

## Investigating Interrogation in the Northern Italian Area - or What Corpora Can Tell Us and Which Questions they Leave Open

Franziska Maria Hack

---

**Electronic version**

URL: <http://journals.openedition.org/corpus/1871>

ISSN: 1765-3126

**Publisher**

Bases ; corpus et langage - UMR 6039

**Printed version**

Date of publication: 1 November 2010

Number of pages: 137-164

ISSN: 1638-9808

**Electronic reference**

Franziska Maria Hack, « Investigating Interrogation in the Northern Italian Area - or What Corpora Can Tell Us and Which Questions they Leave Open », *Corpus* [Online], 9 | 2010, Online since 04 July 2011, connection on 02 May 2019. URL : <http://journals.openedition.org/corpus/1871>

---

# Investigating Interrogation in the Northern Italian Area – or What Corpora Can Tell Us and Which Questions they Leave Open –

Franziska Maria HACK  
University of Oxford

## 1. Introduction<sup>1</sup>

The aim of this article is twofold: it intends on the one hand to provide critical reflections on methodological issues on the basis of research into question formation, and on the other to show some interesting data elicited in recent fieldwork studies. Section two addresses question formation in the Northern Italian area, with particular reference to varieties of Dolomitic Ladin, and presents some interesting instances of syntactic variation. Section three deals with corpora available for the Northern Italian area, illustrating their advantages but also their limitations and some problems they pose with regard to research on question formation. Section four presents the methodology used in some recent fieldwork studies of my own, and section five concludes the paper arguing that corpus-based research and fieldwork studies may fruitfully complement each

---

<sup>1</sup> I gratefully acknowledge the support for various fieldwork studies provided by the University of Konstanz; Balliol College, Oxford and the Faculty of Medieval and Modern Languages, University of Oxford. I would also like to thank the staff of the Istitut Cultural Ladin ‘Majon di Fascegn’ for helping me with finding informants and Hans Goebel for providing me with sample maps of the *ALD-II*. I am indebted to the audiences of the *Workshop on particles*, Cambridge (30-31/10/08) and the *XXXI. Romanistentag*, Bonn (27/09-01/10/09) as well as to Josef Bayer, Georg A. Kaiser, Aditi Lahiri, Louise Mycock, Sandra Paoli, Sabrina Rasom and two anonymous reviewers for helpful suggestions. Thanks to Steven Kaye for proofreading the manuscript. My greatest thanks go to my informants in Val di Fassa and Val di Non.

other. In order to discuss methodological problems syntactic issues will be dealt with in a certain depth. However, for reasons of space it will not be possible to treat these issues in an extensive way.

## 2. Question formation

The languages of the world make use of different strategies to mark a clause as a question ('interrogative clause typing'), which may be syntactic (change in word order such as inversion of subject and verb (henceforth SVI) or fronting of the wh-expression), morphological (e.g. dedicated question markers) or prosodic (e.g. rising intonation contour in yes/no-questions or particular type of 'tune').

### 2.1 Question formation in the Northern Italian area

The traditional and most widespread question formation strategy in the Northern Italian area is inversion of subject (or subject clitic, henceforth SCL) and verb (SVI) (Renzi & Vanelli 1983, Benincà & Poletto 1997, Poletto 2000).

- (1) Subject-verb inversion (Benincà & Poletto 1997: 7)  
 Cosa dice=SCL?  
 what say-3SG=SCL  
 'What is he saying?'

However, it has been observed that traditional SVI in questions is being lost in many of these varieties and replaced by other question formation strategies that allow maintenance of the 'direct' word order SVO, leading to enormous structural variation in interrogatives in the varieties of the Northern Italian area.

- (2) Structural variation in interrogatives in the Northern Italian area<sup>2</sup> (cf. Benincà & Poletto 1997: 7f.)  
 (2a) Cosa è che SCL=dice? (wh-cleft)  
 what be-3SG that SCL=say-3SG

---

<sup>2</sup> I cite only the constructions with direct word order here. Benincà & Poletto (1997) give a number of further question formation strategies involving SVI or the presence of both preverbal and postverbal subject clitics at the same time.

*Investigating Interrogation in the Northern Italian Area*

- (2b) Cosa che SCL=dice? (wh-che-SVO)  
what that SCL=say-3SG
- (2c) Cosa SCL=dice? (wh-SVO)  
what SCL=say-3SG
- (2d) SCL=dice cosa? (wh-in-situ: SVO-wh)  
SCL=say-3SG what

In this paper, I mainly concentrate on the Dolomitic Ladin varieties of Rhaeto-Romance, as this group of closely related varieties exhibits enormous (morpho)syntactic variation in question formation in a relatively small geographical area. Therefore, these varieties represent a particularly promising area of investigation for research on (morpho)syntactic microvariation.

**2.2 Question formation in Dolomitic Ladin**

‘Dolomitic Ladin’ is the term commonly used in order to refer to the Rhaeto-Romance varieties spoken in the valleys around Mount Sella in the Northern Italian provinces of Bolzano, Trento and Belluno, namely Gherdëina, Badiot/Marè, Fascian, Fodom and Anpezan (figure 1).<sup>3</sup> According to some scholars (and politicians), the varieties of Agordino, Val di Zoldo, Cadore, Comelico and Val di Non (Nònes) / Val di Sole (Solandro) should also be considered Dolomitic Ladin varieties. It is not the purpose of this paper to make any statement on this issue: however, for ease of exposition, the former varieties around Mount Sella will be referred to in this paper as ‘Central Dolomitic Ladin’.

---

<sup>3</sup> Dolomitic Ladin is since Ascoli (1873) and Gartner (1883) classified together with Swiss Romansh and Friulian under the umbrella term “Raetoromanisch” (engl. “Rhaeto-Romance”).

