SYNTACTIC MICROVARIATION IN CATALAN. THE *MICROVARIACIÓ.CAT* INITIATIVE[1]

ARES LLOP[2], ANNA PINEDA[3]

Abstract. In this paper we present the *Microvariació.cat* initiative (www.microvariacio.cat), whose main goal is to identify, compile and systematically organise all the research done regarding Catalan syntax in order to design a website available to specialists and the general public. We reflect on the value of investigating the syntactic variation of languages and we prove how the *Microvariació.cat* initiative is the starting point for a major challenging study on the microsyntax of Catalan. We characterise the different aspects of the initiative and we describe the methodology followed to build up a database integrated by an exhaustive list of Catalan microvariation phenomena and a set of reference works on the different topics and areas listed.

Keywords: syntactic variation, microsyntax, Catalan, database.

1. INTRODUCTION

This paper presents the *Microvariació.cat* initiative (www.microvariacio.cat), whose main aim is to identify the linguistic structures that are subject to syntactic variation across Catalan dialects, in order to organise them in a website available to specialists and the general public. In order to provide the appropriate context of the initiative, in section 2 we reflect on the value of investigating the syntactic variation of languages. Then, in section 3, we analyse the state of the field in the particular case of Catalan, in order to make clear how the *Microvariació.cat* initiative aims at filling an existing gap in the study of the syntactic variation of Catalan. We devote section 4 to a description of the different aspects of the initiative, from its methodology to its intended functions, including the configuration of the database. We set forth our conclusions and present prospective research issues in section 5.

---

[2] Universitat de Barcelona, aresllop@gmail.com.
[3] Universitat Pompeu Fabra, anna.pineda@upf.edu.

2. THE NATURE OF SYNTACTIC RESEARCH

Research on syntactic variation has three broad directions: (i) the discovery of syntactic isoglosses, (ii) the utility of syntactic dialectal data for making typological comparison and theoretical points of interest and (iii) its repercussions for codification.

First of all, given that "traditional dialectology concentrated on the relationship between language and geography, and on the spatial differentiation of language" (Chambers and Trudgill 1998: 57), its earliest purpose was to gather "pure" and thorough evidence of variation in different territories. Nowadays, the ability to plot the geographical distribution of variation in terms of certain variables is still one of the most obvious benefits of dialectological research, with the drawing of isoglosses across maps to mark the boundaries between areas of differing syntactic usage the most visual way to represent such variation. According to Bucheli and Glaser (2002), "it is a question of general linguistic interest whether syntactic isoglosses within given dialect areas do exist". This in-depth knowledge of the distribution of syntactic variants allows us to better explore and understand the nature of syntactic isoglosses, the factors that contribute to their configuration, and the possible interaction between the distribution of syntactic areas and the extralinguistic characterisation of the varieties. That said, it must be borne in mind "that syntactic areas normally are much larger than phonological ones, and that, perhaps as a result, regional syntactic variants often are not recognized as such even by conscious speakers" (Bucheli and Glaser 2002: 46).

Studying variation in syntax is also of paramount importance from the formal point of view. This value was emphasised by Hale (1997: 71), who noted that linguistic diversity "is a natural consequence of the very nature of linguistic structure", and it is therefore necessary to understand "both the truly elemental principles of human linguistic knowledge and the full range of possibilities permitted by them". Nowadays, assuming this premise, generative linguists claim that studying variation is one route to achieving a state of linguistic theory that captures all natural languages and dialects—and, as a consequence, the Language Faculty (see Picallo 2014). In this multi-pronged endeavour to identify language universals as well as the patterns and limits of variation, microsyntactic research runs parallel with broader typological studies (see Anderwald and Kortmann 2002), but "with the advantage that the field of investigation is magnified by the close similarity of the languages under investigation" (Cornips and Poletto 2005: 2). Kayne (1996: xvii) also highlighted the importance of research in dialect syntax for theoretical linguistics:

> "[M]icrocomparative syntax is a powerful tool, whose growth is perhaps to be compared with the development of the earliest microscopes, that allows us to probe questions concerning the most primitive units of syntactic variation. And since the invariant principles of UG [Universal Grammar] can hardly be understood in isolation from syntactic variation, this tool promises to provide invaluable evidence that will shape our understanding of those principles themselves."[4]

---

[4] According to Benincà and Poletto (1991: 80): "Il lavoro non si configura quindi come una raccolta di datti fatta allo scopo di semplice documentazione ingenua o preteorica, ma come un esame di alcuni fenomeni linguistici in base a dei parametri di variazione sintattica prestabiliti, individuati sulla scorta di una teoria". Therefore, a symbiosis between field linguistics and formal research, such as the one presented here, must be the horizon for microsyntactic studies.

