Languages in Contrast aims to publish contrastive studies of two or more languages. Any aspect of language may be covered, including vocabulary, phonology, morphology, syntax, semantics, pragmatics, text and discourse, stylistics, sociolinguistics and psycholinguistics.

Languages in Contrast welcomes interdisciplinary studies, particularly those that make links between contrastive linguistics and translation, lexicography, computational linguistics, language teaching, literary and linguistic computing, literary studies and cultural studies.

Languages in Contrast provides a home for contrastive linguistics. It will enable advocates of different theoretical linguistic frameworks to publish in a single publication to the benefit of all involved in contrastive research.

Languages in Contrast provides a forum to explore the theoretical status of the field; stimulates research into a wide range of languages; and helps to give the field of contrastive linguistics a distinct identity.

Language in Contrast is covered in the following indexing and abstracting services: IBR/IBZ, Linguistic Bibliography/Bibliographie Linguistique, LLBA, European Reference Index for the Humanities and Translation Studies Abstracts Online.

Guidelines for Contributors

LiC invites contributions relevant to the aim and scope of the journal. Submission instructions and guidelines can be found on the journal's website: www.benjamins.com/catalog/lic

All correspondence concerning editorial matters – inquiries regarding suitable topics, length of manuscripts, reviews, etc. – should be sent directly to the editor:

Silvia Bernardini
SITLeC – University of Bologna
Corso Diaz 64
47121 Forlì, Italy
e-mail: lic@sslmit.unibo.it
Tel.: +39 0543 374736 Fax: +39 0543 374717

Table of contents

Geschichte und Gegenwart. Betreut von Steffen Krogh, Simon Neuberg und Gilles Rozier (Jahrbuch für Internationale Germanistik, Reihe A 78); 37–43. Bern u.a.: Peter Lang.

Gabelentz, H. C. von der. 1846. *Grammatik der gotischen Sprache*. Leipzig: F.A. Brockhaus.

Gelderen, E. van. 2004. *Grammaticalization as Economy*. Amsterdam: John Benjamins.

Harbert, W. 1992. "Gothic Relative Clauses and Syntactic Theory". *On Germanic Linguistics: Issues and Methods*. I. Rauch et al. (eds), 109–146. Berlin: Mouton de Gruyter.

Harbert, W. 2007. *The Germanic Languages*. Cambridge: Cambridge University Press.

Helgander, J. 1971. *The Relative Clause in English and Other Germanic Languages: a Historical and Analytical Survey*. PhD. Dissertation, University of Gothenburg.

Klinghardt, H. 1877. "Die Syntax der gotischen Partikel ei". *Zeitschrift für deutsche Philologie* 8:142–180, 289–329.

Koopman, H. 1996. "The Spec Head Configuration". Reprinted in *The Syntax of Specifiers and Heads: Collected Essays of Hilda Koopman* (2000). London: Routledge.

Lockwood, W. 1995. *Lehrbuch der modernen jiddischen Sprache*. Hamburg: Helmut Buske.

Lowenstamm, J. 1977. "Relative Clauses in Yiddish: A Case for Movement". *Linguistic Analysis* 4:197–216.

Lumtzer, V. 1896. "Die Leibitzer Mundart, II". *Beiträge zur Geschichte der deutschen Sprache und Literatur* 21:499–539.

Lundskær-Nielsen, T. and P. Holmes. 2010. *Danish: A Comprehensive Grammar*. 2nd edition. London: Routledge.

Mráz, G. 1909. *A Dobsinai német nyelvjárás*. Budapest: Magyar Tudományos. Akadémia.

Ponelis, F. 1993. *The Development of Afrikaans*. Frankfurt am Main: Peter Lang.

Rizzi, L. 1997. "The Fine Structure of the Left Periphery". *Elements of Grammar*. L. Haegeman (ed), 281–337. Dordrecht: Kluwer Publications.

Seppänen, A. and G. Kjellmer. 1995. "The Dog That's Leg Was Run Over: On the Genitive of the Relative Pronoun". *English Studies: A Journal of English Language and Literature* 76:389–400.

Soden, H. Freiherr von. 1913. *Griechisches Neues Testament: Text mit kurzem Apparat*. Göttingen: Vandenhoeck & Ruprecht.

Streitberg, W. 2000. *Die gotische Bibel*. 7th edition. Heidelberg: Universitätsverlag.

Streitberg, W. 1920. *Gothisches Elementarbuch*. Heidelberg: Carl Winter.

van der Auwera, J. 1985. "Relative 'that' — A Centennial Dispute". *Journal of Linguistics* 21:149–179.

Vries, M. de. 2002. *The Syntax of Relativization*. Utrecht: LOT.

Author's address

Wayne Harbert
Department of Linguistics
Cornell University
Ithaca NY 14853
USA

weh2@cornell.edu

Contrastive linguistics and micro-variation

The role of dialectology

Cecilia Poletto
Goethe University, Frankfurt am Main (Germany)

This article deals with a very general problem, namely the origin of the well-known distinction between dialectal and typological variation. It is argued that the fact that the possible grammatical choices are more restricted within a dialectal domain is not due to a supposed principled difference in the parameters that rule variation. Rather, they are a function of the originally unitary lexicon dialects share. If language variation is essentially located in the functional items, and they are derived from the same lexicon, then they will share some core properties that make dialectal variation so restricted. I propose that the fact that the lexicon is similar can give us clues about the internal structure of syntactically complex elements which are represented by a single word, like quantifiers, *wh*-items, modal verbs, etc. Within a homogenous domain, structural complexity correlates with a higher number of lexical roots: the higher the number of the lexical roots found, the more complex internal structure the functional item will display.

1. Introduction

The problem I intend to address in this work has to do with the claimed independence of dialectology with respect to contrastive linguistic research in general. Since the first attempts to use dialectal data as an empirical basis for theoretical syntactic work[1] there has been a whole current of studies which assume that micro-comparative research cannot be captured on the basis of the old Government and Binding standard notion of parameter, but that some other mechanism must be involved and be held responsible for the peculiar more subtle type of variation found when investigating dialects (see among others Poletto, 2000; Manzini and Savoia, 2011). This assumption entails a number of consequences that have not, to my knowledge, been spelled out and that I will discuss here. The first is that

dialectological work should in principle be qualitatively different from contrastive work conducted on the basis of genetically totally unrelated languages or even between languages which belong to the same family but are not so strictly related. But is this really so? Does dialectal variation really differ from typological variation in a qualitative way? And if so, what is the reason for the difference(s) we observe? I will try to answer this question with respect to syntactic phenomena only. It might turn out that for other fields of grammar the answer is different, but for the moment I limit the empirical domain of research to syntactic data.

A second consequence of the above-mentioned assumption is that dialectology should be entitled to have its own methods of investigation, which aim at highlighting the more fine-grained type of variation found among related dialects. Within the research community working on syntactic facts, this has brought about the conception of several enterprises which only partly resemble the old basic tool of a linguistic atlas, as they are mainly in the form of data bases freely available on the web (see among others the ASIt project based in Padua,[2] the SAND project at the Meerstens Instituut in Amsterdam and the Scandiasyn project based in Tromsoe). The issue of how to gather, classify and retrieve a considerable amount of syntactic data coming from non-linguist informants has become an explicit research question, which has led to acknowledging that a standard has to be set as to the empirical methodology (see Cornips and Poletto (2004) for an explicit discussion of field methods), ranging from the choice of informants to the type of tests used for different stages of the field work and to the tagging of the data, including also the possibility of a general European platform for questioning several of these data bases at the same time and with the same interface.