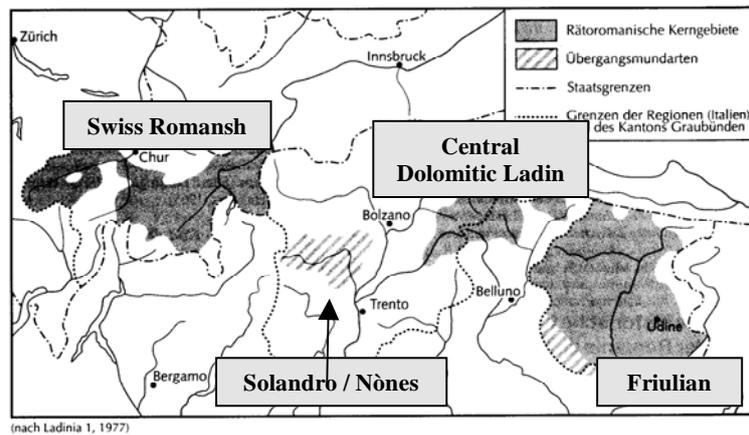


Figure 1. The language areas of Rhaeto-Romance  
(adapted from Kristol 1998: 938)

One particular phenomenon of interest found in Central Dolomitic Ladin as well as in Nones and Solandro is the particle *po* (*pa* being a phonological variant) which originally derives from the Latin temporal adverb *POST* ('then', 'afterwards') (cf. Pellegrini 1972) and developed into a focus particle by means of a grammaticalisation process (cf. figure 2), thus largely losing its original temporal meaning. Furthermore, in some Dolomitic Ladin varieties the particle has developed into an obligatory question marker in *wh*-questions (Gherdëina, Badiot/Marè); in Gherdëina, it is nowadays obligatory in both *wh*- and *yes/no*-questions (cf. Hack 2009, *subm.*).<sup>4</sup>

<sup>4</sup> For a similar development observed for the German particle *denn* and its clitic counterpart *-n* in Bavarian see Bayer (to appear).

*Investigating Interrogation in the Northern Italian Area*

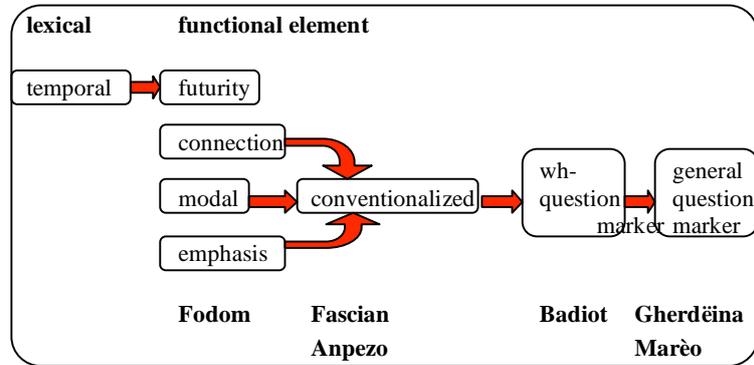


Figure 2. Grammaticalisation stages of the particle *po* (*pa*) in the ‘Central Dolomitic Ladin’ varieties (cf. Hack 2009, *subm.*)<sup>5</sup>

As regards the syntactic position and semantic contribution of the particle, there is huge variation among the Dolomitic Ladin varieties. In what follows, we will focus on the syntactic variation in *po*-questions.

The ‘default position’ of the particle in all varieties is directly after the complex of the finite verb and the enclitic subject pronoun (SVI) in simple tenses (3).

- (3) wh-SVI-*po* (Fascian) (Benincà 1995: 67)  
 Olà vas=to *pa*?  
 where go-2SG=SCL PA  
 ‘Where are you going?’

In compound tenses, however, we find variation. In the ‘Central Dolomitic Ladin’ varieties, the particle occurs between SVI and the participle (4a) whereas in the varieties of Val di Non/Val di Sole, the particle always occurs in postparticipial position (4b).<sup>6</sup>

<sup>5</sup> The interlocking boxes mean that a higher stage includes the previous ones, i.e. in Badiot/Marè, the particle *pa* is a marker of wh-questions but also exhibits the function of a reinforcing particle (focus/emphasis), and there are also cases in which it retains its original temporal meaning.

<sup>6</sup> When not otherwise indicated, the examples are data elicited in recent fieldwork studies of my own (cf. Hack 2009, *in prep.*). See section 4 for details on the fieldwork methodology used.

- (4a) wh-SVI-*po*-PTCP (Fascian Brach)<sup>7</sup>  
 Olà à=le *po* magnà la torta  
 where have-3PL=SCL PO eaten the cake  
 Francesca e Marianna?  
 Francesca and Marianna  
 ‘Where did Francesca and Marianna eat the cake?’
- (4b) wh-SVI-PTCP-*po* (Nònes, Cles)  
 Ndo à=le magnà *po* la polenta  
 where have-3PL=SCL eaten PO the polenta  
 la Francesca e la Marianna?  
 the Francesca and the Marianna  
 ‘Where did Francesca and Marianna eat the polenta?’

We shall come back to this phenomenon in section 3.2. As illustrated in (5), the particle may also be found directly after the *wh*-expression.

- (5) Olà *pa* tu vas? (Benincà 1995: 67)  
 where PA you go-2SG (Fascian Brach & Cazet)

According to native speakers of Fascian, greater stress is laid upon the *wh*-expression when this is directly followed by the particle. It is not the purpose of this paper to give a detailed account of the differences in meaning triggered by the presence vs. absence of *po* in questions in the individual varieties or by the various syntactic positions the particle may assume.<sup>8</sup> Instead, let us concentrate in the following on the purely syntactic properties of *wh-po*-questions.

In fact, in contrast to questions with *po* in postverbal position (cf. (3), (4)), which are characterised by SVI, *wh*-interrogatives featuring *po* in the position directly after the *wh*-

---

7 Fascian comprises three subvarieties: Cazet in the highest part of the valley from Penìa to Ciampestrin; Brach in the central part between Pera and Soraga; and Moenat in the lowest part of Val di Fassa (Chiocchetti & Iori 2002: 9).

8 For further details on the syntactic and semantic properties of the particle see Poletto (2000), Poletto & Zanuttini (2003), Hack (2009, subm.).

expression involve direct word order (cf. 5).<sup>9</sup> The latter case, i.e. *wh-po*-questions with direct word order, is the most frequent one in the Dolomitic Ladin varieties. However, we observe an interesting phenomenon in Moenat, the subvariety of Fascian spoken in Moena, where the sequence *wh-po* is not followed by direct word order SVO as in the other varieties (6a) but by SVI (6b).

(6) Variation in Val di Fassa: *wh-po*-SVO in Cazet and Brach (a) vs. *wh-po*-SVI in Moenat (b)

(6a) Olà    *pa*    le    à                    magnà    angèrn  
where PA they have-3PL eaten yesterday  
Francesca e Marianna?  
Francesca and Marianna  
'Where did Francesca and Marianna eat yesterday?'

(6b) Olà    *po*    à=le                    magnà    na    torta  
where PO have-3PL=SCL eaten a cake  
Francesca e Marianna?  
Francesca and Marianna  
'Where did Francesca and Marianna eat a cake?'