Similarly, Rigau (2004: 1) points out that it is only if we study comparatively the small and big differences across languages that we will be able to explain how they are formed and the extent to which they are similar or diverge, and also what the limits of such variation may be. Additionally, synchronic microsyntactic differences displayed in dialects can be taken as a mirror of diachronic language patterns and their historical dynamics (see Benincà 2004)

Finally, the study of syntactic variation can play a useful role in the process of linguistic codification, that is, the standardization and development of prescriptive norms for a language. This is crucial in the case of minority varieties, such as those that we will focus on below. The codification of Catalan is centred around what is known as Central Catalan, spoken in the more densely populated central and north-eastern areas of Catalonia, including Barcelona, and, although Standard Catalan is compositional and polymorphic, the truth is that, as usually happens, the social prestige of varieties spoken outside this area is often undermined. However, if the different dialectal variants are codified, this can have the effect of boosting the self-confidence of these dialects as well as value that they associate with the use of their native variety (see Llop and Pineda 2017). As Rigau (1998: 63) explains, syntactic variation is sometimes regarded by grammarians or even by speakers as a vulgarism, a deviation from the standard language, whereas on other occasions it may be seen as a sign of linguistic richness. Needless to say, the presence of dialectal syntactic variants in the standard or normative varieties (as is the case with Standard Catalan) is a sign of a conciliatory and respectful model of codification.[5]

3. STUDY OF SYNTACTIC VARIATION IN CATALAN: WHERE ARE WE?

As occurs with many other languages, research on syntactic variation in Catalan is still relatively sparse. That is, works or research projects that focus specifically on such variation have been – at least until very recently – practically non-existent. In what follows we will give some idea of the stage at which this slim body of research finds itself. As is well known, the study of linguistic variation was traditionally undertaken from a perspective that granted privilege to the lexical, morphological and phonetic domains (see, among others, Labov 2001; Chambers and Schilling-Estes 2013), whereas syntactic phenomena were often left out. As stated by Rigau (1998), although Catalan dialectology enjoys a fairly extensive bibliography at this point, this language is no exception to the traditional tendency to neglect syntactic phenomena when assessing dialectal variation. Rigau's (1998: 79) conclusion is categorical: the truth is that we know very little about Catalan syntactic variation. This becomes immediately clear if one looks at the type of data gathered in the existing sources for the study of Catalan dialectal variation, such as the *Atles Lingüístic del Domini Català* (see also Badia i Margarit 1993), or the materials compiled by Alcover (the so-called *Diccionari català-valencià-balear*, available online, and the *Quaderns de camp*, which are also accessible online through the data bases included in the Portal Alcover (https://alcover.iec.cat/). Such sources usually offer variant

---

[5] The fact that the new prescriptive grammar of Catalan published by the Institut d'Estudis Catalans (GIEC) accepts the use (in restricted contexts) of many syntactic dialectal variants embodies an acknowledgement that variation constitutes a source of linguistic richness that must be preserved.

names of a given object (i.e., lexical variation), different realizations of a verbal form (i.e., morphological variation) and different pronunciations of a given word (i.e., phonetic variation), whereas syntactic variation is reflected by the inclusion of several variants that, though conceivably useful for a tentative preliminary speculation, would be grossly inadequate for an in-depth microsyntactic study. The same thing is true of the *Corpus Oral Dialectal* (COD) of Catalan (Viaplana *et al.* 2007), developed at the Universitat de Barcelona. This corpus, which consists of a compilation of questionnaires and semi-directed interviews involving speakers from all over the Catalan-speaking territories, has proven an outstanding and invaluable tool for the study of the lexical, morphological and phonetic levels of the language, but it fails to provide an adequate set of data for the study of the syntactic level. The same holds for smaller corpora such as the *Corpus Aixalda* of the Pyrenean area, developed by the Universitat de Lleida's Oficina de Llengua i Literatura de Ponent i del Pirineu. Although a long list of syntactic variants can be obtained from the transcriptions of semidirected interviews contained in this corpus, it is not a tool especially designed for the study of syntax. Finally, there exist several monographs devoted to the description of particular dialects and varieties. Here again, however, though a few references to syntactic phenomena can be found, these are usually indirect, and fall more properly under the heading morphosyntax rather than syntax as such.

