

NEW METHODS FOR THE STUDY OF GRAMMATICAL VARIATION AND THE *AUDIBLE CORPUS OF SPOKEN RURAL SPANISH*

Inés Fernández-Ordóñez
Universidad Autónoma de Madrid

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 issues which were partially known or fully ignored have been researched, the geographical distribution of these features has been sometimes considerably broadened, and traditional explanations have been replaced by new ones based on a better knowledge of the data. Thus, the index of grammar phenomena deserving further research has been enlarged. In addition, grammar variation phenomena have showed new areal configurations in Spanish dialectology, and moreover, the study of dialect grammar has also revealed itself as an important source for a better understanding of many cross-linguistic principles.

Key words: *Corpora of oral interviews vs atlases, dialect grammar*

Until recently, the study of dialectal variation of Spanish in the Iberian peninsula has been based on various regional atlases and those scarce dialectal monographs which devoted particular attention to Spanish (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*, *Corpus Oral y Sonoro del Español Rural*) is a corpus made up by recordings of rural speech which started to be compiled in 1990 to supplement those traditional sources and it has been growing since then.

1. An overview description

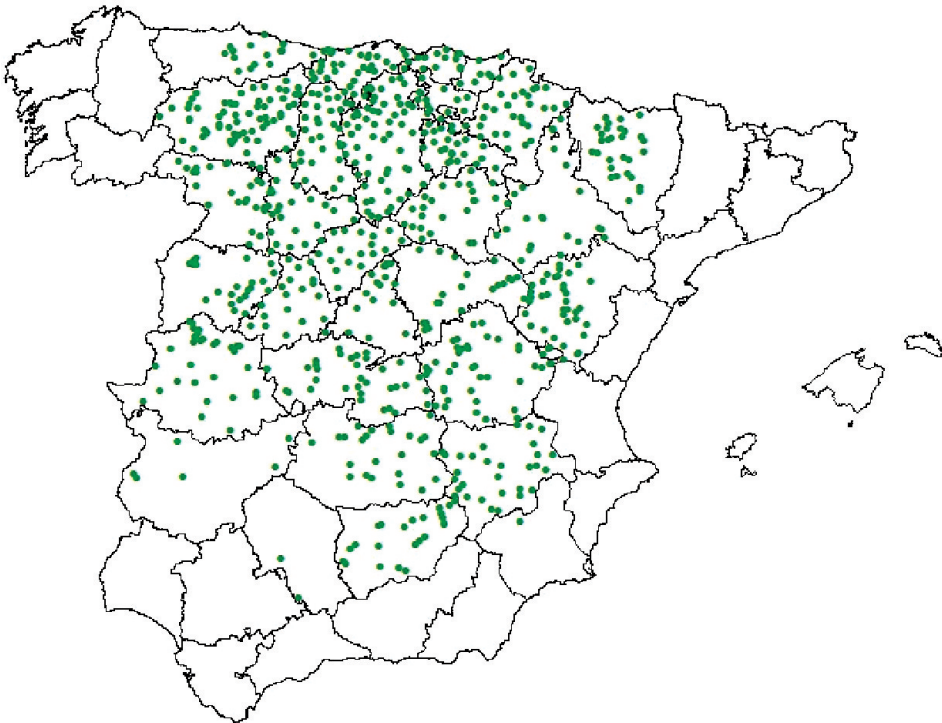
COSER is a corpus of oral interviews restricted to the speech of informants who were deemed interesting for traditional dialectology: rural speakers, elder if possi-

ble, of low education and natives of the place where they were interviewed. Actually, *COSE*R 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).

Regarding the number of informants of each *enclave*, in general one single person has preferably been thoroughly interviewed in *COSE*R, 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 *COSE*R, most of the times only one informant per *enclave* has actually been thoroughly surveyed as desired (almost the half). But sometimes we have made more than one interview per location.

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

*COSE*R consists presently of 940 hours of recording, but this number increases every year thanks to new survey campaigns. The final objective is to obtain record-



Geographic distribution of *COSE*R enclaves (2009)

ings of Spanish spoken in rural areas of the whole Iberian Peninsula. As can be seen in the map, the South is the main area still needed to be interviewed.

The average duration of the recordings is one hour and fifteen minutes (75 minutes) per location, 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. Methodology

The methodology used in *COSE*R has consisted in oral 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 need to recognise themselves as experts on 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 to document 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 locations, 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) it was decided to add subsequently the condition of loquacity ("informants who like talking") to the informants' selection protocol.