### 3. Corpora for the Northern Italian area

There are various kinds of corpora available to researchers interested in the varieties of the Northern Italian area, but I shall discuss here only those which are relevant for syntactic purposes. Of course, each dialect description or dialect grammar on a particular local variety may represent some kind of corpus. However, we shall concentrate on two main types of corpus here: linguistic atlases and electronic corpora.

#### 3.1 Linguistic atlases

First of all, linguistic atlases covering the Northern Italian area such as Karl Jaberg and Jacob Jud's seminal work *AIS* can be exploited for syntactic purposes.

---

<sup>9</sup> A straightforward way to account for these differences would be to assume that in *wh-po*-questions, the particle occupies the position targeted by verb movement, hence inhibiting SVI (cf. Poletto 2000, Hack 2009).

Particularly useful for our specific area of interest is the *ALD* which covers 217 localities in upper and lower Engadin (Switzerland), western Lombardy, Trentino, the Ladinophone parts of South Tyrol, northern and central Veneto and western Friuli. The first part of the project, *ALD-I*, covers phonology as well as fundamental nominal and verbal morphology whereas the second part, *ALD-II*, still in preparation at the University of Salzburg under the direction of Hans Goebel, focuses on elaborated morphology and on syntax.

Another project to be mentioned is the *ASIS*, which was originally intended to register and analyse syntactic variation in Northern Italian dialects but has now developed into the *ASIt*, covering the whole of Italy. As a result of this project, situated in the Department of Linguistics at the University of Padova, a large database has been compiled which is accessible on-line (<http://asis-cnr.unipd.it>). It is possible to search the database according to geographical region, linguistic phenomenon (e.g. interrogatives, syntax and morphology of *wh*-expressions, negation, quantifiers, particles, auxiliary selection, modality etc.) or linguistic annotation. However, as there are no geolinguistic maps available the project is not a linguistic atlas in the narrow sense, but rather an enormous syntactic database.

### 3.1.1 Benefits

The major benefit which can be gained from linguistic atlases is the very reason for their existence: geolinguistic maps indicate the geographical distribution of a particular phenomenon and may thus direct the researcher to one or several specific areas characterised by the presence of the phenomenon in question.

Regarding the phenomenon of *wh-in-situ*, i.e. *wh*-questions in which the *wh*-expression does not appear in sentence-initial position but stays in its base-generated/argumental position, for instance, *ALD-II* maps allow a more or less precise geographical localisation of the phenomenon. On the basis of a number of sample maps we can observe two distinct areas characterised by the constant occurrence of *wh-in-situ* in interrogatives: on the one hand Val Camonica in Lombardy and on the other an area extending from Falcade and Valle Agordina in the North/Northeast to Valdobbiadene in the

*Investigating Interrogation in the Northern Italian Area*

South and Val Primiero in the West to the Alpego region in the East.

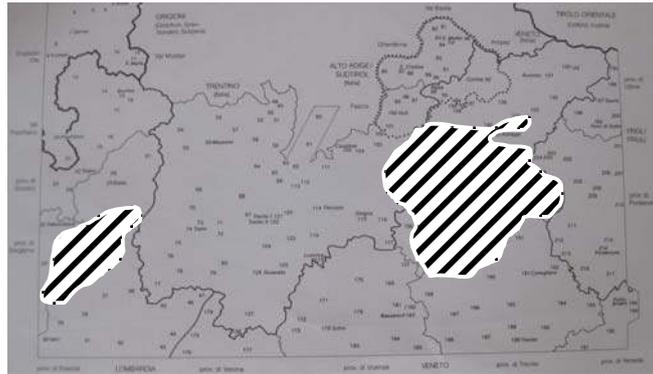


Figure 3. The areal distribution of *wh-in-situ* in the Northern Italian area (on the basis of *ALD-II* sample maps)

Another case is the geographical distribution of the particle *po* (*pa*) in *wh*-questions. Again, geolinguistic maps help to determine the diffusion area of the particle in Northern Italy. As *ALD-II* sample maps reveal, the particle occurs in two main areas, in the ‘Central Dolomitic Ladin’ valleys around Mount Sella on one side and in Val di Non and Val di Sole on the other. Furthermore, there are some individual occurrences in Valtellina, Val Camonica, Val Giudicarie, Val di Cembra, Valsugana, Valle Agordina and Val Tagliamento.

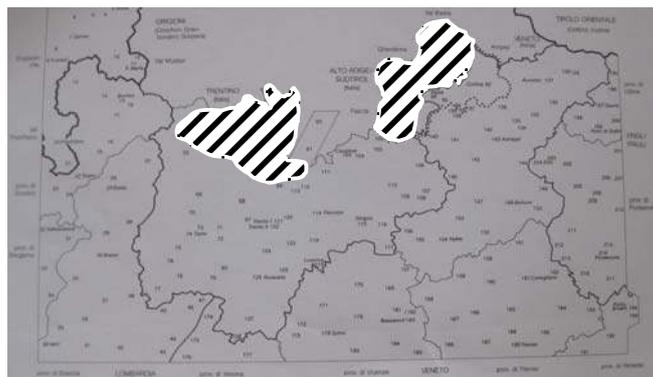


Figure 4. The main areas of occurrence of the particle *po* (*pa*) in *wh*-questions (on the basis of *ALD-II* sample maps)

### 3.1.2 *Limitations and problems*

Linguistic atlases essentially pose two different kinds of problem, one of them related to the methodology used for data elicitation and the other regarding the presentation of these data.

#### 3.1.2.1 *Methodology used for data collection*

In order to illustrate the problems related to the methodology used in the data collection process for linguistic atlases, let us first refer to subject pronoun use.<sup>10</sup> The use or non-use of subject clitics is an interesting topic with respect to Northern Italy in particular, given that most of the varieties spoken in this area exhibit so-called split-pro-drop paradigms, i.e. subject clitics are used or exist only for some grammatical persons whereas they are dropped or simply non-existent for others. Much research (e.g. Renzi & Vanelli 1983, Poletto 2000, Thiele 2001, Manzini & Savoia 2005, Hack 2007, Hack & Gaglia 2009, Kaiser & Hack *subm.*) has been undertaken to establish the paradigms of subject pronouns and subject clitics for the individual varieties and to determine the conditions for their use.