In sum, syntax has tended to be neglected by dialectology. This tendency is due in part to the (untrue) belief that syntax is a domain in which little variation occurs, but it is also a result of the decades-old lack of formal machinery to appropriately analyse the syntactic variation we find not only across languages but also across the different dialects of a given language. In this context, recent formal approaches to language have become crucial for the study of syntactic variation, most particularly the Generative framework (Chomsky 1986, 1995). In this framework, language is conceived as a biologically innate endowment of human beings that consists of universal Principles and language-specific Parameters that can be set in different ways, thus yielding variation. Specifically, *micro*parameters have small-scale consequences and vary across typologically close languages or varieties, yielding what is known as *microvariation* (see, among others, Barbiers *et al.* 2002; Barbiers 2012). The Generativist approach constitutes a powerful tool to account for syntactic variation across and within languages, and it is for this reason that the *Microvariació.cat* initiative, which will be described in fuller detail below, is based on this theoretical model.

At the same time, over the past few years, the fact that dialect syntax has received very little attention relative to other levels of language has not passed unnoticed, and this imbalance is now beginning to be corrected. The present project forms part of that effort. Interestingly enough, in the field of Catalan linguistics, the transition to a new syntactic understanding of dialectal variation has catalysed an upsurge in research (including some doctoral dissertations, journal articles and papers in conference proceedings) dealing with particular syntactic phenomena. Two important milestones within this tendency are the *Gramàtica del català contemporani* (Solà *et al.* 2002) and the *Gramàtica de la llengua catalana* (IEC, 2016), wherein the concept (and specific instances) of syntactic variation is discussed on several occasions and illustrated with specific examples. Nonetheless, there are no specific works or projects centred on the field of Catalan dialect syntax. One must reluctantly conclude, therefore, that the study of Catalan syntactic microvariation remains at a preliminary stage, with the work done and data available thus far constituting merely

the starting point for far more detailed formal studies. All in all, a great deal remains to be investigated, and the initiative we will describe now is intended as a significant first step in that direction.

4. FILLING A GAP: THE *MICROVARIACIÓ.CAT* INITIATIVE

Microvariació.cat is an unprecedented initiative in the field which is intended to make a substantial contribution to the systematic study of syntactic variation in Catalan. The initiative is the continuation of an earlier exploration of the state of affairs in the study of Catalan dialect syntax, carried out by two of the team members (see Llop and Pineda 2017). The motivation was clear: although Catalan has a long tradition of dialectal studies focusing on phonology, morphology and lexicology, research on micro-syntactic variation is still at a preliminary stage. General properties of the language have been described and some syntactic variants have been noted, but the phenomena of dialectal syntax remain relatively unexplored, data are sparse and unsystematic and there exists as yet no dialectal syntax-oriented annotated corpus.

In its initial phase the *Microvariació.cat* initiative aims to fill this gap by collecting and making universally accessible a constantly updated online database (www.microvariacio.cat) that incorporates all existing and future studies dealing with any aspect of Catalan syntactic microvariation.[6] Before detailing the features of the database (section 4.2), we will first describe the methodology underlying its creation (section 4.1). We will then go on to present the main goals of the initiative (section 4.3) and finally discuss its links with similar projects in the field of syntactic variation (section 4.4).

4.1. Methodology

The *Microvariació.cat* initiative follows the methodological protocols designed and consolidated in the framework of other micro-comparative syntax projects that have successfully gathered data on syntactic variation in other European languages and dialects. In particular, we have applied what is known as a *layered methodology* for the collection of large amounts of data (Cornips and Poletto 2005, 2007), a kind of stepwise procedure for the study of linguistic structures used to determine "how [microsyntactic] phenomena themselves and relations among phenomena can be discovered and brought to the attention of the linguistic community" (Cornips and Poletto 2007: 4).

As represented in Figure 1, layered methodology approaches the study of the syntactic properties of a dialect starting with a broad survey, and then progressively narrows the target until it detects something of potential interest for microcomparative linguistic research. Working on the basis of intuition, the preliminary survey consists of a literature review on the grammar of the language in question, with each work being carefully checked for any mention of syntactic phenomena subject to variation. This process leads to a careful description of the general properties of the language.