Interviewers are recruited among the students of Dialectology at the Autonomous University Autónoma of Madrid, after a methodological preparation. The students work in groups of four people, as most, to collect the data, but just one person acts as interviewer each time while the rest write down all external information about the informant's characteristics that could be relevant for linguistic analysis, and a linguistic profile of the phenomena documented. These notes are kept together with the recording.

Regards the selection of survey points, it has always intended to cover the communication network of the surveyed area —denser or looser, depending on the number of existing locations—. At the same time, geography and population have

been taken into account. In general, small villages have been preferred, and all the geographical major divisions of a certain area have been surveyed.

3. Aims

COSER is a corpus that aims to study grammatical variation. The study of dialect grammar of Peninsular Spanish has been almost non-existent until recently. Linguists have not paid much attention to it, mainly because there were no sources available to tackle these issues. Morphosyntax is traditionally an aspect hardly represented in dialectal monographs and in questionnaires of linguistic atlases. Therefore, it was generally presumed an apparent uniformity in Peninsular Spanish syntax together with the overall existence of certain non-standard uses, which were just mentioned to be avoided.

In this respect, *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. Standard languages seem to have a lower tolerance towards grammatical variation. Thus, this type of variables is frequently subject to a sociolinguistic filtration which may alter the linguistic principles that explain their original function. 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. This is the case, for instance, of the non-standard uses of the unstressed pronouns known as *leísmo*, *laísmo* y *loísmo*.¹ 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 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.

In addition, 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 sen-

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.

tences uttered in a real speech, in which it is possible to investigate contrastive values, and pragmatic motivations and inferences related to a specific structure. Thus, for instance, data from *COSEER* enable to understand better a structure which existed in ~~the old~~ Spanish and is only found nowadays in some specific rural varieties with a clear focal value: the use of the article followed by the 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.

Moreover, dialect grammar phenomena need a large amount of sentences (and possible variants of sentences) to be studied in order to get a fine analysis of the contrasting uses and to have the possibility of quantifying data: given a specific linguistic variable, the interview enables to quantify the variants in a specific location as well as distinguishing contexts of occurrence.

Instead, in traditional atlases this quantifying is not usually possible since one single answer is normally given for each location 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 ~~Penin-~~sula, i.e., the use of the conditional indicative (*-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). 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 indicative *-ría* (majority variant), but also by the imperfect indicative *-ba* (minority variant). Although both variants exist, 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 replaced 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 *COSEER*. First of all, because the phenomenon is sometimes recorded in contexts that were unexpected when atlas questionnaires were designed. This was indeed the case for *leísmo*, *laísmo* and *loísmo*.²

² Thus, the atlas for all the Romance languages in the Iberian Peninsula, *ALPI* (see Heap 2003) 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*

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*),³ while in fact the phenomenon appears in complement, relative, final, concessive, causal clauses, etc: i.e., in any subordinate clause where the imperfect subjunctive is likely to be found in Spanish. 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. The quantity of data makes it possible to identify focal areas for a linguistic phenomenon, which is not always possible with data coming from atlases, and to apply statistical tests like logistic regression, ena-

(*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*)). Regional atlases do not improve much ALPI questionnaire. *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 feminine personal *leísmo* (*A la madre la vio en la calle* (*The mother was seen in the street*), map 1713). Only *ALECan* 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*), *ALECan* 1195, *El paquete lo olvidé* (*I forgot the parcel*), *ALCyL* 116). These two regional atlases also reproduce questions 350, 352-353, 356, 359 and 362 of ALPI (*ALECan*, 1243, 1245-1247, 1192, 1197; *ALCyL*, 111-114, 117, 120) and 1713 of *ALEANR*. In *ALECMAN* the questions 350-353, 356, 359 y 362 from ALPI are included, and new questions are added to research *leísmo*: *A las niñas no (les/las) gusta estudiar* (*The girls do not like studying*), *La torre desde aquí se (le/la) ve* (*The tower, from here it is seen*). 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 feminine mass objects (*agua* (*water*), *miel* (*honey*), *mantequilla* (*butter*), etc), not even *ALECan*, in spite of the fact that Cantabria is a region where the existence of the mass neuter was well-described.

³ 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 *ALECan* (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 *ALECan* and *ALCyL* also inherited (maps 1218 and 150, respectively). *ALECan* 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*)).

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

COSER can measure the differences which may be found in the speech of socio-cultural groups with a lower education in rural areas. It therefore complements both linguistic atlases and 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 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.