Attention to the methods of empirical research is a relatively new issue in the generative domain, which has often contented itself with an empirical basis drawn directly from the competence of the author and maybe some other linguists. The reason for this is that dialectological research has as one of its strong points the fact that we can build more solid empirical generalizations basing our observation on a wide set of closely related languages. For instance, Kayne (1996) explicitly claims that investigating dialects is the kind of empirical research that most closely resembles a scientific experiment, where the majority of the variables is controlled and only one factor varies. Therefore, we can state that dialectological research is at the moment developing a new standard in the gathering and treatment of data, which only partly resembles the standard created by closely related fields of investigation as language acquisition or psycho-linguistics. However, this does not yet mean that micro-comparative research is in itself a different discipline or that micro-variation is qualitatively different from macro-variation.

It is important to stress from the very beginning that the peculiar type of linguistic variation found across dialects discussed here, which I dub 'leopard spots',

borrowing the terminology from traditional etymological studies on Italian dialects, cannot possibly be explained away as facts pertaining to language contact, migration or mobility between dialects, as it is highly improbable that, to make an example, the dialect of Semogo, a village in Northern Lombardy, has had any influence on the Emilian dialect of Ferrara, as the two varieties are not in contact, speakers use the standard Italian variety when they are in a different region, and there is no relevant immigration between the two communities. The same argument can be replicated for the vast majority of the dialects presented here, and recent work by Christine Lamarre on entirely different dialects, as the Mandarin varieties in Northern China, has shown exactly the same leopard spots distribution found in Italian dialects. The argument presented here is precisely that the same type of syntactic developments is found among dialects that are not close to each other and for functional elements, which, as is well known, are rather difficult to borrow, contrary to lexical words.

Notice furthermore that the reasoning presented here is valid even in case of sociolinguistic variation within a community, as it is based on the sheer existence of the relevant forms for functional items in a given dialect, not on the usage, interpretation or distribution.

In what follows I will first show (Section 2) that at first sight there seems to be no real qualitative difference between these two types of variation, but that there is actually only a quantitative difference, though a massive one, in the way variation spreads in what looks like a more homogenous set of languages.

In Section 3, I will provide a possible answer to the type of distinctions which are actually found between micro and macro-comparative syntax and propose that the responsible factor for the clustering of similar dialects in one spectrum of the possible variation range is the (functional) lexicon, which has similar properties in similar dialects.

In Section 4, I will discuss some implications of this apparently trivial observation and take into account what this could mean for the general theory of language variation.[3]

The empirical field which I will draw examples from are the Romance dialects spoken on the Italian territory and in the Italian-speaking part of Switzerland (including varieties that have the status of minority languages), and will leave the extension of these observations to other language groups for future research.

2. Micro-comparative and macro-comparative variation compared

The linguistic 'distance' among dialects has been object of investigation since dialectology exists, a whole discipline — dialectometry[4] — aims at determining how

dialectological work should in principle be qualitatively different from contrastive work conducted on the basis of genetically totally unrelated languages or even between languages which belong to the same family but are not so strictly related. But is this really so? Does dialectal variation really differ from typological variation in a qualitative way? And if so, what is the reason for the difference(s) we observe? I will try to answer this question with respect to syntactic phenomena only. It might turn out that for other fields of grammar the answer is different, but for the moment I limit the empirical domain of research to syntactic data.

A second consequence of the above-mentioned assumption is that dialectology should be entitled to have its own methods of investigation, which aim at highlighting the more fine-grained type of variation found among related dialects. Within the research community working on syntactic facts, this has brought about the conception of several enterprises which only partly resemble the old basic tool of a linguistic atlas, as they are mainly in the form of data bases freely available on the web (see among others the ASIt project based in Padua,[2] the SAND project at the Meerstens Instituut in Amsterdam and the Scandiasyn project based in Tromsoe). The issue of how to gather, classify and retrieve a considerable amount of syntactic data coming from non-linguist informants has become an explicit research question, which has led to acknowledging that a standard has to be set as to the empirical methodology (see Cornips and Poletto (2004) for an explicit discussion of field methods), ranging from the choice of informants to the type of tests used for different stages of the field work and to the tagging of the data, including also the possibility of a general European platform for questioning several of these data bases at the same time and with the same interface.

Attention to the methods of empirical research is a relatively new issue in the generative domain, which has often contented itself with an empirical basis drawn directly from the competence of the author and maybe some other linguists. The reason for this is that dialectological research has as one of its strong points the fact that we can build more solid empirical generalizations basing our observation on a wide set of closely related languages. For instance, Kayne (1996) explicitly claims that investigating dialects is the kind of empirical research that most closely resembles a scientific experiment, where the majority of the variables is controlled and only one factor varies. Therefore, we can state that dialectological research is at the moment developing a new standard in the gathering and treatment of data, which only partly resembles the standard created by closely related fields of investigation as language acquisition or psycho-linguistics. However, this does not yet mean that micro-comparative research is in itself a different discipline or that micro-variation is qualitatively different from macro-variation.

It is important to stress from the very beginning that the peculiar type of linguistic variation found across dialects discussed here, which I dub 'leopard spots',

borrowing the terminology from traditional etymological studies on Italian dialects, cannot possibly be explained away as facts pertaining to language contact, migration or mobility between dialects, as it is highly improbable that, to make an example, the dialect of Semogo, a village in Northern Lombardy, has had any influence on the Emilian dialect of Ferrara, as the two varieties are not in contact, speakers use the standard Italian variety when they are in a different region, and there is no relevant immigration between the two communities. The same argument can be replicated for the vast majority of the dialects presented here, and recent work by Christine Lamarre on entirely different dialects, as the Mandarin varieties in Northern China, has shown exactly the same leopard spots distribution found in Italian dialects. The argument presented here is precisely that the same type of syntactic developments is found among dialects that are not close to each other and for functional elements, which, as is well known, are rather difficult to borrow, contrary to lexical words.

Notice furthermore that the reasoning presented here is valid even in case of sociolinguistic variation within a community, as it is based on the sheer existence of the relevant forms for functional items in a given dialect, not on the usage, interpretation or distribution.

In what follows I will first show (Section 2) that at first sight there seems to be no real qualitative difference between these two types of variation, but that there is actually only a quantitative difference, though a massive one, in the way variation spreads in what looks like a more homogenous set of languages.

In Section 3, I will provide a possible answer to the type of distinctions which are actually found between micro and macro-comparative syntax and propose that the responsible factor for the clustering of similar dialects in one spectrum of the possible variation range is the (functional) lexicon, which has similar properties in similar dialects.

In Section 4, I will discuss some implications of this apparently trivial observation and take into account what this could mean for the general theory of language variation.[3]

The empirical field which I will draw examples from are the Romance dialects spoken on the Italian territory and in the Italian-speaking part of Switzerland (including varieties that have the status of minority languages), and will leave the extension of these observations to other language groups for future research.