---

[6] Though at the moment the primary focus of the project is Syntax-Semantics, we plan to include Phonology and Morphology as well.
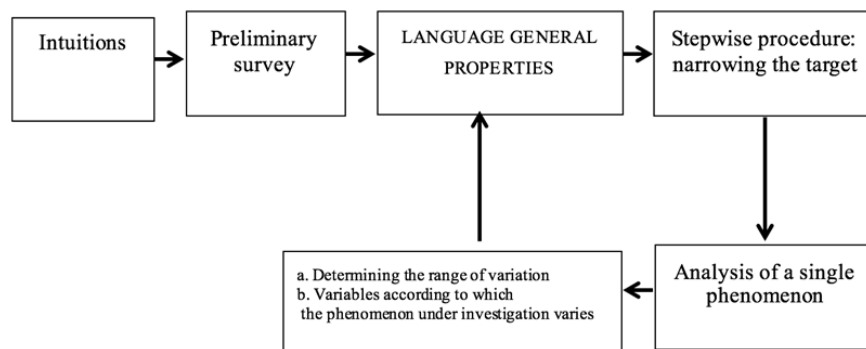
Figure 1. The layered methodology template. Adapted from Cornips and Poletto (2000).

In the layered methodology approach, once this initial broad overview is completed and a number of interesting phenomena have been identified, the target of study is progressively narrowed down and detailed analyses of each phenomenon deemed of interest is undertaken. In other words, once the general properties of a language have been studied, it is time to single out phenomena and subject it to what Cornips and Poletto (2007) call *in vitro* observation, without any other phenomena interfering. To this end, specific questionnaires are developed and other data collection techniques are deployed that will allow the researchers to delimit the distribution, extension and intricacies of each phenomenon, as well as its connections with other linguistic variables via the identification of clusters of similar properties. Given the great heterogeneity of syntactic phenomena, it is important to determine which specific techniques will be most appropriate to obtain data relevant for the study.

In our application of the layered methodology approach to the case of Catalan, informed intuition has already led to the identification of a considerable number of syntactic phenomena of potential interest, and the literature review is already well underway (recall section 3). In the next steps, a wide-focus survey of the syntactic phenomena displaying syntactic variation in Catalan will be undertaken, followed by a increasingly narrow-focused compilation of data.

4.3. The database

As noted at the beginning of section 2, the first phase of the present initiative consists of collecting and organising all the existing references dealing with any phenomenon of Catalan syntactic variation. These references, of all sorts, are organised into a database. For each of them a set of descriptive information is specified (see Figure 2) in order to make it easier to search the online database according to different variables. Before going into detail about the makeup of the database, it is important to note that, in accordance with the layered methodology approach, this initial phase will be followed by a second one, whose point of departure will be not references, but rather phenomena. In other words, once the full set of reference works has been included in the database (albeit updated regularly), additional descriptive information will be added as further syntactic phenomena subject to variation across Catalan dialects are detected.

Label	General descriptor	Specific descriptor	Variety	References	Approach
VP	Unaccusativity	Subject position, agreement and other related phenomena	Not specified	Rigau (1991) Rigau (1992) Rigau (1994) Rigau (1998)	F
VP	Unaccusativity	Subject position, agreement and other related phenomena	North-Western	Alonso and Suïls (1998)	NF
VP	Unaccusativity	Subject position, agreement and other related phenomena	North-Western	Alonso and Suïls (1998) Massanell (2011)	NF
VP	Unaccusativity	Subject position, agreement and other related phenomena	Varieties of Cambrils	Navarro (2012)	NF
VP	Unaccusativity	Subject position, agreement and other related phenomena	Varieties of the Ribera d'Ebre	Cubells (2005/2009)	NF
VP	Unaccusativity	Subject position, agreement and other related phenomena	Ribagorçan	Suïls, Sistac, Rigau (2006)	F
VP	Unaccusativity	Subject position, agreement and other related phenomena	Roussillonese	Gómez (2012)	NF

Figure 2. Sample of the information included in the database for each phenomenon studied.

Figure 2 includes all the descriptors used in the database. Recall that references constitute the source of data in the first phase: for each item, the particular linguistic phenomenon or phenomena is identified, and then one of the linguistic *macrotags* listed in (1) is assigned, depending on the particular area of grammar each phenomenon belongs to. The same will apply in the second phase, where phenomena will be identified through research strategies other than scanning the existing references (recall section 4.1).

