## Background

This project focuses on a striking parallelism between two macro-groups of languages: southern Italian (Romance) dialects and the so-called split-ergative languages, like Basque, Georgian, Dyirbal, Hindi/Urdu, and many others. Surprisingly, these two groups of languages, which have otherwise very little in common, both organize the grammatical material in a very peculiar fashion, by grouping some elements together to the exclusion of others.

Consider the following data from Abruzzese, a southern Italian dialect (1), and then Diyrbal, an Australian, Pama-Nyungan language (2):

(1)

| | | | |
|---|---|---|---|
| a. So'/si | magnate | c. Seme/sete | magnite |
| be-1st sg/2nd sg | eaten | be-1st pl/2nd pl | eaten |
| 'I have | eaten' | 'We have | eaten' |
| b. A | magnate | d. A | magnite |
| have-3rd | eaten | have-3rd | eaten |
| 'He has | eaten' | 'They have | eaten' |

In (1a,c) the auxiliary *be* is used, whereas in (1b,d) the auxiliary *have* is used. In Abruzzese, 1st and 2nd person subjects (i.e. *I, you, we, you(pl)*) require the use of the auxiliary *be*, while the 3rd person subjects (*he, she, Mary, they, Mary & John, ...*) require the auxiliary *have*. This is true for all verbs (Rohlfs 1969, Tuttle 1986, Manzini&Savoia 2005). Hence, we can conclude that there is a *split* in the person information (1/2 vs. 3) which triggers the selection of *be* (for 1/2) or *have* (for 3). Let us now turn to Dyirbal:

(2) a. Nana banaga-n$^{yu}$
1st pl-Nom return-Nfut
'We returned.'

b. Numa banaga-n$^{y}$u
3rd sg.father-Abs return-Nfut
'Father returned.' [from Dixon 1994:14]

In Dyirbal, a different case system altogether is used if the subject is 1st/2nd or 3rdperson. 3rd person 'subjects' (for the exact definition of subjects in ergative languages see Dixon 1972), like *father* in (2b), have an ergative/absolutive case system, 1st and 2ndsubjects, like *we* in (2a) have a Nominative/Accusative case system.

Dyirbal is obviously very different from Abruzzese, but the comparison of (1) and (2) strongly suggests that we are facing the same pattern: 1st and 2nd person on the one hand and 3rd person on the other. In other words, in (2) we are facing again a person *split*(1/2 vs 3). Auxiliary selection (1) or the use of different case (2) are just flags that tell us that a split has been performed in grammatical information between 1st and 2nd person on the one hand and 3rd person on the other. This person split is quite common in southern Italian dialects as well as in split-ergative languages, but quite rare in other languages. Observe furthermore that (2) is illustrating only one of the possible alternations that can be found in ergative languages (see Nichols 1992 and Dixon 1994 for a complete list).

Aside from the person split, there are many other possible splits that cause a *have-be*alternation in Romance and Germanic languages and case alternation (or other kinds of alternations) in split-ergative languages (Dixon 1972, 1994, Silverstein 1976, DeLancey 1981, Marantz 1984, Mahajan 1994, Bittner&Hale 1996, Nash 1996, Butt 2006, van de Visser 2006). The most commonly found *have-be* alternation in Romance is triggered for example by the kind of verb used (see Perlmutter 1978, Hoekstra 1984, Burzio 1986, Tuttle 1986, Kayne 1993, Cocchi 1995, Van Valin 1990, Sorace 2000, forth. Bentley&Eythórsson 2003, Aranovich 2007, Loporcaro 2007 among others). There are however many other possible triggers, like the tense/aspect/mood of the verb (*have* or *be*is selected depending on whether we are in the indicative or in the subjunctive, for instance) (Rohlfs 1969, Giammarco 1979, Tuttle 1986, Kayne 1993), or sentence-subordination (*John has watched football* vs *I think that John has watched football*) (Cennamo 1999, 2002, 2007, Ledgeway 2007, Formentin 2001).
In what follows, we refer to person, tense, aspect, mood, etc. with the cover term 'grammatical information'. We use the term *split* to refer to the divisions in this grammatical information (like the person split in (1)-(2)). Observe that the term split is sometimes used in the literature to indicate the *have-be* alternation. We will not use it this way in this project.
Of all Romance and Germanic languages, splits are most common in southern Italian dialects. No other Romance or Germanic group exhibits an equivalent wealth.