2. Micro-comparative and macro-comparative variation compared

The linguistic 'distance' among dialects has been object of investigation since dialectology exists, a whole discipline — dialectometry[4] — aims at determining how

distant two linguistic systems are in terms of objective linguistic measures drawn from a set of lexical, phonological and morphological features. A well-known example of this approach is Pellegrini's (1977) map of Italian dialects, where the distinctions among dialects are encoded in terms of colours and are based on an arbitrary set of 30 measures including lexical, phonological and morphological features. One might ask what the relevance is with respect to the proximity of the internal languages that have these dialects as output, are they really more similar than, say Italian and French, or Italian and Chinese? At a first naïve look, everyone would say that they are, but one interesting observation that comes from comparing measures concerning the linguistic system and the lexicon is that there is an astonishing convergence between the two components: dialects that are more similar in terms of lexicon are also more similar in their phonological and morphological system.[5] This convergence must be far from accidental.

2.1 Subject agreement

Although any dialectologist (as well as any typologist) has the impression that you can never find the same degree of variation in dialectology as found in typology, proving or disproving this first impressionistic belief is actually extremely difficult.

For instance, Van Koppen (2011) has shown that if we concentrate on just one syntactic phenomenon and investigate the range of variation that can be found across Dutch dialects, this turns out to be similar to the one we can find across a typological research: the phenomenon she discusses is the expression of the subject.[6] She notes that across typological work we can find the same four types of subject realizations that we find looking at different Dutch dialects:[7]

a. no subject pronoun necessary (pro drop languages like standard Italian)
b. a full pronoun or a NP necessary (non pro drop languages like English or standard Dutch)
c. a reduced pronoun necessary ("intermediate cases" like Dutch dialects)
d. doubling: when a NP is present a (reduced) pronoun also is (like a subset of Dutch dialects)

She concludes that the range of variation is in principle not different across dialects from what is found in typological research.

The same experiment can be replicated for Italian dialects: taking the possible variation range concerning the distribution of the subject, we find exactly the four possibilities provided above. Southern Italian dialects are generally pro-drop, as no subject is necessary, as shown in (1).

(1) *Vole cu bbene*   (Salentino)
Wants that comes
"He wants to come"

Friulian dialects have a system where six pronouns out of six are obligatory:

(2) a. *I mangi S.*   Michele al T. (Friulian)
deict. SCL eat (I)
"I eat"

b. *I ti mangis*
deict. SCL SCL eat (you)
"You eat"

c. *A l mangia*
deict. SCL SCL eat (he)
"He eats"

d. *I mangin*
deict. SCL eat (we)
"We eat"

e. *I mangè*
deict. SCL eat (you)
"You eat"

f. *A mangin*
deict. SCL eat (they)
"They eat"

The third case is exemplified by several Northern Italian dialects, where subject clitics are necessary only for some persons or for all depending on the dialect:

(3) a. **(El) vien*   (Venetian)
(he) comes

b. *Vegno*
(I) come

The last type is found in other Northern Italian dialects where a subject clitic is obligatory even when a DP subject is already present, a phenomenon which has been dubbed 'redetermination of subject agreement' by traditional dialectologists:

(4) *La Maria *(la) vien*   (Trentino)
The Mary (she) comes

Several other phenomena display the same wide spectrum of variation, which shows that there is no principled reason why dialectal variation should be different from typological variation. If one looks close enough, it is possible to find syntactic phenomena that are typical of typological variation, though they are quantitatively very rare.

In what follows, I will first make the same point as van Koppen (2011) and consider some of the typical claims made about variation within the Romance domain and show that the space of *a priori* possible syntactic variability found across dialects is actually the same that can be found across distant languages, though some 'types' are extremely rare.[8] Again, I make no claims on phonology, hence types of variation like tone marking of the subject are not taken into account. In the following, I will also try to go a bit further and find a reason for the rarity of some structures in micro-variation.

2.2 OV sentences

The first feature to be discussed is one of the typical syntactic properties that are always cited when talking about the property of 'stability', i.e. the fact that no Romance variety is nowadays an OV language, though Latin was (see Renzi, 1986 and Oniga, 2004 on this). However, if we consider the possibility of OV orders in modern Romance, we can find several examples of OV even leaving aside the OV orders obtained through clitic or quantifier movement already discussed in Kayne (1975). One of these cases is Friulian, which tolerates some limited cases of OV with DP objects (and crucially obligatory past participle agreement with the object).[9]

(5) a. *O ai lis sigaretis desmenteadis*
I have the cigarettes+FEM.PL forgotten+FEM.PL
"I have forgotten the cigarettes"

b. *O ai desmenteadis lis sigaretis*
I have forgotten+FEM.PL the cigarettes+FEM.PL
"I have forgotten the cigarettes"

c. *O ai desmenteadis/desmenteat lis sigaretis*
I have forgotten-AGR the cigarettes+FEM.PL[10]
"I have forgotten the cigarettes"

d. **O ai lis sigaretis desmenteat*
I have the cigarettes+FEM.PL forgotten-AGR
"I have forgotten the cigarettes"

The fact that informants find (5d) impossible attests that the construction is by no means 'frozen' (whatever this might mean), but that it is part of the linguistic system and integrated into the agreement patterns of the past participle. Therefore, although it seems generally true that no Romance dialect is an OV language in the sense languages like Japanese (or even German) are, OV orders are possible with regular DP objects, not only clitics or quantifiers.

2.3 Agglutinative morphology

Another typical claim about dialects is that they have more or less the same type of morphology: all Romance dialects do not have agglutinative morphology, but display morphology which belongs to the flexive type. For instance, no Italian dialect has the morphological makeup of a language like, say, Turkish.

(6) *oku- ya- ma- yabil- ir- im*
read Abil NEG Abil Aor 1SG
"I might be unable to read; It is possible that I shall be unable to read"

Turkish displays two clear modal morphemes, while Italian dialects generally never display a clearly separable morpheme which only encodes modality. The inflectional morpheme encodes modality, tense and (in some cases) person agreement fused in a single non-analyzable element. Nevertheless, even in this case there are some sporadic cases of what actually looks like agglutinative forms adjoined to the right of the 'usual' inflectional morphology (Benincà, 1996), and has indeed been analyzed as such: Benincà shows that a mood morpheme for subjunctive etymologically derived from the adverb *bene* 'well' is used in several Northern Lombard and Veneto dialects. What looks striking is that the modal morpheme is agglutinated after the subject agreement marker, a very unusual case in Romance, where the subject agreement marker generally 'closes up' the word.

(7) a. *Son* (I am PRES. ind.)
b. *Son-be* (I be PRES. Subjunctive)
c. *Sie-be* (you be PRES. Subjunctive)

Again, the general claim that Italian (and Romance) dialects do not possess agglutinative morphology is falsified by what are clearly minority choices in the Romance area, but still exist as a possibility. This means that the processes that lead to an OV grammar or to an agglutinative language are not blocked *per se*, they are simply 'rarer', i.e. the problem of dialectal variation looks more like a quantitative problem. A modern theory of grammar as an 'optimal' construct does not leave space for percentages or quantitative phenomena, although we cannot ignore facts about the distribution of 'rare' phenomena. On the one hand, we have to express the fact that in principle the range of variation of closely related dialects is similar to the one of distant languages; on the other hand, we have to express the fact that all related dialects 'tend to converge' towards similar grammars. One further astonishing property of dialectal variation is that even diachronic development is similar within a dialectal area (as we will see below) and proceeds along similar lines for a long time even when dialects are subject to the influence of different standard languages.