4. Problems

Although *COSER* offers new possibilities for the research of dialects, it also has shortcomings and problems. First of all, 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' skills as well as the interaction between them; however, no testimony should be disregarded for that reason, since there is always some valuable information. Secondly, *COSER* methodology cannot avoid the problem of accommodation between the informant and the interviewer, or the challenging representativeness of the informant randomly chosen. Nevertheless, our experience suggests that the quantity of the data allows to circumvent these potential problems, since the data always show geographical coherence and make it possible to identify those informants who could be considered anomalous with their area and scarcely representative. Thirdly, the long time elapsed between the first interviews (1990) and the last (2009) and future ones (2010-) does not assure a total intercomparability of the data. However, the planning of the interviews has intended to research areas related to each other (for historical or linguistic reasons). For example, North and Center Castile has been researched between 1990 and 1995; East Castile and Aragón between 2001 and 2008, and so on. The documentation of an issue related to an area has usually finished in a reasonable time. Finally, 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. On the other hand, linguistic atlases offer a type of information which is not provided by corpora like *COSER*. Oral interviews have proved especially productive

to record phenomena of grammatical character but not as far as lexis is concerned. Since the data come from almost open conversations, the words recorded in COSER are not always repeated and no conclusions are drawn comparable to those of an atlas as regards vocabulary. Although the development of sociolinguistics has shown multiple limitations of atlas methodology, 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 and Benito 2009). Therefore, *COSER* aims to supplement the material collected in linguistic atlases as well as in other type of dialectal sources. *COSER* data are a supplement which opens up enriching prospects for the study of dialectal grammar.

It always has to be born in mind that success and adequacy depend on aims and results considered together with resources. Firstly, as regards aims, it is not *COSER* aim to study the lexicon or to obtain the elicitation of a precise structure or use in every place —which just could be done by means of a questionnaire—. Instead, we try to register spontaneous speech by following a similar thematic protocol in order to document, identify and study dialect grammar phenomena (known and unknown). Intercomparability is expected to be achieved between areas and not between locations. Secondly, as regards the relationship between aims and resources, *COSER* recordings started as a modest activity linked with the teaching of Dialectology, which aimed to document and understand grammar phenomena absent in linguistic atlases and traditional dialect monographs. Therefore, it can prevent neither the accommodation problem nor the time elapsed because it was not globally designed and planned as a research project, although *COSER* has also received support of several research projects. *COSER* interviews were collected by each year Dialectology students and annual campaigns have surveyed as many locations as possible depending on the availability of number of students, time and funding.

Despite all this, *COSER* is still a valuable source to study dialect grammar since anything alike exists for rural Spanish. By now only *COSER* offers a collection of oral interviews for the almost whole rural Spanish speaking territory. Moreover, the interest of *COSER* recordings reveals itself in the progress made in our knowledge of Spanish dialect grammar since 1990. This is true regarding the study and better understanding of dialect phenomena which were hitherto partially known —or even completely ignored by grammarians and dialectologists up to now—. The index of dialect grammar issues deserving further research has been considerably enlarged too. I will deal with these issues in the following sections.

5. Outcomes

There are hitherto three kinds of results coming out from *COSER* recordings: specific progress made in our knowledge of dialect grammar issues, general progress made in Spanish dialectology regards areal configuration, and finally theoretical outcomes.

Regarding the specific results, a number of dialect grammar issues so far have been researched and their geographical areas have been defined —confirming or dis-

carding areas already known, or describing areas for the first time—, a better description and understanding of the dialect structures have been achieved, and an explanation has been proposed. I will present three examples. Firstly, the analysis of the data from rural speech recordings has enabled to establish the exact geographic delimitation of the areas where the phenomena traditionally known as *leísmo*, *latísmo* and *loísmo* are 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). Secondly, 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 neuter agreement has been considerably enlarged to the South and a global explanation has been developed (Fernández-Ordóñez 2007, 2006-2007). Finally, COSER data quantifying has also made possible to demonstrate that the syntactic locus of origin for the replacement of subjunctive by indicative are complement clauses in Northern Spanish, and not the conditional clauses, and that the focal area is located in Castile, and not in the Basque Country, as usually believed (Pato 2003, 2004).

Our general knowledge of Spanish dialectology has improved so far in that new dialect areas of Peninsular Spanish have clearly emerged, namely a Western Spanish vs. an Eastern Spanish both layed out from North to South (Fernández-Ordóñez 2006). Also contrast between urban speech and rural speech regards grammar have made clear sociolinguistic selection of some grammatical features. On the other hand, contrast between data from old atlases and present oral data has showed the historical development of some dialect aspects, whether maintenance (subjunctive replacement Pato 2004 or inflected infinitives Pato / Heap 2009) or decline and loss (of some pronominal phenomena, see Heap 2006, Pato 2009).