At first sight, the *have-be* alternation seems to be the result of some historical accident, and does not appear to tell us anything about the complexity of linguistic systems. However, quite surprisingly, exactly the same splits that determine the *have-be* alternation in southern Italian dialects determine other kinds of systematic alternations in split-ergative languages (Basque, Georgian, Hindi/Urdu, Dyirbal, and many others). The person split in (1)-(2) has been noted before in comparative studies by Mahajan (1994), Cocchi (1995), Nash (1997), but no cross-linguistic comparative study of all splits has been performed so far, mainly because most southern Italian dialects still await documentation. Thus, the immense wealth of these varieties for what concerns splits has not been exploited yet.
This project shifts the focus of investigation from the traditional *have-be* alternation to what triggers it, namely the splits in grammatical information. The final aim of this project is to unveil their function in the linguistic system and the mechanism that creates them.
It needs to be added that southern Italian dialects and split-ergative languages also exhibit clusters of splits, i.e. combinations of two or more splits at the same time. For instance, in Abruzzese, (1), the *have-be* alternation disappears in the subjunctive, so one needs to be 'in the right mood' and 'in the right person' to have *be*. Strikingly, the clusters of these splits seem to be exactly the same in the two macro-groups of languages. In other words, Italian dialects and ergative languages exhibit exactly the same splits and the same combinations of splits and only those (Cocchi 1995, Mahajan 1994). Again, this perfect correspondence cannot be accidental. This parallelism suggests that splits do not represent an exceptional phenomenon that accidentally emerged in one language group, as they are usually considered, but that they are the reflex of a more general language device, whereby some languages split and cluster grammatical information in a particular fashion.

## Overall aim and key objectives

This project aims at providing answers to one data question and four central questions:

***Data question:*** **Is the generalization on splits in the given languages complete or are there more splits than the ones we know about?**
**Question 1:** ***What is the function of splits in linguistic systems*** **?**
**Question 2:** ***Why these splits and not others?***
**Question 3:** ***Which splits can combine and which ones cannot?***

**Question 4:** ***Do alternations in case, auxiliation, and the rest, reflect underlyingly the same operation?***

Regarding question 4, it needs to be added that some potential triggers never determine any splits. For instance, given that we have a subordination-driven split, we could expect an interrogative-driven, or a negative-driven split. However, nothing like this is found: No language chooses the auxiliary depending on whether the sentence is negative or not, or it is interrogative or not. Why is that? This observation is crucial to understand the mechanisms of language.

The *starting hypothesis* is that splits have a function, possibly that of giving more salience to a piece of information with respect to the rest. This hypothesis is substantiated by studies on animacy hierarchies (Silverstein 1976 and ff.) that are often present in languages that exhibit splits. In these languages, splits are often found between the most 'animate' element and the rest.

Answering the four central questions will shed new light on the complexity of linguistic systems. It will help us understand how the grammatical information is organized and handled in different languages but also the underlying driving force or syntactic mechanisms that determine these splits. By understanding the function of these splits and the manner in which grammatical information is split and clustered, as well as the ways in which splits are instantiated, new light can be shed on a fundamental building block of linguistic theory.

## Methodology

In order to perform a thorough study of splits, it will be necessary to tackle the 4 central questions and the data question from three different perspectives: historical, micro-comparative and macro-comparative; each of these will be addressed by a dedicated subproject, to achieve different insights on the function and distribution of splits.

Splits are often addressed 'from within', i.e. not by looking at the whole set of possible splits in one language or more, but by looking at one instantiation at a time. Cross-linguistic comparison is rare and often selects, again, only one or two splits (see Mahajan 1994, Loporcaro 2007 and Manzini&Savoia 1999, 2005). At this point, it seems necessary to set up a comparative and integrated study of splits and their combination (or clusters), where these linguistic phenomena are observed cross-linguistically and from different perspectives. The methodology that this project will follow will be innovative in several respects. It will be:

**A. Comparative**
**B. Integrated**

A. The methodology used will be both micro- and macro- comparative. Comparative research will be pursued in three ways: by comparing all instances of splits in SIDs; by comparing the splits found in SIDs to those found in split-ergative languages; by comparing languages with splits with languages without splits.

B. This approach will also be integrated, in that each subproject will feed the others and every subpart will be integrated into the bigger picture. Each of the subprojects will need to adjust both its area of investigation and its working hypothesis according to the input provided by the others, thereby achieving a maximally complementary and integrated perspective.