Summing up, we are faced with three problems: the first one is that micro-variation, being so fine-grained, seems to fall outside the pattern of usual research and cannot be captured by the standard tools we use in accounting for language diversity in general. The second one has to do with the way variation is distributed across dialects: although they tend to cluster towards very similar grammars, there are still cases that show that there is no a priori principled reason that excluded the same type of variants found in typological variation. The third is that there seems to be a sort of very long 'drift' as similar dialects tend to develop the same type of properties, which are often not shared by the standard language, so that they cannot be attributed to language contact (at least not with the standard).

In a sense, we are facing a paradox, as what is surprising in dialectology is not what changes, but what remains constant, though the range of 'possible', even if rare, variation across dialects is virtually identical to the one found across languages.

3. On the peculiarity of micro-variation

3.1 Theories on variation

Although in the recent minimalist framework the 'theoretical space' is severely limited by the assumption that all languages are constrained by the same requisite to be optimal with respect to the interfaces, there are at least two opposite views that have been suggested to provide a theoretical status to the phenomenon of linguistic variation: the first one has been originally proposed by Mark Baker in his 1997 book, where he tries to derive 'big' typological distinctions (like incorporating languages) from macro-parameters and leaving 'smaller' distinctions to micro-parameters, whose value only becomes relevant if a particular setting of the overarching macro-parameter has been chosen. More recently Longobardi's work has also taken this direction, as he develops a theory where certain parameter settings depend on others, arguing for the existence of clusters of parameters.

The opposite view is represented by Kayne's work (see, among others, Kayne, 1996, 2004), and work done by many other linguists, like Rizzi and Cinque and the so-called 'cartographic school'. They assume that macro-parameters and micro-parameters are qualitatively alike, as they descend from settings related to the lexical items that realize functional projections, only the quantity of different settings is dramatically different, because the number of functional projections is not limited to the standard CP-TP vP-VP layering of the minimalist program.[11]

The facts that we have discussed above point towards the second hypothesis: it seems that the potential range of variation found across dialects is the same found in typological work, so there are no 'macro-parameters' that are set for all dialects,

although the majority of dialects are indeed very close from the point of view of the parametric choices they make, though not always.

We are thus left with the uncomfortable problem of having to embed an (imprecise) statistical concept, namely 'rarity' into our theory of language variation. I will propose that the reason why the majority of historically related dialects have very similar grammars depends on the fact that they generally have a very similar lexicon: if language variation is determined by a change in the properties of lexical items expressing functional projections, then dialects that have a similar lexicon will also have similar morpho-syntactic properties. Thus, the answer to our problem is rather trivial; however, its consequences are not. Given that Italian dialects have inherited the majority of their lexicon from Latin, they clearly share a common core, but this common core must have remained pretty stable across two thousand years. The problem does not seem to simply involve stability: if we look at functional words across time, i.e. to grammaticalization processes, how come several dialects have evolved in exactly the same direction? As the theory of grammaticalization looks precisely at what changes but also at what remains constant throughout time when the process applies, it is just what we need in order to understand the phenomenon of dialectal variation.

I will adopt here again the term 'leopard spots' typical of traditional Romance dialectological work to define the peculiar type of variation found among dialects. What is meant is that the distribution of a given feature (be it phonological, morphological or lexical) is scattered in a non-homogeneous way across dialects, i.e. it does not extend over a whole geographical area, but appears here and there in different spots, although the neighbouring dialects do not display it. It is very interesting to observe how variation distributes across a relatively homogeneous domain (as a group of related dialects clearly is), since the distributional pattern can tell us a lot about the way a given change proceeds and the features it is sensitive to. As already noted in the introduction, the geographical and sociolinguistic situation of Northern Italy is not such that we can attribute the data I present below to migration or to language contact.

3.2 Leopard spots and interrogative clauses

In what follows I will exemplify how variation distributes across neighbouring dialects looking at the different structures we can find in main interrogative sentences: we know from written sources that all medieval Northern Italian dialects consistently had subject inversion and that subject clitic inversion has been preserved in main interrogative sentences in a way that is similar to standard French, and that is still the regular way to build main interrogatives in several dialects. I will use as an example the translation of a sentence like "What is he doing?":[12]

(8) *Cossa fa-lo?*   (Paduan)
   What does he?

This possibility is still attested only in some but not all dialects. Many of them have developed alternative strategies to subject clitic inversion, one of which is cleft clauses:

(9) *Dove   ze che el va?*   (Venetian)
   Where is that he goes?
   "Where is he going?"

The second strategy found across the domain is the one that realizes a complementizer after the *wh*-item not only in embedded but also in main interrogatives:

(10) *Cossa che el fa?*   (Portogruaro)
    What that he does?

Other dialects display an in-situ strategy (which is in the majority of cases only available in main but not in embedded interrogatives):

(11) *Fa-lo   che?*   (Bellunese)
    Does he what?

Looking at a map, we see that all these dialects belong to the same Veneto area, but if we move, say from Veneto to Piedmont, we see that the type of variation is essentially the same: so we do not find for example an Eastern area (Veneto) which has developed an interrogative strategy different from the Western area (Piedmont) but the same type of scattered variation which ranges over at least four different constructions is found in both East and West. This is notably the same type of variation found in non-standard French varieties (see among others Elsig and Poplack, 2006 for an overview of the distribution of the construction in French).

On the other hand, it is surprising to see how uniform the direction of the change is: dialects of the whole area are abandoning the 'traditional' strategy of subject clitic inversion although they choose different ways for doing so. Notice that this cannot be attributed to the influence of standard Italian, as standard Italian does not display any of these constructions. It is rather the opposite, as Lombard speakers, who use cleft constructions very frequently, also use it in standard Italian in contexts where it is not allowed or necessary in standard Italian. The change cannot be attributed to influence of French, as French does not have any in the Italian domain. Therefore, we have to state that the diachronic drift is autonomous, and manifests itself uniformly in all Northern Italian dialects in the loss of subject clitic inversion and non-uniformly in the way the old strategy is being substituted. Leopard spots variation describes exactly this non-uniform substitution.