(1)   a. VP: Verb Phrase
         b. DP: Determiner Phrase
         c. AdjP: Adjective Phrase
         d. NP: Noun Phrase
         e. PP: Preposition Phrase
         f. PolP: Polarity Phrase
         g. AdvP: Adverbial Phrase
         h. Other: Other sentential particles

Additionally, in the description accompanying each database entry, every phenomenon is related to both a general descriptor (e.g., 'unaccusativity') and a specific descriptor (e.g., 'subject agreement'), as shown in Figure 2.[7] In all cases, information is given about the dialects (and subdialects) in which the construction presenting microsyntactic variation occurs. The database also provides a list of all the references, that is, all the articles, books, chapters, dissertations or other documentary sources, in which the phenomenon is studied or mentioned: this field actually constitutes the source of information in the first phase of the initiative. This list of references includes information about the type of the approach applied in the study, specifically whether it is formal (F) or non-formal (NF).

4.3. Intended functions of Microvariació.cat

It is hoped that the *Microvariació.cat* initiative will primarily serve to shed light on a wide range of under-studied or completely unexplored data related to syntactic variation in Catalan, in order to make them interpretable, interesting and relevant for scholars, from dialectologists to theoretical syntacticians, as well as sociolinguists, typologists, and even teachers and students of Catalan. Thus our intention is that this tool will pave the way for countless monographs, theses, journal articles and conference papers, each of them devoted to a detailed analysis of an important syntactic phenomenon.

Ideally, specific research projects associated with the initiative will be able to develop a detailed and innovative characterization of the syntactic variation found across Catalan dialects, in order to assess the main foci of variation, the associated geographic areas, and the linguistic variables involved in each case. By basing themselves on the *Microvariació.cat* initiative, subsequent funded research projects may achieve a more precise understanding of variation in Catalan dialects, on the one hand, and the general properties of the language, on the other, thereby informing linguistic theory and ultimately refining the existing proposals and analyses. More specifically, *Microvariació.cat* should contribute significantly to the defining of syntactic isoglosses in the Catalan linguistic domain. Ultimately, the typological comparison of the syntax of Catalan dialects with that of other languages is also a feasible outcome of the initiative. Finally, contributing to the codification of Standard Catalan (and the way it is taught) may be possible in those cases where an accurate study of dialectal variation leads to a better comprehension of linguistic structures and to the removal of the social stigma associated with some of them.

4.4. Connection with other projects and initiatives

The potential for linkage with other projects and initiatives is great. For example, the *Microvariació.cat* initiative can directly contribute to the Formal Approach to Romance Microvariation (FARM) initiative. We think that the particular set of references marked as F in our data set (i.e., those that specifically deal with a phenomenon within a formal perspective) could easily be included within the pan-Romance FARM initiative. The

---

[7] Nowadays the list of general descriptors includes 93 different items and the list of specific descriptors, 191.

contribution of *Microvariació.cat* to the FARM initiative would thus be twofold: it would contribute an organised set of references and it would provide an exhaustive classification of Catalan microvariation phenomena.

That said, it is worth noting that this ambitious initiative is for the time being a non-funded collaboration altruistically started by a team of young researchers headed by the authors of this paper, who are also contributing to the development of in-depth studies and data collection through their own research on various microsyntactic phenomena (see for example the dissertation on the microsyntax of negative minimizers by Llop (2017) and the dissertation on dative/accusative case alternations and ditransitivity by Pineda (2014, published in 2016); the research on clitic climbing in verbal complexes by Paradís (in preparation) and on the interaction between argument structure and participial agreement by Salvà (in preparation); the joint work by Llop and Paradís (2017) on the interaction between presuppositional negation and clitic climbing in verbal complexes). Their work constitutes a sample of the sort of fine-grained studies that that will be made possible in the near future thanks to the data collected in the *Microvariació.cat* project. However, given the open-ended nature of the project, it is clear that the non-funded model is unsustainable in the long run. For this reason, the *Microvariació.cat* initiative is currently making a call for funding.

5. CONCLUSIONS

Summing up, we have presented an initiative that aims to compile and systematically organise all the research done regarding Catalan syntax studies. The database presented is conceived as the starting point for a major project devoted to the in-depth study of the microsyntax of Catalan dialects. The initiative is now ready to present an exhaustive classification of Catalan microvariation phenomena and to provide an organised set of reference works related to microsyntactic studies as well, in order to provide the linguistic community with an exhaustive database for further research work and projects in this area.

REFERENCES

Anderwald, L., B. Kortmann, 2002, "Typology and dialectology: A programmatic sketch", in: J. Berns, J. Marle (eds), *Present-Day Dialectology. Problems and Findings*, Berlin, Mouton de Gruyter, 159–171.