Studies coming out from COSER have confirmed some theoretical generalizations concerning dialectology and linguistic typology. As for dialectology, concepts like mixed lect, fudged lect and scrambled lect, suggested by Trudgill (1986) as typical phonetic solutions of transition areas, have been confirmed to be operating in grammar variation (Fernández-Ordóñez, under review). As for typology, dialect grammar has revealed as an important source for a better understanding of many cross-linguistic principles which opens up new ways to test their validity and to achieve their refinement (see Kortmann 1999, 2004a, 2004b). For instance, out from COSER data and research, a proposal of refinement of the Agreement Hierarchy (Corbett 2006) has been proposed (Fernández-Ordóñez 2006-07), and two different syntactic pathways for the development of gender as a lexical category have been identified (Fernández-Ordóñez 2009).

6. Prospects

A small sample of COSER materials is now available in the Internet (www.uam.es/coser) as audio and text files: the recordings of 32 locations from 8 prov-

inces (c. 40 hours). The ~~recently~~ project “Variation and change in Peninsular Spanish Syntax” (FFI2009-10817) (2010-2012), funded by the Spanish National Research Funds, opens new prospects for COSER. Aiming to integrate COSER into the European Dialect Syntax (EDISYN) network, the project envisages two areas of activity, as is the case in the other European projects concerning dialect syntax (for example DYNASAND, see Barbiers 2006).

~~One area~~ will concentrate on the collection, processing and storage interview recordings which will allow for the study of syntactic variation in Spanish. To that end, our goal will be to digitalize, transcribe and web-publish at least 150 hours of recordings from COSER. So the interviews of 120 localities from 30 provinces —4 per province— will be available in the Internet as audio and texts files. It is also aimed the alignment of sound and text and the morpho-syntactic tagging of those 150 hours. At the end, there will be a searchable database in the Internet with several searching possibilities (by lemma, area or location, sex, age, and tags). The searched data will be retrieved in small listenable paragraphs.

~~The other area~~ will focus on the study of a set of syntactic variation phenomena within the framework of comparative and typological linguistics. This approach combines modern and historical data in order to achieve a unified analysis of the mechanisms which underlie and condition language change. To this end we will endeavour to integrate into our overall explanation current and historical linguistic perspectives with geolinguistic and, to some degree, sociological aspects of a set of phenomena which vary grammatically in Peninsular Spanish and which to date have only been partially investigated. The following grammatical issues will be part of ongoing research for the next three years: quirky dative marking in Eastern Spanish; possessives in Northern and Western Spanish, quirky quantifiers agreement; dative objects (with verbs such *help*, *call*, *follow*, *warn*, *scold*) and structures alternating dative and accusative; reflexive, passive, impersonal and middle diathesis associated with reflexive verbs and reflexive structures; inflected infinitives; diminutive suffixation related to word classes, semantic proprieties and syntactic functions, as well as to geographic areas; and ~~anterior and aoristic drift~~ in Peninsular Spanish dialects.

References

- ALCyL: Alvar, M., 1999, *Atlas lingüístico de Castilla y León*, 3 vols, Junta de Castilla y León, Valladolid.
- ALEANR: Alvar, M. et al., 1979-83, *Atlas lingüístico y etnográfico de Aragón, Navarra y La Rioja*, with the collaboration of A. Llorente, T. Buesa and E. Alvar, 12 vols, Institución Fernando el Católico / CSIC, Zaragoza.
- ALECanr: Alvar, M. et al., 1995, *Atlas lingüístico y etnográfico de Cantabria*, 2 vols, Fundación Marcelo Botín, Madrid.
- ALECMAN: García Mouton, P. & F. Moreno Fernández (dirs), *Atlas Lingüístico (y etnográfico) de Castilla-La Mancha*, Universidad de Alcalá, <<http://www2.uah.es/alecman>>
- ALPI: Navarro Tomás, T. (dir.) et al., 1962, *Atlas lingüístico de la Península Ibérica*, vol. 1, *Fonética*, with the collaboration of F. de Borja Moll, A. M. Espinosa [junior], L. F. Lindley Cintra, A. Nobre de Gusmão, A. Otero, L. Rodríguez-Castellano and Manuel Sanchis Guarner, Madrid, CSIC.