This long drift in the change process shows that dialects do not simply follow the pressure of the standard language when they change. Moreover, what is astonishing here is the clear correspondence between Northern Italian dialects and French, which cannot be due to any recent influence, but evidently to very long-lasting properties which made these two domains different from the rest of the Romance world already in the middle ages (see Vanelli *et al.*, 1985 on this). One might ask how properties determining the change can last so long, and in the view of the brief discussion about the role of the lexicon in language variation, we could propose that the correspondence we see here is due to the properties of the functional lexicon: lexical items that become grammaticalized maintain at least some of their original properties, and this drives the diachronic cline as much as the general conditions on (semantic, phonological and syntactic) impoverishment, as formulated by the standard grammaticalization theory and that this persistence is responsible for long-lasting drifts like the one discussed here. Therefore, what we should investigate in dialectal variation is the pertinacity of very basic properties of elements which become functional, not only the features each lexical item progressively loses during the process. The reason why all Northern Italian dialects are losing subject clitic inversion and are developing the three strategies noted above on a par with French is ultimately that the original properties of functional elements (like the complementizer, the copula of clefts and interrogative pronouns) restrict the possible output structures a given element can enter. This is the reason why dialects are all so similar and why we find the so-called 'leopard spots' distribution, where the same set of constructions is distributed in an uneven way over a whole territory.

4. Clustering variation

A second property which emerges very clearly from dialectal variation is that it clusters around some grammatical areas or lexical items but not others. In what follows I will provide three examples of clusters concerning functional items: modal verbs, quantifiers and *wh*-words. I will propose that the reason why this is so has to do with the level of internal syntactic complexity a given item displays.

4.1 Modal verbs

Among modal verbs we see an interesting split among the modal verb representing 'will, want' and the modal verb representing 'need' in the sense of a pure necessity. In what follows I will present examples of the two verbs showing that the modal verb of volition is rather stable and does not vary much in the overall

majority of dialects, while the necessity verb uses a varieties of lexical forms which make several different constructions possible. Virtually all Italian dialects investigated in the ASIt project (except for Sardinian) display the same word meaning 'will/want', which is derived from Latin *volo* and has adapted to the phonological changes that have happened in the different dialects, though the etymological root has remained the same.[13]

The ASIt data base represents an essential tool to check this fact: in what follows I simply provide a list of dialects from different areas of the Northern part of Italy, which all display the same etymological root. I take here the third person singular of the present indicative as an example:

> The form *vol* is found in the following dialects, among others: Albosaggia (Ligurian), Aldeno (Trentino), Bologna, Borgofranco d'Ivrea (Piedmont), Casalmaggiore (Lombard), Farra di Soligo (Veneto).
>
> The form *vo* is found in Alassio, Calizzano (Ligurian), Comano (Northern Tuscan).
>
> The form *vole* in the whole central Veneto area and Florence.
>
> The form *vor* is found in the Piedmont and Lombard areas (for instance in the dialects of Lecco (Lombard), Livorno Ferraris (Piedmont)).

Other dialects present forms that have been eroded in different ways, like *ul* (Remanzacco, Friulian) or *o* found in the Liguria dialects of Favale di Malvaro, Arzeno.

The list could be much longer; however, in all cases the form can be reduced to the same Latin etymology. If we take into account the verb indicating necessity, we observe a variation of different verbs that are used in order to express this meanings (see Benincà and Poletto, 1994 for a syntactic analysis of the impersonal form *bisogna* 'need' and its cognates, which shows that this verb is indeed different from other modal verbs like *volere* 'want/will', *potere* 'can' and *dovere* 'must').

I provide here some examples of the various etymological types: the first type corresponds to the standard Italian one *bisogna* (etymologically and probably also syntactically related to the noun *bisogno*) and is found scattered in the whole of Northern Italy. Here I provide the forms for some of the dialects where the form occurs:

> Aldeno (Trentino), Altavilla Vicentina (Veneto) *bisogna*, Bologna (Emilian), *bisagna*, Alassio (Ligurian), Albosaggia (Lombard) *besogna*, Cairo Montenotte (Ligurian/Piedmont) *zogna*, Cicagna (Ligurian), Monno (eastern Lombard) *mia*.

The second type is the one also found in Italian, and uses forms of the verb *toccare* 'to touch', found for instance in:

> Aquileia (Friulian) *tocia*, Carrara (Tuscan) *al toc*, Cimolais (Friulian) *a toscia*, Cordenons (Friulian) *al tocia*, Novi Ligure (Piedmont) *u tuca*, Poirino (Piedmont) *antuca*.

A third form comes from Latin *convenit* and is found (among other places) in:

> Livorno Ferraris (Piedmont), Torino (Piedmont) *a venta*, Campitello di Fassa (Ladin) *se con*, Cencenighe (Northern Veneto) *cognon*, Falcade (Northern Veneto) *ne cognon*, Rocca Pietore (Ladin) *el cognon*, S. Michele al Tagliamento (Friulian) *scuin*, Rodoretto (Provençal) *entò*.

Another form, corresponding to the impersonal form *ci vuole*[14] in standard Italian is the one found in:

> Farra di Soligo (Veneto) *ghe vol*, Trieste (Colonial Venetian) *ghe vol*, Milano (Lombard) *ghe vor*.[15]

Another type corresponds to *aver da* 'have to', and is found in:

> Semogo (Northern Lombard) *eres de*, Valfurva (Northern Lombard) *ares da*, Villa di Tirano (Northern Lombard) *aris da*, Frontale di Sondalo (Northern Lombard) *s'a da*, Calcinate (Lombard) *an ga de*, Ferrara (Emilian) *al gh'è da*, Bologna (Emilian) *avan da*.

A borrowing from German is also found, though only in the Ladin varieties in contact with German dialects (and with the standard):

> Laste (Ladin) *el moza*, Selva di Val Gardena (Ladin) *messon*, Corvara (Ladin) *an mess*.

We can conclude that the verb corresponding to 'it is necessary' uses a variety of forms (some of them impersonal, others construed with a subject clitic), while the volitional verb is extremely stable. Moreover, the type of scattering we find across the whole area is also worth noting.

4.2 Quantifiers

The same dichotomy observed for modal verbs between very stable roots and verbs that display several different realizations is found within the domain of quantifiers: all dialects have the same lexical element meaning 'all, everything', though adapted to the phonology of the language. I provide here some examples from the whole area:

> *tutu* Alassio (Ligurian),
>
> *tut* Bondeno (Romagnol), Aldeno (Trentino), Malonno (Eastern Lombard), Cuneo (Western Piedmont),

tot Cesena (Emilian), Lecco (Western Lombard),

tuto Vicenza (Veneto),

tout Rodoretto (Provençal),

dut S. Michele (Friulian).

On the contrary, the quantifier 'much' varies a lot: again we find several types. The first one is formed by *tanto* 'much' and its compound forms:

Alassio (Ligurian) *tantu*, Andreis (Friulian) *mitant*, Altare (Ligurian) *atant*, Cimolais (Friulian) *betant*, Erto (Friulian) *matant*.

The second type is formed by the adverb *bene* 'well' and its compounds:

Rodoretto (Provençal) *bien*, Livorno Ferraris (Lower Piedmont) *motoben*, Poirino (Piedmont) *motobin*, Vallecrosia (Ligurian) *benben*, Sondalo, Grosio (Northern Lombard) *begè*.

The third type corresponds to an indefinite determiner (sometimes followed by the adjective 'nice') and a noun of quantity:

Albosaggia Bormio (Lombard) *un bel po'*, Lecco (Lombard) *un bel pitt*, Chioggia (Veneto) *mondo*, Carpi (Emilian) *dimondi*, Forlì (Romagnol) *namasa*, Cesena (Romagnol) *una masa*, S. Michele (Friulian) *una vora*.