Badia i Margarit, A., 1993, *Atles Lingüístic del Domini Català: qüestionari*, Barcelona, Institut d'Estudis Catalans.

Barbiers, S., 2013, "Where is syntactic variation?", in: P. Auer *et al.* (eds), *Language. Language Variation – European Perspectives IV*, Amsterdam, Philadelphia, John Benjamins Publishing Company, 1–26.

Barbiers, S. *et al.* (eds), 2002, *Syntactic Microvariation*, Amsterdam, Electronic publication of the Meertens Instituut.

Benincà, P., 2004, "Dialetti d'Italia e dialetti d'Europa", *Quaderns d'Italià,* 8/9, 11–26.

Benincà, P., C. Poletto, 1991, "Il modelo generativo e la dialettologia: un'indagine sintattica", *RID*, 15, 77–97.

Bucheli, C., E. Glaser, 2002, "The Syntactic Atlas of Swiss German Dialects: empirical and methodological problems", in: S. Barbiers, L. Cornips, S. VanderKleij (eds), *Syntactic Microvariation,* Amsterdam, Meertens Institute Electronic Publications in Linguistics, 41–74.

Chambers, J.K., N. Schilling-Estes, 2013, *The Handbook of Language Variation and Change*, Chichester, Wiley.

Chambers, J.K., P. Trudgill, 1998, *Dialectology*, Cambridge, Cambridge University Press.

Chomsky, N., 1986, *Knowledge of Language,* New York, Praeger.

Chomsky, N., 1995, *The Minimalist Program*, Cambridge, MIT.

Cornips, L., C. Poletto, 2005, "On standardising syntactic elicitation techniques" (part 1), *Lingua*, 115, 939–957.

Cornips, L., C. Poletto, 2007, *Field linguistics meets formal research*, ICLaVE.

Kayne, R.S., 1996, "Microparametric syntax. Some introductory remarks", in: J.R. Black, V. Motapanyane (eds.), *Microparametric Syntax and Dialect Variation*, Benjamins, Amsterdam, ix-xviii.

Labov, W., 2001, *Principles of Linguistic Change*, Chichester, Blackwell.

Llop, A., 2017, *La reanàlisi dels minimitzadors negatius en el continuum romànic pirinenc,* PhD dissertation, Universitat Autònoma de Barcelona.

Llop, A., A. Paradís, 2017, "Clitic climbing and presuppositional negative markers: a meeting point in restructuring contexts", Linguistic Simposium on Romance Languages, Newark, Delaware, 19-23 April 2017.

Llop, A., A. Pineda, 2017, "L'estudi de la variació sintàctica en català: On som i cap on anem?", in: M. Pérez Saldanya, R. Roca Ricart (eds), *Actes del XVIIè Col·loqui Internacional de Llengua i Literatura Catalanes* [online], 527–542.

Hale, K., 1997, "Some observations on the contributions local languages to linguistic science", *Lingua,* 100, 71–89.

Institut d'Estudis Catalans, 2016, *Gramàtica de la llengua catalana*, Barcelona, Institut d'Estudis Catalans.

Paradís, A., in prep., *L'ascens dels clítics en català,* PhD dissertation, Universitat Autònoma de Barcelona.

Picallo, M. C. (ed.), 2014, *Linguistic Variation in the Minimalist Framework*, Oxford, Oxford University Press.

Pineda, A., 2016, *Les fronteres de la (in)transitivitat. Estudi dels aplicatius en llengües romàniques i basc* [published and revised version of the PhD dissertation defended in 2014], Barcelona, Institut d'Estudis Món Juïc. Col·lecció Cum Laude, 6.

Rigau, G., 1998, "La variació sintàctica: Uniformitat en la diversitat", *Caplletra*, 25, 63–82.

Rigau, G., 2004, "El quantificador focal *pla*. Un estudi de sintaxi dialectal", *Caplletra*, 36, 25–54.

Salvà, S., in prep., *La concordança del participi i la selecció de l'auxiliar en el català de Mallorca: la rellevància de l'aspecte intern*, PhD dissertation, Universitat Autònoma de Barcelona.

Solà, J. *et al.* (dirs), 2002, *Gramàtica del català contemporani*, Barcelona, Empúries.

Viaplana, J. *et al.,* 2007, *Corpus Oral Dialectal*, Barcelona, Promociones y Publicaciones Universitarias.