As far as this research topic is concerned, linguistic atlases such as the *ALD* prove less useful, given that subject pronouns have been systematically elicited in the data collection process. Even if this was not the case, the methodology usually used in data collection for linguistic atlases may be problematic and may provide a distorted image of the linguistic reality. Linguistic atlases normally use *questionnaires* asking informants how they would express a given word or sentence in their local variety. For this type of task one might expect the structure of the (Italian) stimulus sentence to affect the speaker's response, especially in view of the diglossic or even polyglossic situation in these areas. A native speaker of a Northern Italian variety, when confronted with an Italian stimulus sentence, might thus transpose the Italian sentence structure into his local variety and hence might

---

<sup>10</sup> We have to distinguish here between subject pronouns, i.e. pronominal argumental elements, on the one hand and subject clitics, i.e. inflectional elements, on the other.

not use the subject clitic although it would normally be used in this particular variety.<sup>11</sup>

Another problem with the *questionnaire* methodology concerns the fact that syntactic phenomena are much more difficult to deal with than phonological, lexical or basic morphological data in the sense that the syntactic structure may largely depend on the context, i.e. information structure and pragmatics, which a translation task is not able to take into consideration. On these grounds, it may emerge that a translation task is unable to capture the subtleties of syntactic structures and their interaction with contextual factors.

### *3.1.2.2 Presentation of data*

Linguistic atlases may also pose problems regarding the presentation of data.

First of all, the responses presented on geolinguistic maps might not be the only constructions used in a particular locality. There could in fact be various other constructions in use whose existence escapes notice for various reasons, be it the methodology used in the data collection process<sup>12</sup> or typographic restrictions regarding the representation of responses on the maps.<sup>13</sup>

Another problem may arise, however, when it comes to the interpretation of the data provided – though obviously this is not a problem specific to data from linguistic atlases but affects all kinds of corpus data. In order to illustrate the problems and

---

11 As shown by Cornips & Poletto (2005), and pointed out by an anonymous reviewer, the standard (in this case Italian) may affect the construction provided by the informant when his local variety exhibits both variants (here: presence and absence of SCL). According to these authors, the variant that is closer to the standard is likely to be chosen. However, we might actually also observe the opposite: the informant might deliberately produce the variant which is less common in the standard in order to highlight the peculiarity of his local variety.

12 As regards linguistic atlases, usually no more than two informants have been interviewed per locality; therefore it may well be that not all constructions used in one particular locality have been registered.

13 The linguistic atlases *AIS* and *ALD* have addressed this problem by including the data that could not be accommodated on the maps in (sometimes extensive) legends and/or appendices.

challenges in interpreting data from linguistic atlases which could lead researchers astray, let us consider the case of the particle *po* (*pa*) and its relation to subject pronoun use.

In the 19<sup>th</sup> century, the Dolomitic Ladin variety of Val Gardena, Gherdëina, exhibited enclitic subject pronouns for the first and second person plural (*-i*) (7a). Later these clitics underwent phonological reduction, changing to *-e* [ə] (7b) before being lost completely (7c) (cf. Anderlan-Obletter 1991: 38).

(7) Loss of the 2PL postverbal subject clitic in Gherdëina

(7a) ciantëis=i  
sing-2PL=SCL

(7b) ciantëis=e  
sing-2PL=SCL

(7c) ciantëis  
sing-2PL

In contrast to this development characterised by the loss of subject clitics in certain grammatical persons, we observe a reverse development in Gherdëina: the particle *pa* became obligatory first in *wh*-questions and then also in *yes/no*-questions. Hence, nowadays *pa* serves as a general question marker (cf. figure 2) and according to native speakers a sentence is not perceived as a question if this particle is lacking. Interrogatives involving the particle *pa* are already to be observed in the 19<sup>th</sup> century –and they display a quite remarkable distribution in Vian (1864). In *wh*-questions with the main verb in the second person plural, we observe either the verb with the subject clitic (8a) or the verb without the clitic but instead followed by the particle *pa* (8b).<sup>14</sup>

(8a) Co staseis=e? [ə] (Vian 1864: 176)  
how be-2PL=SCL  
'How are you?'

---

<sup>14</sup> We can conclude from the examples in Vian (1864) that in 19<sup>th</sup> century Gherdëina *wh*-questions, the enclitic subject pronouns was in complementary distribution with the particle *pa*. I found only one single example in which SCL and *pa* co-occurred.

*Investigating Interrogation in the Northern Italian Area*

- (8b) Ulà staseis *pa*?  
 where live-2PL PA  
 ‘Where do you live?’

In the modern language, however, the particle has phonologically reduced to *a* [ə] after *s* (probably concomitantly with its cliticisation onto the verb), giving rise to structures such as those in (9).

- (9) Phonological reduction of the particle *pa* in Gherdëina (Anderlan-Obletter 1991: 104)

- (9a) Ulà compres=*a* n liber?  
 where buy-2SG=PA a book  
 ‘Where are you buying a book?’

- (9b) A chi cumprëis=*a* n liber?  
 to who buy-2PL=PA a book  
 ‘For whom are you buying a book?’

Hence, both the postverbal subject clitic of 2SG and 2PL and the reduced form of the particle *pa* after a dental fricative are realised as [ə] and the development in Gherdëina can be summarised as follows:

- (10a) lascëis=e [ə] (2PL enclitic subject pronoun)  
 let-2PL=SCL
- (10b) lascëis (loss of the clitic)  
 let-2PL
- (10c) lascëis *pa* (particle *pa*)  
 let-2PL PA
- (10d) lascëis=*a* [ə] (cliticisation on the verb and phonological reduction of the particle)  
 let-2PL=PA

There are reasons to believe that, in this case, the particle *pa* actually compensates for the loss of the subject clitic (cf. Hack 2009, subm.). Let us note here only that besides both occurring in postverbal position, the 2PL subject clitic and the reduced form of the particle are homophonic.

In conclusion, linguistic atlases are without any doubt helpful and even indispensable in allowing one to localise a given phenomenon and determine its geographical distribution.

However, one has to bear in mind that data from linguistic atlases may have been elicited in a way that might be of little use to one's own particular research purposes and might give a distorted impression of the real situation. Furthermore, when it comes to data interpretation, one has to be very cautious not to be led astray. Hence, the consultation of atlas data should always be accompanied and/or followed by further investigations which allow to look at linguistic phenomena with a 'geographical or geolinguistic magnifying glass'.<sup>15</sup>

### **3.2 Electronic corpora**

Nowadays, digital corpora are an essential basis not only for research but also for language planning and standardisation, in particular as far as lesser-used languages such as Dolomitic Ladin are concerned.