Other less common types are the following ones:

Campitello di Fassa (Ladin) *muie*, Lonato (Eastern Lombard),

fes ('thick'), La Spezia (Ligurian),

assé ('very'), Ferrara (Romagnol),

puras ('also + very'), Corvara (Ladin),

trop ('too much'), Monno (Eastern Lombard),

gran ('big') Casarza (Ligurian),

dubelu ('of + beautiful').

Again we can state that while the quantifier 'everything' has a single etymological origin, the quantifier 'much' has several. Moreover, the geographical distribution of the forms is not according to areas, but in 'leopard spots'. Let us take for example the Lombard area: the forms found are not homogeneous, but rather represent several of the forms seen above: Lecco (Lombard) *un bel pitt*, Lonato (Eastern Lombard) *fes*, Monno (Eastern Lombard) *gran*, Albosaggia Bormio (Northern Lombard) *un bel po'*, Sondalo, Grosio (Northern Lombard) *begè*.[16]

4.3 *Wh*-items

The third example is provided by *wh*-elements, where we observe the same etymological stability for the forms of some *wh*-words, while others vary much more, without an apparent reason. For instance, the *wh*-item corresponding to 'who' is extremely stable: the form is always the same modulo the phonological rules of the dialect.

The form *chi* is attested in the majority of the dialects.

The form *ci* is attested in Illasi (Western Veneto), Livigno (Northern Lombard) and Tassullo (Trentino), due to palatalization.

The form *cui* is attested in Friulian dialects and is due to diphtongization.

The form *che* is attested in Chioggia, Frontale di Sondalo, S. Leonardo and is due to lowering of the vowel.

The situation of the *wh*-item corresponding to 'where' is very different. The various forms found look like different morphemes that are realized or not according to the dialect. Interestingly, the lexical variation is not random, but at least some of the various formatives can be identified as functional elements (see Munaro and Poletto (in prep.) for a detailed analysis of all the formatives that are part of the *wh*-item 'where' in Northern Italian dialects).

The first formative is a preposition corresponding to 'in', visible in the dialects that have forms like the following (again, I quote here only some of the numerous examples):[17]

endo (Aldeno, Trentino) *indulà* (Aquileia, Friulian) *indunda* (Borgomanero, Eastern Piedmont).

The second formative that can be identified inside the word 'where' corresponds to the complementizer *che/ca: engleca* (Albosaggia, Northern Lombard), *olache* (Campitello di Fassa, Rhaetoromance).

The third formative is a locative tonic form 'there' *là*, visible in Friulian dialects, which generally do not have a clitic counterpart for the locative, while other Northern Italian dialects do: *ulà* (Andreis) *dulà* (S. Michele) *indulà* Aquileia, *daulà* (Barcis).

The fourth formative is the morpheme *d-* visible in some of the cases reported above as well as in: *onde* (Calalzo), *dove* (central Veneto).

The last formative is the copula, which is rarer, but reported in Piedmontese and Veneto dialects: *andé* (Cirvoi, Tignes d'Alpago) *anté* (Poirino, Piedmont).

Again, we conclude that lexical variation concentrates onto some elements while sparing others, and that this phenomenon cuts across classes like modal verbs, quantifiers and *wh*-items. In Section 5 I will try to provide some suggestions for a possible solution of the puzzle.

5. Lexical and syntactic complexity

In the previous sections we have seen that there is no reason to believe that micro-variation is qualitatively different from macro-variation, as the span of possible structures that can be found is essentially the same. In this sense, van Koppen (2011) is right in her hypothesis that the range of possible syntactic structures found concerning, for instance, the realization of subjects is the same found in a dialectal area (she uses the Dutch dialects as the empirical field on which she tests her claim, I replicated her findings with Italian examples here) or across distant languages. However, the fact that statistically dialects concentrate on a much more limited range of variation (and for instance agglutinative morphology is much rarer in the Italian domain than inflectional morphology, or OV orders are generally much less widespread, though possible, than VO) is also true and has to be taken into account. The present generative theory of syntax, taken in its minimalist or cartographic realizations, does not permit encoding of statistical facts into the syntactic principles themselves, and most probably correctly so.[18] Therefore, we have two main problems to deal with:

a. On the one hand, the lack of variation that we find across dialects, which have indeed for the vast majority very similar grammars.
b. On the other hand, the fact that in principle we can find the same type of variation found across distant languages, though it is distributed statistically in a different way. All possibilities exist, but some are realized only in very few cases or in a very limited fashion.

In order to solve the first problem, we could resort to ideas about linguistic change that have been circulating in the recent literature: some authors (see recent work by Longobardi and Guardiano, 2009; Bortolussi *et al.*, 2011) have tried to develop the idea that there are 'clusters of parameters', in the sense that the setting of one parameter causes automatically a given value of other parameters. Although this view is definitely worth exploring, and probably correct for several clusters of phenomena,[19] I think that much of the lack of variation we see in dialectal areas is due to a different mechanism, namely the persistence of lexical and categorial properties of the elements which realize functional heads. This is the only way to explain why dialects of the same area still develop similar strategies which have nothing to do with the standard language as we have seen to be the case for French and Northern Italian dialects.

If the grammaticalization process that turns a lexical item into a functional one still retains some of the basic properties of the element itself, it is clear why dialects move in the same direction: they change according to the same 'path', because the element that realizes a given functional head still keeps some of its basic properties (be it the complementizer in main interrogatives or the verb 'be' in cleft constructions or any other functional head). This implies that grammaticalization processes have a very heavy inertial weight to carry with them. Not only is the (semantic, phonological and syntactic) impoverishment important in language change, but so are the formal properties that remain constant of the original lexical element selected to spell out a given functional head.

If dialectal variation were to be explained within a theory that formalizes the existence of 'clusters' of parameters (see again Longobardi and Guardiano, 2009), we would expect no exceptions, hence no 'rare phenomena',[20] but the fact that we do find the same range of variation also found in macro-variation tells us that parameter clustering is not responsible for the strange statistical effect found among dialects, which are on the one hand very often so similar, but on the other can vary as much as distant languages can.

The common syntactic basis of dialects is due to their common lexical basis, and to the fact that elements that are used to spell out functional features keep several of the original properties through time. Since parameters are conceived as formal properties of the functional lexicon (see the discussion in Section 2, and for a formal approach to diachronic change Roberts and Roussou, 2003), the functional lexicon plays an essential role in language variation. The whole approach to grammaticalization is actually based on this assumption. On the other hand, the elements reinterpreted as functional originally belong to the non-functional part of the lexicon, and therefore, in the end also the non-functional part of the lexicon matters or at least some of its original properties do.

It is always possible that the functional lexicon varies, as a new element is selected to represent a functional head. At this point the special type of 'leopard spots' variation that we have noticed in the previous section comes in handy, as it shows that within natural classes like *wh*-items, quantifiers and modals, variation concentrates around some elements but not others.