### Subproject 1 (historical). The raise and fall of splits

Language change does not affect only one area of a language or only one item, but the change of one feature in a grammatical system triggers many other changes in the same system. Observing the

status of a language system while splits were emerging or disappearing is crucial to see what functions were being transferred to splits by other parts of the grammar or viceversa. In order to understand the role of splits, and how and why they combine, we need to isolate the moment in which they emerged or disappeared, the change they caused in the linguistic system, i.e. we need to study their history related to the system to which they belonged. By looking at the development of splits and their reflex in the language system, it will be possible to provide an answer to

***Question 1: What is the function of splits in linguistic systems*** **?**
The historical approach can also be very fruitful to analyze the reasons why some splits are never found. We can certainly expect some language changes to give rise to splits. But are these 'expected' splits actually found? By analyzing the development of splits this subproject will be able to provide a clue for answering

**Question 2:** ***Why these splits and not others?***
A very interesting example of historical research on splits is provided by the study of Ledgeway (2003) on *Libro de la destructione de Troya*, a Neapolitan work from the XIV c. In one sentence, repeated in (3) the verb *bastare* 'to suffice' selects *have* in an *irrealis*/conditional context and *be* in a factive context. This has led Ledgeway to conclude that HAVE emerges first in *irrealis* contexts, and not, like Formentin (2001) maintains, in perfective contexts.
(3) E se eo non avesse avuta in me questa potestate **averria*me***
and if I not had-subj had-past part in me this power it-would=to-me

ben potuto bastare, commo **èy** bastato ad onnuno de quist'altri signuri
well could suffice like it-is sufficed-past part to each of these=other men

'And if I didn't have this power in me, it would have sufficed me, like it has to each of these other men'

(14th c. Neap, *Libro de la destructione de Troya* 201.35-6, in Ledgeway 2003:6)

Old Neapolitan is relatively well documented but most documents in old Italian dialects still await documentation. A historical study of these dialects will also help us provide an answer to several subquestions:

- In which context does one or the other auxiliary appear?
- What were the conditions of the language before and after the appearance/disappearance?
- Did the system determine the split or did the split emerge for some specific reason causing a readjustment in the system? What kind of readjustment?

This PhD project will undoubtedly offer contribution to the data question, in that it will document older stages of southern Italian dialects, most of which are not well documented. Despite the potential difficulty in finding suitable material, this subproject will be able to make use of selected material from the following archives in southern Italy: L'Aquila and Chieti (for Abruzzese), Campobasso (for Molisan), Salerno (for Campanian), Foggia (for upper Apulian), Bari (for Apulian), Potenza (for Lucanian), Reggio Calabria and Cosenza (for Calabrian), and Catania (for Sicilian). Sicilian no longer shows a *have-be* alternation but used to show it in the past (Loporcaro 2007). As such, it can be an object of investigation for a historical project. The varieties of Naples, Lecce, and Palermo,

already well documented, will be left out from our search and kept as a background. For reasons of time, minor varieties will not be addressed but occasionally. The retrieval of old southern Italian texts will constitute an added value to this project, which will be available not only to linguists for future research, but also for historians and humanists in general.

## Subproject 2 (microcomparative). The syntax of splits in Southern Italian varieties

The micro-comparative subproject will deal with the comparison of linguistic systems in which splits are present. In particular, it will compare modern southern Italian dialects (direct descendants from the old dialects) with each other.

This study will approach southern Italian systems in a 'synchronic' fashion, not considering the evolution of splits but their place within the rest of the linguistic system. A substantial part of this project will be devoted to the documentation of contemporary dialects, for many of which we lack data. This subproject will focus on the following questions:

**Question 1:** ***What is the function of splits in linguistic systems*** **?**
**Question 3:** ***Which splits can combine and which cannot?***
Within this sub-project, all split phenomena attested in southern Italian dialects will be studied and classified. Then, splits will be compared with each other and, occasionally, with old southern Italian vernaculars documented by subproject 1.