The cultural institutes of the Dolomitic Ladin area have undertaken a number of projects in this respect, such as the corpus project *TALL*<sup>16</sup> *CorpusLad*, whose goal is to cover the entire written production of all the subvarieties of Dolomitic Ladin. The subproject *Corpus Ladin leterar* ('corpus of literary Dolomitic Ladin'), under preparation at the Free University of Bolzano, aims to complete this corpus with literary texts of a high linguistic level. The general goal of the project is to make available a digital database which should be representative for the use of the Dolomitic Ladin language and which should also be available for statistical analysis (cf. Videsott 2009: 19f.). The database, which currently comprises more than 14,000 texts of various genres in all codified varieties of 'Central Dolomitic Ladin' and is constantly being expanded, can be accessed online. This direct access to data from closely related varieties is particularly useful for investigations on microvariation, above all in situations where native speakers of these varieties are not directly available. Let us consider the usefulness of this kind of corpus again with respect to our investigations on question formation in the Northern Italian area.

---

<sup>15</sup> I am indebted to Hans Goebel for this very apt expression.

<sup>16</sup> *TALL* = *Tratament Automatic dl Lingaz Ladin* ('automatic treatment of the Dolomitic Ladin language')

*Investigating Interrogation in the Northern Italian Area*

As mentioned in section 2.2, we observe variation in varieties of the Northern Italian area as regards the syntactic position of the particle *po* in questions involving compound tenses. While *po* appears after SVI and is followed by the participle in the Central Dolomitic Ladin varieties, the particle occurs in postparticipial position in other varieties such as Nònes, Solandro, Pagotto. The examples in (4) are repeated here under (11) for convenience.

- (11a) wh-SVI-*po*-PTCP (Fascian, Vigo di Fassa)  
Olà à=le *po* magnà la torta  
where have-3PL=SCL PO eaten the cake  
Francesca e Marianna?  
Francesca and Marianna  
'Where did Francesca and Marianna eat the cake?'
- (11b) wh-SVI-PTCP-*po* (Nònes, Cles)  
Ndo à=le magnà *po* la polenta  
where have-3PL=SCL eaten PO the polenta  
la Francesca e la Marianna?  
the Francesca and the Marianna  
'Where did Francesca and Marianna eat the polenta?'

The variation in the syntactic position of the particle *po* is surprising. Why should the particle, in the course of its grammaticalisation process, have become fixed in different syntactic positions?

From the perspective of the cartographic approach within generative grammar and under the assumption that the particle itself (on a par with adverbs) occupies a fixed syntactic position, the variation in (11) may be linked to different functional projections targeted by the participle.

Cinque (1999) departs from the assumption that adverbs occupy fixed positions in the sentence structure. By determining the relative position of the finite verb with respect to different classes of adverbs on the basis of a cross-linguistic survey he proposes a universal hierarchy of functional projections within the IP layer. According to Cinque (*ibid.*), adverbs are located in the unique specifier position of functional projections whose head positions serve as landing sites for verb movement. Under the assumption that Cinque's hierarchy is

correct and that the particle occupies a fixed position in the sentence structure, the differences regarding the relative positions of *po* with respect to the participle might result from the (active) past participle targeting different landing sides in the varieties considered. The hypothesis is thus that in wh-SVI-*po*-PTCP-varieties (cf. 11a), the participle does not move as high as it does in the wh-SVI-PTCP-*po*-varieties (cf. 11b), given that the participle follows the particle in the former whereas it precedes it in the latter. In order to test this hypothesis, we have to determine the relative position of the participle with respect to Cinque's (1999) lower sentence adverbs in these varieties.

In fact, on the basis of the observation that in various Northern Italian dialects the (active) past participle obligatorily occurs to the left of various classes of adverbs, Cinque (1999: 146) concludes that in these varieties the past participle raises to different heads. In (standard) Italian, the active past participle must raise higher than *tutto* ('everything') (and *bene* ('well') and *presto* ('early')) (Cinque 1999: 146). In Bellunese, active past participles raise higher than *del tut* ('completely'). Cinque's (*ibid.*) observations concerning standard Italian and various Northern Italian varieties are summarised in table 1.

Italian:	√ mica	√ già	√ più	√ sempre	√ completamente	√ tutto	* bene	*
Friulian:	√ mingul	√ zà	√ plui	√ simpri	√ completaminti	√ dut	* ben	*
Venetian:	* miga	* già	?? più	? sempre	√ del tuto	√ dut	* ben	*
Bellunese:	* mia	* za	* pi	?? sempre	√ del tut	* tuto	* ben	*

Table 1. Raising of the active past participle in Italian and Northern Italian varieties (Cinque 1999:147)<sup>17</sup>

We now need to determine the relative position of the (active) past participle with respect to these adverbs in wh-SVI-*po*-PTCP (e.g. Gherdëina) and wh-SVI-PTCP-*po* (e.g. Nònes) varieties.

It is in this respect, among others, that electronic corpora like *CorpusLad* prove very useful. Access to native

17 √: position can be occupied by the past participle; \*: position cannot be occupied by the past participle.

*Investigating Interrogation in the Northern Italian Area*

speakers of a number of varieties might be difficult, and thus as a first approach one can resort to corpus data in order to get a first impression of the situation. As far as our research question is concerned, *CorpusLad* allows us to search for the individual adverbs in the whole corpus or in the specific subcorpus for the Central Dolomitic Ladin variety under consideration. The search results are provided along with the immediate environment, i.e. an (adjustable) number of words at the left and right margins of the target lemma. One can also directly access the specific text in which the target word has been found. Comparing the relative position of the (active) past participle with respect to the adverbs in table 1, the search on *CorpusLad* for the wh-SVI-*pa*-PTCP variety Gherdëina has yielded the results summarised in table 2. As there are no (electronic) corpora available for the wh-SVI-PTCP-*po* variety Nònes, results are drawn from a manual search in Bertagnolli (1912), the anthology of the variety of Val di Non, as well as from judgments by native speakers.

Gherdëina: * nianca * bele * plu * for √ cumpletamënter √ dut √ bën *
Nònes: * mǐgja * dza * pu * semper √ daltut √ tut √ ben *

Table 2. Raising of the active past participle in Gherdëina and Nònes

There does not seem to be a difference between Gherdëina and Nònes regarding the landing site of the (active) past participle, which in both varieties raises higher than *cumpletamënter/daltut* ('completely') but cannot precede *for/semper* ('always'). It seems thus that a purely syntactic analysis within the cartographic approach is not able to account for the differences between these varieties regarding the position of the particle *po* in questions involving compound tenses.<sup>18</sup>

---

<sup>18</sup> For space limitations I cannot discuss alternative syntactic analyses of this issue here, but I refer to Hack (subm.). However, there is reason to believe, that the variation regarding the postverbal vs. the postparticipial position of the particle in compound tenses cannot be accounted for in (purely) syntactic terms. The position of the particle rather seems to be

### **3.3 General problems with corpora**

An important drawback of corpora is the fact that they do not provide negative evidence, i.e. information regarding the impossibility of certain structures, which, however, is often essential in order to verify or falsify a hypothesis.