The reason why this is so becomes clear if we consider the last example mentioned above, namely the one of the *wh*-item 'where': in this case we face a much greater variation, but at least some of the formatives used to lexicalize the *wh*-item are identifiable as prepositions, the complementizer, the verb 'be', or different types of locative deictic elements. This shows that the element 'where' is made up by a number of different formatives that spell out different features most probably

located on different functional projections, hence the semantics/syntax, morphology of the element 'where' is more complex than the one of other *wh*-items, like for instance 'who'. In this light, I would like to propose that the type of variation we have identified as 'leopard spots' concentrates precisely on those functional items whose internal structure is actually more complex and that languages (and dialects) can vary according to the number and the type of internal features/projections they can realize of a functional item like for instance a *wh*-item. I will not try to technically develop this line of thought any further here for space reasons (see on this Munaro and Poletto (in prep.)), but I will propose as a working hypothesis the assumption that functional elements that are internally more complex are also those that present a greater degree of lexical variation (with great etymological oscillations, as we have seen) and this peculiar distribution in leopard spots in dialectal areas.

We can conclude that variation concentrates where the syntax and most probably the semantics are more complex; the more semantic features there are, the more functional projections in the syntax, the more we will find different elements that spell out different portions of a complex internal structure. Other elements which have a less complex internal functional structure, because they are semantically less complex, are more stable in the lexicon and do not present leopard spots variation.

This assumption can lead us to discover that elements that we believe *prima facie* to be more complex, in fact are not. For instance, this could be the case of the modal verb: one might wonder why a verb like 'need' should be more complex than a verb like 'want/will'. However, Cattaneo (2009) has proposed that the verb meaning 'need' in Italian dialects actually derives from a noun, the noun *bisogno*, and that the noun enters the derivation as such and is incorporated into a null verb.[21]

Other cases, which might seem implausible at first sight, might reveal an unexpected internal complex structure at a closer scrutiny. In other words, what I propose here is a new type of empirical research that capitalizes on the geographical distribution of forms to single out which elements might be more complex.

On the theoretical side, if the preliminary observations made here are correct, we have a handle to understand why dialectal variation is the way it is and the two problems mentioned above without having to embed statistics into grammar. Given that dialects share the same lexicon (or a great portion thereof), the words that can serve to spell out functional heads will only allow for a limited set of variations, as they tend to keep their original properties, and therefore behave alike; even when the language changes, it changes through the same path. This solves the first problem, namely why dialects are often so similar to one another and also why they change in a similar way, often in a direction that is not similar to the one of the standard language putting pressure on them.

Given that some functional items are internally complex, and contain several pieces of information that can be spelled out in order to 'represent' the whole functional element, it can happen that different dialects select different features (hence a different lexicon) to spell out the whole functional item, and therefore they can vary as much as distant languages can. Whether all cases of dialectal variation can be reduced to this picture is an empirical question, that requires a lot of field work to be answered, and I will not attempt to do so here. The aim of this article was to show that we should never overlook empirical facts, even statistical facts, because they do not fit our theory of syntax. What odd facts can tell us is often much more than standard cases of language change, and a broader perspective is often needed to see where we have to aim at in our empirical research. On the other hand, if the line of thought I have proposed here is correct, etymology has still to be considered as one of the basic tools for dialectological research.

Notes

1. See the seminal work of the Paduan group directed by Paola Benincà and carried out through the ASIt project, Atlante Sintattico d'Italia, and related activities from 1991 until today (http://asis-cnr.unipd.it).

2. When not otherwise indicated, all the data presented in this article stem from the ASIt project, whose questionnaires are only partially available online. To write this article I could also profit from material as yet unpublished on the ASIt website.

3. Needless to say, I will take for granted the fact that there is no language internal difference between a dialect and a language in the sense that the degree of computational complexity of a language and a dialect are similar. It is clear that dialects do not undergo any pressure to standardization and (especially in the European domain) are subject to language contact with the standard (whatever this might mean in terms of the internal system), but it is far from evident that this influences the level of complexity of these varieties.

4. Dialectometry has been developed by H. Goebl and colleagues; among his vast production I quote here Goebl (2010) that seems to me the most relevant article with respect to the points discussed here, but see also Goebl (1982) for the general framework of dialectometry.

5. Pellegrini's work does not explicitly take into account syntactic variables, but some of his morphological variables are indeed syntactic in more modern terms, like the existence of some inflectional paradigms in the verbal and pronominal system.

6. She abstracts away from purely phonological or morphological distinctions, like tone, and concentrates on the syntactic differences; I will follow her on this point.

7. An anonymous reviewer points out that Italian dialects do not display parasynthesis, tone or ergative systems. Actually, D'Alessandro (2010) shows that Abruzzese dialects display some phenomena pertaining to the split ergative system. I think that what van Koppen points out here

is not the type of variation found in morphology, but only the syntactic distribution of subject realization.

8. An anonymous reviewer points out that there are no cases of incorporation in Italian dialects. Again, although this is evidently true, there are phenomena that lead us to think that the possibility of incorporation is not in principle excluded, it is only extremely rare: Longobardi (1991) analyzes the behaviour of the noun *casa* 'house/home' in standard Italian and Italian varieties and proposes that this is a case of incorporation of the noun into the position of the determiner.

9. These examples have been provided by Laura Vanelli and are discussed from a theoretical point of view in Poletto (2006). The phenomenon was first noted by Benincà (1994:83 endnote 4).

10. Here I have glossed the unmarked form of the masculine singular as [-agreement].

11. But see Mensching and Remberger (2011) for a treatment of dialectological variation in minimalist terms.

12. The data I use come from the ASIt Project, see note 1.

13. As an anonymous reviewer pointed out, Sardinian is an exception, as it uses *kerrere*, derived from Latin *quaerere*. Clearly, this is an exception to the generalization, but still, on a quantitative basis, the verb corresponding to 'want' is much more stable than the one corresponding to 'need'.

14. The form literally means 'there wants' but it is used, as all the others quoted here, in the meaning of 'need'.

15. For a detailed analysis of this extension of *volere* 'want/wish' to a deontic meaning see Benincà and Tortora (2009).

16. I will not try to provide an etymological origin for all the various types found in the sample, as this would imply a detailed analysis of each variant and would be a whole article by itself. Munaro and Poletto (in prep.) are doing precisely this for the *wh*-item *dove* 'where', but this should clearly be attempted also for the quantifier 'much'. I hope that the point concerning variation is clear although the single etymological origins and processes are not developed here.

17. We have highlighted the relevant morpheme within the form.

18. There is a longstanding debate between sociolinguists, who claim that optionality should be embedded into grammatical principles and the generative approach, where optionality is not a feature of human grammar per se: when we see optional phenomena, they are derived from the coexistence of two distinct grammars, see Kroch (1989) for a proposal in this sense.

19. It is one of the possible ways in which we could translate the old concept of 'parameter' from the Government and Binding classical framework into newer theories.

20. As noted by an anonymous reviewer the theory of parameter clustering could include the notion of fuzziness at the edges of the cluster. However, the two formal versions of a theory of parameter clustering I am aware of (namely Baker, 1997; Longobardi and Guardiano, 2009) do not formalize this notion, they are both conceived in a way in which rather the opposite happens, namely, the setting of a parameter determines automatically the setting of the parameters related to it, as they are in a hierarchy. Until the theory of parameter finds a principled way to embed the notion of fuzziness (which probably means giving up or at least weakening the idea of a hierarchy of parameters) the problem remains.