Micro-comparison is of crucial importance in linguistic research. To put it with Kayne (1996), micro-comparative syntax "is a powerful tool, whose growth is perhaps to be compared with the development of microscopes, that allows us to probe questions concerning the most primitive units of syntactic variation" (see also Longobardi 2003). As an example, consider what follows: In recent papers, D'Alessandro&Roberts (forth.) and D'Alessandro&Ledgeway (forth.) attribute the person-driven split in Abruzzese to the presence of a person feature on the verbal functional head (*v*). This implies that some other phenomena take place, such as the agreement of the past participle with the subject of transitive verbs, which is unknown in Romance. Interestingly, D'Alessandro&Roberts find that Abruzzese does show past participle agreement of this kind, as expected. Thus, a tiny difference in the location of a person feature has produced a visible syntactic reflex in participial agreement. The documentation of modern southern Italian dialects will bring to light new data, previously unknown, which will offer an important contribution to the solution of the problems at issue. Modern southern dialects are spoken as native varieties by fewer and fewer people, and the majority of native speakers are over 60. This means that these varieties are classifiable as endangered, and that the output of this project also has the added value of contributing to the documentation and study of endangered varieties. Similar to subproject 1, this subproject will address a selection of dialects. These dialects will however not be city dialects, heavily suffering from the superstratal influence of Italian, but countryside dialects, optimally one for each province of the southern Italian regions.

## Subproject 3 (macrocomparative). The syntax of splits and clusters

The structural subproject will be concerned with syntactic macro-variation and the configuration of the grammatical material on functional heads.

Several studies have addressed the nature of split-ergativity by looking at it from the perspective of ergative languages (Dixon 1972, 1994, Silverstein 1976, Marantz 1984, Müller 2004 among others). Many explanations have been proposed for the ergative patterns itself, which clusters verbal

arguments in a different fashion than the Romance and Germanic languages. However, we have seen that the same splits arise in a group of languages (southern Italian dialects) which do not express morphological case. In these varieties, splits are mainly reflected in the *have-be* alternation, whereas in ergative languages there are more ways to reflect them. If we take a step back and look at the two macro-groups, it seems that parallel mechanisms are at work. With this assumption, this subproject will tackle the

**Question 4:** ***Do alternations in case, auxiliation, and the rest, reflect underlyingly the same operation?***
This project will provide a new vision on studies on split-ergativity and will at the same time profit from insights coming from ergative languages, better documented than southern Italian dialects.

We have seen that there are splits in person, Tense/Aspect/Mood, finiteness, subordination. These elements are represented, in syntactic studies, as functional categories and their projections (see Muysken 2008 for a recent cross-linguistic study on functional categories, and Longobardi's studies on global parametrization). The exact nature/model of functional categories is not uncontroversial.In recent years there has been an ongoing debate on whether each piece of information should be assigned to a dedicated head (mainly following the insightful works of Cinque (1999) and Rizzi's (1997) split-CP hypothesis), or whether it is necessary to minimize the structural apparatus in a clause and hence to minimize the number of functional projections (Minimalist Program, Chomsky 1993 and ff.). These splits seem to provide a clear indicator of how information is divided and organized on functional heads. If an auxiliary split is triggered by person, for instance, this will mean that 1st and 2nd person on one hand and 3rd person on the other will be 'separated' units, and possibly separated functional heads. This study will help us build a more informed syntax, by answering the question:

- *What is the exact content of the functional heads that are involved in splits?*

Even more interestingly, if for instance person interacts with mood to determine auxiliary in Abruzzese, this will mean that the head hosting the person feature and the head hosting mood interact with each other. The question is then:

- *How does this happen*? *How do these pieces of grammatical information interact? And to what extent?*

It is also possible to hypothesize that some splits and clusters are not possible because of structural constraints. This subproject will try to identify the structural constraints that facilitate/forbid some splits and clusters and not others, both in Romance and ergative languages. This methodology and the assumption that functional heads reflect splits will also allow us to tackle

**Question 2:** ***Why these splits and not others?***
From the perspective of syntactic theory this question can be rephrased as

- *which material can combine on functional heads and which cannot*?

It goes without saying that this study will have to include those languages that do not present splits and also those languages, like Sicilian or Old Spanish, which used to exhibit splits at some stage but lost them. Addressing the problem from this macro-comparative perspective which also includes languages that do not exhibit splits will also offer possible answers to

**Question 1:** ***What is the function of splits in linguistic systems*** **?**

By comparing all possible instances of splits in southern Italian dialects and in ergative languages we will be able to see how the grammatical information is handled and organized on functional heads. This study will thus offer an important contribution to syntactic theory, in that it provides a clearer understanding of the nature and shape of functional heads on the basis of straightforward data.