As far as our study on the syntactic positions occupied by the (active) past participle in Dolomitic Ladin varieties (cf. section 3.2) is concerned, corpora like *CorpusLad* can provide data allowing us to determine the relative position of the participle with respect to adverbs. However, they can do this only on the basis of positive evidence. We do not know whether some speakers would also accept alternative structures and we cannot say whether some speakers may allow optional movement of the participle to a higher syntactic position.

Moreover, question formation is clearly an interface issue, involving not only the (morpho)syntactic but also the phonological and the semantic-pragmatic levels of linguistic analysis as well as information structure. Furthermore, socio-linguistic variables may also be involved, e.g. in the choice of a certain question formation strategy.

In sum, as argued above for linguistic atlases in particular, atlas data are in fact extremely useful –if not essential– as a starting point, but at the end they may prove too limited and superficial for a comprehensive investigation of a particular phenomenon.

On these grounds, I argue in the following section that fieldwork –provided that it is conducted in a systematic, thoroughly elaborated and fine-grained fashion– can be a means to effectively complement corpus- or linguistic atlas-based studies.

### **4. Fieldwork – the construction of one's own corpus**

This section sets out the methodology used to investigate question formation in varieties of the Northern Italian area in some recent and future fieldwork studies of my own (cf. Hack 2009, subm., in prep.).

---

governed (also) by prosodic factors. Hack (in prep.) develops an alternative account of the phenomenon by investigating the syntax-prosody interface.

*Investigating Interrogation in the Northern Italian Area*

An important principle underlying this methodology is that it combines several tasks. On the one hand this allows the researcher to mutually validate the results obtained from different elicitation tasks such as a ‘question formation’ task (4.1), a ‘questions in context’ task (4.2), a translation task (4.3) and a grammaticality judgment task (4.4). On the other hand the tasks are complementary in the sense that they are used to elicit both positive and negative evidence.

**4.1 The ‘question formation’ task**

In the ‘question formation’ task, statements are presented to informants in their local variety. One and the same statement is presented several times, first of all without any particular constituent highlighted, and then in several further versions each marking a different constituent in bold face characters, as shown in (12) on the basis of an example in Fascian.

- (12) Angérn Tone à dat n pom de èlber a tia  
yesterday Tone has given an apple of tree to your  
sor.  
sister.  
‘Yesterday Tone gave an apple to your sister.’
- (12a) Angérn **Tone** à dat n pom de èlber a tia sor.
- (12b) Angérn Tone à dat **n pom de èlber** a tia sor.
- (12c) Angérn Tone à dat n pom de èlber **a tia sor**.

The task of the informant is thus first to read out the unmarked version and then to produce orally a question for each of the marked versions whereby the answer to the question the informant has to provide should correspond to the bold face constituent. Table 3 shows a number of interrogative structures obtained from native speakers of Fascian in response to the stimulus (12b), repeated below as (13).

- (13) Angérn Tone à dat **n pom de èlber** a tia  
yesterday Tone has given an apple of tree to your  
sor.  
sister.  
‘Yesterday Tone gave an apple to your sister.’

construction	Subvariety	type
Che ge à=l dat angérm Tone a tia sor?	Brach	wh-SVI
Che ge à=l <i>pa</i> dat Tone a tia sor?	Cazet	wh-SVI- <i>po</i>
Che <i>po</i> ge à=lo dat inier Tone a to sorella?	Moenat	wh- <i>po</i> -SVI
Che che l ge à dat angérm Tone a tia sor?	Brach	wh-che-SVO
Che è stat che ge à dat a tia sor?	Cazet	cleft

Table 3. Interrogative constructions obtained in Val di Fassa in response to the stimulus sentence in (13)

For reasons of space we cannot further discuss variation in question formation in Val di Fassa here; the reader is referred to Chiochetti (1992), Hack (2009; *subm.*; *in prep.*) for more details.

The ‘question formation’ task has a number of advantages. First of all, it allows the researcher to control for various factors such as the length of the utterance to be elicited, and to take into consideration all possible *wh*-expressions and grammatical persons. Moreover, as regards prosodic analysis, *i.e.* the analysis of intonation contours, a direct comparison between statement and question intonation is possible.

However, there are also a number of drawbacks to bear in mind.

To begin with, the design of this task requires a fairly deep knowledge of the variety under consideration, as the statements should be presented in the informant’s local variety. This may also cause orthographical problems given that local varieties are not normally codified. As speakers are thus not used to being confronted with their local variety in writing, they may feel confused, especially when given an orthography which they would not expect or with which they do not agree.

A related problem concerns the vocabulary used in the task: a single word which is uncommon in the particular local variety may be sufficient to confuse the informant.

#### ***4.2 The ‘questions in context’ task***

Another task which proved useful for the elicitation of interrogatives asks native speakers to provide a question on the basis of a given context. These short descriptions of situations contain an information gap that should incite informants to ask

*Investigating Interrogation in the Northern Italian Area*

a question.<sup>19</sup> (14) gives an example for a stimulus context and table 4 contains some of the interrogative constructions elicited on the basis of this task.

- (14) ‘Questions in context’ task  
 La signora Bertamelli ha tre figli: Giancarlo, il più grande, fa il dottore; Paolo, il secondo, è studente di linguistica.<sup>20</sup>

construction	Subvariety	type
Che fèj=el l terzo fi della signora Bertamelli?	Brach	wh-SVI
Chel più pìcol, che à=l po studià?	Brach	wh-SVI-po
Che che l à inom, l terzo?	Brach	wh-che-SVO

Table 4. Interrogative constructions obtained in Val di Fassa in response to the stimulus sentence in (14)

As specific situational contexts are used as back-grounds, this task is particularly useful for the investigation of information structure and its impact on question formation.<sup>21</sup>

**4.3 The ‘translation’ task**

Despite its caveats (cf. section 3.1.2.1), I have also included a translation task in the multitask methodology, in particular in order to be able to directly compare the elicited data with those found in linguistic atlases such as the *ALD-II*.

On the basis of a number of *ALD-II* sample maps for wh-questions, we observe that most of the localities in Val di Fassa investigated for the *ALD-II* exhibit a ‘typical’ question formation strategy,<sup>22</sup> illustrated here by the responses to *ALD-II* stimulus 207/2 *Cosa hai fatto?* (‘What did you do?’).

---

19 For practical reasons the contexts were presented in Italian, whereas the informants were of course asked to provide a question in their local variety.