21. Cattaneo (2009:313) provides a complex structure for need in Bellinzonese, which I report here. For the detailed analysis and the discussion of the data I refer to Cattaneo (2009) chapter V.

(i) [Have *Bisogn*] [DPpossessor [PP to [SC P [... eat [something]]]]]

References

Baker, M. 1997. *The atoms of languages*. New York: Basic Books.

Benincà, P. 1994. *La variazione sintattica*. Bologna: Il Mulino.

Benincà, P. and C. Poletto 1994. "Bisogna and Its Companions: the Verbs of Necessity". *Paths towards Universal Grammar, Festschrift for R. Kayne*. G. Cinque et al. (eds), 35–58. Georgetown: Georgetown University Press.

Benincà, P. 1996. "Agglutination and Inflection in Northern Italian Dialects". *Aspects of Romance Linguistics. Selected Papers from the Linguistic Symposium on Romance Languages XXIV*, C. Parodi et al. (eds). Washington: Georgetown University Press.

Benincà, P. and C. Tortora. 2009. "On clausal architecture: evidence from complement clitic placement in Romance". *Papers from the workshops at the Linguistics Institute 2008*. Torrens et al. (eds). 219–237. Barcelona: Cambridge Scholars Publishing.

Bortolussi, L., C. Guardiano A. Sgarro and G. Longobardi. 2011. "How many possible languages are there?". *Proceedings of the 1st International Work-Conference on Linguistics, Biology and Computer Science*. Tarragona: Interplays.

Cattaneo, A. 2009 "It Is All About Clitics: The Case Of A Northern Italian Dialect Like Bellinzonese". PhD Thesis. New York University.

Cornips, L. and C. Poletto. 2004. "Standardisation of syntactic elicitation techniques". *Lingua* 16:1–19.

D'Alessandro, R. 2010. "Introduction. The verbal domain: TP-VP structure and auxiliaries". R. D'Alessandro, A. Ledgeway and I. Roberts. (eds). *Syntactic Variation. The dialects of Italy* (in press) Cambridge: Cambridge University Press.

Elsig, M. and S. Poplack. 2006. "Transplanted dialects and language change: question formation in Québec". *University of Pennsylvania Working Papers in Linguistics* 12(2):77–90.

Goebl, H. 1982. *Dialektometrie. Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie*. Wien, Denkschriften der Österreichischen Akademie der Wissenschaften, phil.-hist. Klasse, Volume 157.

Goebl, H. 2010. "Dialectometry and quantitative mapping". A. Lameli, Kehrein, R. and S. Rabanus (eds). *Language and Space. An International Handbook of Linguistic Variation*. vol. 2: *Language Mapping*. Berlin: Mouton de Gruyter.

Kayne, R. 1975. *French Syntax*. Cambridge: MIT Press.

Kayne, R. 1996. "Microparametric Syntax. Some Introductory Remarks". J.R. Black and V. Motapanyane (eds). *Microparametric Syntax and Dialect Variation*. ix–xviii. Amsterdam: John Benjamins.

Kayne, R. 2004. "Some Notes on Comparative Syntax with Special Reference to English and French". G. Cinque and R. Kayne (eds). *The Oxford Handbook of Comparative Syntax*. 3–69. New York: Oxford University Press.

Kroch, A. 1989. "Reflexes of Grammar in Patterns of Language Change". *Journal of Language Variation and Change* 1(3):199–244.

Lamarre, C. 2011. "Looking for Verbal Suffixes in Northern Mandarin: Less Typological Constraints on Grammaticalization". Talk delivered at the VII EACL Conference, Venice, 13–15 September 2011.

Longobardi, G. 1991. "Proper Names and the Theory of N-movement in Syntax and Logical Form". *University of Venice Working Papers in Linguistics* 9.

Longobardi, G. and C. Guardiano. 2009. "Evidence for syntax as a signal of historical relatedness". *Lingua* 119:1679–1706.

Manzini, M.R. and L. Savoia. 2011. *Grammatical categories*. Cambridge: Cambridge University Press.

Mensching, G. and E.-M. Remberger. 2011. "Syntactic variation and change in Romance: a minimalist approach". P. Siemund (ed). *Linguistic universals and language variation*. Berlin/New York: De Gruyter.

Munaro, N. and C. Poletto. (in prep.). "Etymology as Frozen Syntax" to appear in Benincà et al. (eds) *Festschrift in honor of Mair Parry*.

Oniga, R. 2004. *Il latino. Breve introduzione linguistica*. Milano: Franco Angeli.

Pellegrini, G.B. 1977. *Carta dei dialetti italiani, Commento*. Pisa: Pacini.

Poletto, C. 2000. *The Higher Functional Field: Evidence from Northern Italian Dialects*. New York: Oxford University Press.

Poletto, C. 2006. "Parallel Phases: a study on the high and low left periphery of Old Italian". *Phases of Interpretation. Studies in Generative Grammar* vol. 91. M. Frascarelli (ed). Berlin: Mouton de Gruyter.

Renzi, L. 1986. *Nuova Introduzione alla Filologia Romanza*. Bologna: Il Mulino.

Roberts, I. and A. Roussou. 2003. *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.

van Koppen, M. 2011. "The distribution of phi-features in pronouns". *Natural Language and Linguistic Theory*. 1–43.

Vanelli, L., L. Renzi and P. Benincà. 1985. "Typologie des pronoms sujets dans les langues Romanes". *Actes du 17ème Congrès International de Linguistique et Philologie Romanes* (Aix-en-Provence l983) vol. 3. 164–176.

Author's address

Cecilia Poletto
Institut für Romanische Sprachen
Goethe Universität Frankfurt
Campus Westend
Grüneburgplatz 1
60629 Frankfurt am Main
Germany

poletto@em.uni-frankfurt.de

From contrastive linguistics to linguistic typology*

Johan van der Auwera
University of Antwerp (Belgium)

The paper looks back at Hawkins (1986), *A comparative typology of English and German*, and shows, on the basis of raising and human impersonal pronouns in English, Dutch and German, that contrastive linguistics can be viewed as a pilot study in typology. It also pleads for doing the contrastive linguistics of three languages rather than of two, not least because the third language can teach us something about the other two.

Keywords: contrastive linguistics, typology, raising, human impersonal pronoun, semantic map

1. Introduction

In the wake of Hawkins (1986), *A comparative typology of English and German: Unifying the contrasts*, this paper makes a case for allowing contrastive linguistics the role of pilot typology. It also sketches how 'contrastive linguistics as typology' has to meet the standards of both contrastive linguistics and typology and it pleads for doing more contrastive linguistics with three languages rather than two.

2. Contrastive linguistics as pilot typology

One way of comparing the current state of two fields is to compare what practitioners of these fields did at recent, important, and representative international conferences specifically devoted to these fields. I believe that the Sixth International Contrastive Linguistics Conference held in Berlin in September 2010 and the 8th International Conference of Linguistic Typology held in Berkeley in 2009 are such conferences. At the Berlin Contrastive Linguistics Conference, there was a pronounced preference for studying issues particular to two languages, rather