísmo y loísmo: estado de la cuestión", in: Olga FERNÁNDEZ SORIANO (ed.), *Los pronombres átonos*, Madrid: Taurus, 63-96.

FERNÁNDEZ-ORDÓÑEZ, Inés (1994): "Isoglosas internas del castellano. El sistema referencial del pronombre átono de tercera persona", *Revista de Filología Española*, LXXIV, 71-25.

FERNÁNDEZ-ORDÓÑEZ, Inés (1999) "Leísmo, laísmo y loísmo", in Ignacio BOSQUE, Violeta DEMONTE (eds.), *Nueva gramática descriptiva de la lengua española* (3 vol.), Madrid: Espasa-Calpe, vol. 1, 1317-1397.

FERNÁNDEZ-ORDÓÑEZ, Inés (2001) "Hacia una dialectología histórica. Reflexiones sobre la historia del leísmo, el laísmo y el loísmo", *Boletín de la Real Academia Española* LXXXI, 389-464.

FERNÁNDEZ-ORDÓÑEZ, Inés (2007) "El neutro de materia en Asturias y Cantabria. Análisis gramatical y nuevos datos", in Inmaculada DELGADO COBOS y Alicia PUIGVERT OCAL (eds.), *Ex admiratione et amicitia. Homenaje a Ramón Santiago*, Madrid: Ediciones del Orto, 395-434.

FERNÁNDEZ-ORDÓÑEZ, Inés (2006-07) "Del Cantábrico a Toledo. El "neutro de materia" hispánico en un contexto románico y tipológico", *Revista de Historia de la Lengua Española*, 1, 67-118; 2, 29-81.

FLORES CERVANTES, Marcela (1997) "Individuación de la entidad en los orígenes de *leísmo, laísmo* y *loísmo*", in Concepción COMPANY (ed), *Cambios diacrónicos en el español*. México: Universidad Nacional Autónoma de México, 33-63.

FLORES CERVANTES, Marcela (2002) *Leismo, laísmo y loísmo: sus orígenes y evolución*, México: Universidad Nacional Autonoma de México.

GARCÍA, Erica C. (1975) *The role of theory in linguistic analysis. The Spanish pronoun system*, Amsterdam: North-Holland Publishing Company.

HEAP, David (2000) *La variation grammaticale en géolinguistique: les pronoms sujet en roman central*, München: Lincom Europa.

HEAP, David (2002) "Segunda noticia histórica del *ALPI* (on the 40th anniversary of the publication of its first volume)", *Revista de Filología Española*, LXXXII, 5-19.

HEAP, David (2003-05) *Atlas lingüístico de la Península Ibérica. ALPI searchable database*, London, Ontario: University of Western Ontario. URL: http://www.alpi.ca.

HEAP, David (2006) "Secuencias «invertidas» de clíticos: un cambio (¿?) en tiempo real", in José Jesús de BUSTOS and José Luis GIRÓN ALCONCHEL (eds.), *Actas del VI Congreso Internacional de Historia de la Lengua Española* (Madrid, September 29 - October 3, 2003), Madrid: Arco/Libros, I, 785-98.

KLEIN-ANDREU, Flora (1979) "Factores sociales en algunas diferencias lingüísticas en Castilla la Vieja", *Papers: Revista de Sociología*, 11, 46-67.

KLEIN-ANDREU, Flora (1981) "Distintos sistemas de empleo de 'le, la, lo': perspectiva sincrónica, diacrónica y sociolingüística", *Thesaurus*, XXXVI, 284-304. (Reprinted in: Olga FERNÁNDEZ SORIANO (ed.), *Los pronombres átonos*, Madrid: Taurus, 1993, 337-353.)

KLEIN-ANDREU, Flora (2000) *Variación actual y evolución histórica: los clíticos le/s, la/s, lo/s*, München: Lincom Europa.

KORTMANN, Bernd (1999) "Typology and dialectology", in: B. CARON (ed.), *Proceedings of the 16th international congress of linguists*, CD-ROM, Amsterdam: Elsevier Science.

KORTMANN, Bernd (2004a) "Why dialect grammar matters", *The European English Messenger*, XIII, 24-29.

KORTMANN, Bernd (ed.) (2004b) *Dialectology meets typology. Dialect Grammar from a Cross-Linguistic Perspective*, Berlin /New York: Mouton de Gruyter.

LAPESA, Rafael (1968) "Sobre los orígenes y evolución del leísmo, laísmo y loísmo", in Kurt BALDINGER (ed.), *Festschrift Walther von Wartburg*, Tübingen: Max Niemeyer, 523-551. (Reprinted in: Rafael LAPESA, *Estudios de morfonsintaxis histórica del español*, Madrid: Gredos, 2000, 279-310).

MARTÍNEZ MARTÍN, Manuel (1983) "La sustitución de *cantara* (*cantase*) por *cantaría* en el habla de la ciudad de Burgos", *Lingüística Española Actual*, V/2, 179-204.

PATO, Enrique (2003) “Contextos neutralizadores de la oposición modal y relaciones de alomorfismo desde el español medieval: Las formas *cantase*, *cantara* y *cantaría*”, *Moenia*, 9, 223-252.

PATO, Enrique (2004) *La sustitución de* cantara/ cantase *por* cantaría/ cantaba *(en el castellano septentrional peninsular)*, Madrid: Universidad Autónoma de Madrid. URL: http://joule.qfa.uam.es/coser/publicaciones/enrique/2_es.pdf.

RIDRUEJO, Emilio (1975) “*Cantaría* por *cantara* en la Rioja”, *Berceo*, 89, 123-134.

SILVA CORVALÁN, Carmen (1985) “Modality and semantic change”, in: Jacek FISIAK (ed.), *Historical semantics. Historical word-formation*, Berlin / New York: Mouton de Gruyter, 547-572.