- Barbiers, S. et al, 2006, *Dynamic Syntactic Atlas of the Dutch dialects* (DynaSAND), Amsterdam, Meertens Institute <<http://www.meertens.knaw.nl/sand/>>.
- Benito, C. de, 2009, *Descripción y análisis del pronombre concordado con el sujeto en el Atlas Lingüístico de la Península Ibérica*, Master's degree memory, UAM.
- Chambers, J. K., 1995, *Sociolinguistic theory*, Blackwell, Oxford.
- Corbett, G. C., 1991, *Agreement*, Cambridge U. P.
- COSER: Fernández-Ordóñez, I. (dir.), 2005-, *Corpus oral y sonoro del español rural*, Universidad Autónoma de Madrid, Madrid <<http://www.uam.es/coser/>>.
- Fernández-Ordóñez, I., 1993, «Isoglosas internas del castellano. El sistema referencial del pronombre átono de tercera persona», *Revista de Filología Española* LXXIV, 71-25.
- , 1999, «Leísmo, laísmo y loísmo», in I. Bosque & V. Demonte (ed.), *Nueva gramática descriptiva de la lengua española* (3 vols.), Espasa-Calpe, Madrid, vol. 1, 1317-1397.
- , 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.
- , 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.
- , 2007, «El neutro de materia en Asturias y Cantabria. Análisis gramatical y nuevos datos», in I. Delgado Cobos & A. Puigvert Ocal (ed.), *Ex admiratione et amicitia. Homenaje a Ramón Santiago*, Ediciones del Orto, Madrid, 395-434.
- , 2009, «The development of mass / count distinctions in Indo-European languages», in V. Bubenik, J. Hewson & S. Rose (ed.), *Grammatical Change in Indo-European Languages* (Current Issues in Linguistic Theory 305), John Benjamins, Amsterdam / Philadelphia, 55-68.
- (under review), «Dialect areas and linguistic change: Pronominal paradigms in Ibero-Romance dialects from a cross-linguistic and social typology perspective».
- Heap, D., 2000, *La variation grammaticale en géolinguistique: les pronoms sujet en roman central*, Lincom Europa, München.
- , 2003-, *Atlas lingüístico de la Península Ibérica. ALPI searchable database*. University of Western Ontario, London, Ontario <<http://www.alpi.ca/>>.
- , 2006, «Secuencias «invertidas» de clíticos (?): un cambio (?) en tiempo real», in J. J. de Bustos & J. L. Girón Alconchel (ed.), *Actas del VI Congreso Internacional de Historia de la Lengua Española* (Madrid, September 29-October 3, 2003), Arco/Libros, Madrid, I, 785-98.
- Klein-Andreu, F., 1979, «Factores sociales en algunas diferencias lingüísticas en Castilla la Vieja», *Papers: Revista de Sociología* 11, 46-67.
- , 1981, «Distintos sistemas de empleo de 'le, la, lo': perspectiva sincrónica, diacrónica y sociolingüística», *Thesaurus* XXXVI, 284-304. (Reprinted in O. Fernández Soriano (ed.), *Los pronombres átonos*, Taurus, Madrid, 1993, 337-353).
- , 2000, *Variación actual y evolución histórica: los clíticos le/s, la/s, lo/s*, Lincom Europa, München.
- Kortmann, B., 1999, «Typology and dialectology», in B. Caron (ed.), *Proceedings of the 16th international Congress of Linguists*, CD-ROM, Elsevier Science, Amsterdam.
- , 2004a, «Why dialect grammar matters», *The European English Messenger* XIII, 24-29.
- (ed.), 2004b, *Dialectology meets typology. Dialect Grammar from a Cross-Linguistic Perspective*, Mouton de Gruyter, Berlin /New York.
- Pato, E., 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.
- , 2004, *La sustitución de cantaral/ cantase por cantaríal/ cantaba (en el castellano septentrional peninsular)*. Universidad Autónoma de Madrid, Madrid <http://joule.gfa.uam.es/coser/publicaciones/enrique/2_es.pdf>

-
- , 2009, «Nivelación lingüística y simplificación: el uso de preposición + *tú* en la historia de la lengua», in Emilio Montero et al, *Actas del VIII Congreso Internacional de Historia de la Lengua Española* (Santiago de Compostela, 14-18 de septiembre de 2009), in press.
- & D. Heap, 2009, «Plurales anómalos (el morfema verbal *-n*) en los dialectos y en la historia del español», in E. Montero et al, *Actas del VIII Congreso Internacional de Historia de la Lengua Española* (Santiago de Compostela, 14-18 de septiembre de 2009), in press.
- Trudgill, P., 1986, *Dialects in contact*, Blackwell, Oxford.