20 ‘Mrs. Bertamelli has three sons: Giancarlo, the eldest one, is a doctor; Paolo, the second one, studies linguistics.’

21 However, for space limitations I cannot expand this topic here.

22 ‘Typical’ means here that in the sample maps considered, the given local variety always (or nearly always) exhibited the same question formation strategy.

locality	typical strategy	example
Campitello	wh-che-SVO	Che che te ès fat? what that you have-2SG done
Vigo di Fassa	wh-SVI	Che às=te fat? what have-2SG=SCL done
Moena	wh-SVI- <i>po</i>	Che às=to <i>po</i> fat? what have-2SG=SCL <i>po</i> done

Table 5. Localities in Val di Fassa and their typical question formation strategies (source: *ALD-II* sample maps)

On the basis of these observations, one could be led to assume that this ‘typical’ question formation strategy is in fact the only one used in the local variety under consideration. However, this is not necessarily the case, as is shown by my fieldwork. In Vigo di Fassa, for instance, I observed variation in question formation strategies –using the translation task and some of the *ALD-II* stimuli. The above mentioned *ALD-II* stimulus *Cosa hai fatto?*, for instance, yielded constructions (15a-c). Furthermore, the construction in (15d) was elicited by means of the ‘questions in context’ task (cf. section 4.2)

(15) Question formation strategies observed in Vigo di Fassa for the stimulus *Cosa hai fatto?*

(15a) wh-SVI  
Che às=te fat?  
what have-2SG=SCL done  
‘What did you do?’

(15b) wh-che-SVO  
Che che t’=às fat?  
what that SCL=have-2SG done

(15c) wh-*pa*-SVO  
Che *pa* tu às fat?  
what PA you have-2SG done

(15d) wh-SVI-*pa*  
E che às=te *pa* fat l terzo dì?  
and what have-2SG=SCL PA done the third day  
‘And what did you do on the third day?’

### *Investigating Interrogation in the Northern Italian Area*

While according to *ALD-II* sample maps strategies (15a, b and d) also occur in other localities in Val di Fassa, (15c) does not occur at all in the *ALD-II* data.<sup>23</sup>

Moreover, another question formation strategy commonly used in Moenat that does not occur in linguistic atlases is the *wh-po-SVI* construction illustrated above in (6b) of section 2.2 and repeated here in (16).

- (16) Olà *po* à=le magnà na torta  
where PO have-3PL=SCL eaten a cake  
Francesca e Marianna?  
Francesca and Marianna  
'Where did Francesca and Marianna eat a cake?'

In sum, in these fieldwork studies we have observed much more variation in interrogatives than could be predicted on the basis of linguistic atlas data – be it due to the greater number of informants involved or to the broader variety of tasks which may have been able to cover a greater variety of contexts and to take pragmatic and information structural factors into account.

#### **4.4 Grammaticality judgements**

A grammaticality judgement task was included in the fieldwork methodology in order to allow for native speakers' subjective opinions about the use of one or another question formation strategy as well as individual preferences, and in order to elicit negative evidence.

As exemplified in (17), speakers were presented with several variants of a question in their local variety and were then asked to comment on them, pointing out which one(s) they would use in a day-to-day situation, in which specific contexts they would use the construction(s), and which strategies they could not use at all.

---

<sup>23</sup> In fact, *ALD-II* sample maps show some individual occurrences of the particle *po* in the position after the *wh*-expression ('*wh-po*-position'). However, this is only found with the *wh*-expression *parché* ('why'). It has often been observed in the literature that why-questions show distinctive characteristics in a variety of languages. In Fascian, the *wh-po*-position seems to be possible with all *wh*-expressions, whereas in Nònes, it can only appear with some question words. For more details see Hack (2009).

- (17) Example for a grammaticality judgement task (Fascian)
- (17a) Olà vas-to?
- (17b) Olà vas-to *pa*?
- (17c) Olà *pa* vas-to?
- (17d) Olà che tu vas?
- (17e) Olà él *pa* che tu vas?
- (17f) Olà *pa* tu vas?
- ‘Where are you going?’

However, a procedure of this kind requires particular attentiveness on the part of the researcher. One has to be very cautious as regards the vocabulary used in this task. Just one word in the stimulus sentence which is uncommon in the local variety of the informant (though it may be very frequent in the neighbouring village) might make the speaker reject the whole sentence although the overall construction would otherwise be accepted without question. It has to be noted here that linguistic atlases are very useful as a means of avoiding this kind of slip, as they cover lexical variation in a particular geographical area.

Finally, it should be noted that an evaluation of the interrogative constructions elicited by means of the different tasks did not show any task-based artifacts, i.e. the type of task used to elicit interrogatives has no impact on the kind of interrogative construction produced by the informant.<sup>24</sup>

## 5. Conclusion

On the whole, the view taken in this paper is that corpora are not only important but indispensable as starting points for investigations, as well as bases for the building up of working hypotheses.

---

24 Apart from these ‘individual’ tasks I also used some ‘collaborative’ tasks, i.e. tasks which had to be worked on by two speakers, in order to elicit conversations among native speakers. ‘Collaborative’ tasks included a map task and a picture description task. For space limitations we cannot address these tasks here, but instead we refer the reader to Hack (2009) and to the website of the Human Communication Research Centre (HCRC) (Universities of Edinburgh and Glasgow) for a general introduction to the map task methodology: [<http://www.hcrc.ed.ac.uk/maptask/#top>].

### *Investigating Interrogation in the Northern Italian Area*

However, on the basis of research into question formation in the Northern Italian area it has been argued that, depending on the research aims, a more fine-grained methodological procedure is required which is better adapted to specific research requirements. Data collection through fieldwork can be a means to meet these demands, involving in turn the construction of one's own specialised corpus. Of course it would be desirable that a particular corpus of this kind is designed and made available in a way that it is reusable by other researchers as well.

#### **Linguistic atlases and corpora**

*AIS* = Jaberg K. & Jud J. (1928-1940). *Sprach- und Sachatlas Italiens und der Südschweiz*. Zofingen: Verlagsanstalt Ringier & Co.

*ALD-I* = Goebel H. (ed.) (1998). *Atlant linguistisch dl ladin dolomitich y di dialec vejins, 1a pert/Atlante linguistico del ladino dolomitico e dei dialetti limitrofi, 1a parte/Sprachatlas des Dolomitenladinischen und angrenzender Dialekte, 1. Teil*. Wiesbaden: Ludwig Reichert.

*ALD-II* = *Atlant linguistisch dl ladin dolomitich y di dialec vejins Atlant linguistisch dl ladin dolomitich y di dialec vejins, 2a pert/Atlante linguistico del ladino dolomitico e dei dialetti limitrofi, 2a parte/Sprachatlas des Dolomitenladinischen und angrenzender Dialekte, 2. Teil*. In preparation under the direction of Hans Goebel, Universität Salzburg, Fachbereich Romanistik.

*ASIt* = *Atlante Sintattico d'Italia*. Università di Padova, Dipartimento di discipline linguistiche, comunicative e dello spettacolo. [URL: <http://asis-cnr.unipd.it/>].

*CorpusLad* = *Corpus Ladino*. Vigo di Fassa: Istitut Cultural Ladin "Majon di Fascegn".  
[<http://corpuslad.tall.smallcodes.org/applications/textanalysis/search.jsp>]

**References**

- Anderlan-Obletter A. (1991). *La rujeneda dla oma. Grammatica dl ladin de Gherdëina*. Urtijëi: Istitut Pedagogich Ladin.
- Ascoli G. I. (1873). « Saggi Ladini », *Archivio Glottologico Italiano* 1: 1-537.
- Bayer J. (to appear). « From Modal Particle to Interrogative Marker: A Study of German *denn* », in L. Brugè, A. Cardinaletti, G. Giusti, N. Munaro & C. Poletto (eds), *Functional Heads*.
- Benincà P. (1995). « Agglutination and Inflection in Northern Italian Dialects », in C. Parodi, C. Quicoli, M. Saltarelli & M. L. Zubizzareta (eds), *Aspects of Romance Linguistics. Selected Papers from the Linguistic Symposium on Romance Languages 10-13 marzo 1994*, vol. 24. Washinton, D.C.: Georgetown University Press, 59-72.
- Benincà P. & Poletto C. (1997). « Introduzione », in P. Benincà & C. Poletto (eds), 5-11.
- Benincà P. & Poletto C. (eds) (1997). *Strutture interrogative dell'Italia Settentrionale. Publication of the Centro di Studio per la Dialettologia Italiana*. Padua: Dipartimento di discipline linguistiche, comunicative e dello spettacolo, Centro Nazionale delle Ricerche. (= *Quaderni di Lavoro dell'ASIS* 1)
- Bertagnolli G. (1912). *Poesie e poeti della Val de Non*. 3 vol. Trento: G.B. Monauni.
- Chiocchetti F. (1992). « Evoluzioni sintattiche dell'interrogativa nel fassano », *Mondo Ladino* 26.3-4: 199-219.
- Chiocchetti N. & Iori V. (2002). *Grammatica del Ladin Fascian*. Vich / Vigo di Fassa: Istitut Cultural Ladin 'majon di fascegn'.
- Cinque G. (1999). *Adverbs and functional heads: A cross-linguistics perspective*. Oxford: Oxford University Press.
- Cornips L. & Poletto C. (2005). « On standardizing syntactic elicitation techniques (part 1) ». *Lingua* 115: 939-957.

*Investigating Interrogation in the Northern Italian Area*

- Gartner T. (1883). *Raetoromanische Grammatik*. Heilbronn: Henninger.
- Hack F.M. (2007). *Das Rätoromanische und der Nullsubjektparameter*. Unpublished BA Thesis. Konstanz: Universität Konstanz, Fachbereich Sprachwissenschaft.
- Hack F.M. (2009). *Syntactic variation in interrogatives in Romance languages and dialects*. Unpublished MA Thesis. Konstanz: Universität Konstanz, Fachbereich Sprachwissenschaft.
- Hack F.M. (subm.). « The particle *po* in the varieties of Dolomitic Ladin: grammaticalization from a temporal adverb into an interrogative marker », *submitted to Studia Linguistica*.
- Hack F.M. (in prep.). *Variation in interrogative clause typing – evidence from varieties of the Northern Italian area*. PhD dissertation. Oxford: University of Oxford, Faculty of Linguistics, Philology and Phonetics.
- Hack F.M. & Gaglia S. (2009). « The use of subject pronouns in Rhaeto-Romance. A contrastive study », in G.A. Kaiser & E.-M. Remberger (eds), *Proceedings of the Workshop « Null-Subjects, Expletives, and Locatives in Romance »*. Konstanz: Universität Konstanz, Fachbereich Sprachwissenschaft, 157-181. (= Arbeitspapier 123)
- Kaiser G.A. & Hack F.M. (subm.). « Sujets et sujets nuls en romanche », *submitted to M. Iliescu; H. Siller-Runggaldier & P. Danler (eds), Actes du XXVe Congrès International de Linguistique et Philologie Romanes, Innsbruck, 3-8 septembre 2007*. Tübingen: Niemeyer.
- Kristol A. M. (1998). « Die historische Klassifikation der Romania. III. Rätoromanisch », in: G. Holtus, M. Metzeltin & C. Schmitt (eds), *Lexikon der Romanistischen Linguistik*. Vol. 7. Kontakt, Migration und Kunstsprachen, Tübingen: Niemeyer, 937-948.
- Manzini M.R. & Savoia L. (2005). *I dialetti italiani e romanci. Morfosintassi generativa*. 3. vol. Alessandria: Edizioni dell'Orso.

- Pellegrini G. (1972). *Saggi sul ladino dolomitico e sul friulano*. Bari: Adriatica Editrice.
- Poletto C. (2000). *The Higher Functional Field. Evidence from Northern Italian Dialects*. Oxford: Oxford University Press.
- Poletto C. & Zanuttini R. (2003). « Making Imperatives: Evidence from Central Rhaeto-Romance », in C. Tortora (ed.), *The Syntax of Italian Dialects*. Oxford: Oxford University Press, 175-206.
- Renzi L. & Vanelli L. (1983). « I pronomi soggetto in alcune varietà romanze », in P. Benincà, M. Cortelazzo, A.L. Prodocimi, L. Vanelli & A. Zamboni (eds), *Scritti Linguistici in Onore di Giovan Battista Pellegrini*. Vol. 2. Pisa: Pacini, 121-145.
- Thiele S. (2001). *Gadertalische und grödnerische Pronominalsyntax*. Münster: Nodus Publikationen.
- Vian J.A. (1864). *Gröden, der Grödner und seine Sprache. Von einem Einheimischen*. Bozen: Wohlgemuth'sche Buchdruckerei.
- Videsott P. (2009). « Ladinistische Forschungsprojekte an der Freien Universität Bozen », in V. Lyding (ed.), *LULCL II 2008. Proceedings of the Second Colloquium on Lesser Used Languages and Computer Linguistics (LULCL II), Bozen-Bolzano, 13th-14th November 2008*, Bozen: European Academy Bozen-Bolzano, 11-24.