Language Standardisation and Linguistic Resources: the Case of Central Ladin (Dolomites)

Fabio Ciocchetti* and Fabio Pianesi‡

*Istitut Ladin, Vigo di Fassa, 38 (Trento), Italy
istladin@relay.tqs.it

‡IRST, 38050, Povo, Trento, Italy
pianesi@irst.itc.it

In this paper we describe the efforts and actions undertaken to promote the development of a unifying standard for the Ladin language spoken in the Dolomites (Italian Alps), called *Central Ladin*, within the SPELL project.[1] A great part of SPELL has been, and is currently devoted to the development of linguistic resources for the local communities and institutions. We will start by discussing the main geo- and socio-cultural issues involved, §1, then illustrate the actions taken within SPELL towards standardisation, §2, and the development of linguistic resources for the standard language, §3 and §4.

1. Geo- and Socio-Linguistic Background

The Ladins of Dolomites form a small community located in five valleys of the western Alps, entirely within the Italian borders. The inhabintants speak a Romance language, called *Ladin* or *Rhetoromance*, which is related, on the one hand to northern Italian (Gallo-Romance) dialects, and on the other to Western Romance languages such as French, Occitan, Catalan, etc...

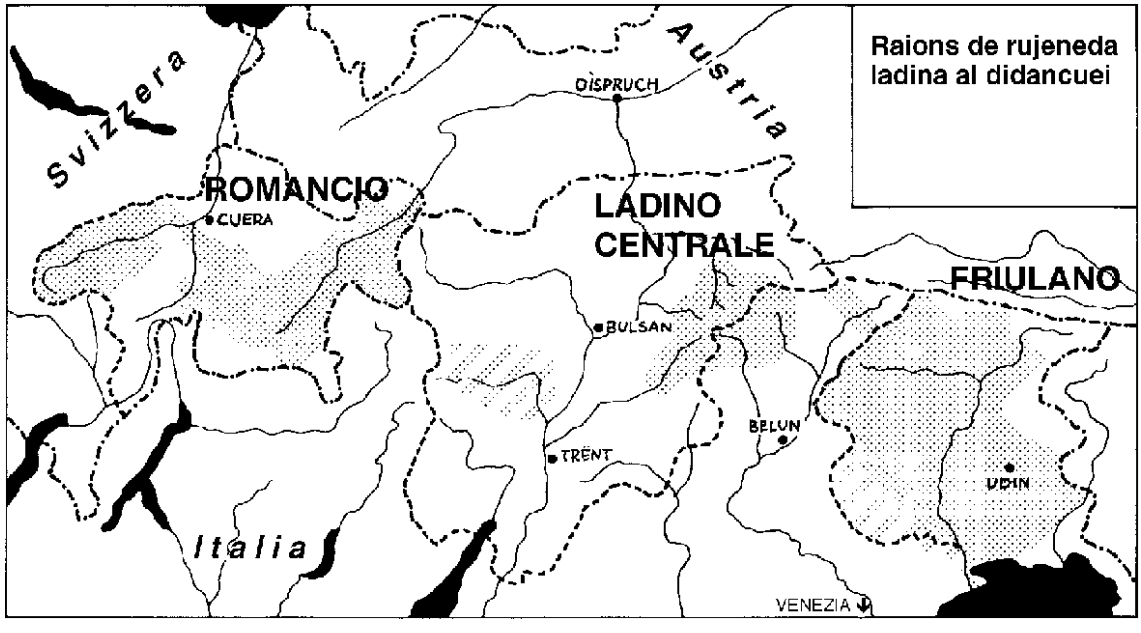

fig.1

[1] The SPELL project has been supported by the CEC, DGXII, Programme for Regional and Minority Languages

Besides this area, Ladin languages are spoken in two more places: the Swiss canton of Grigioni (**Romancio**), and Friuli (**Friulan**), in the north-eastern part of Italy (**Friulan**), see fig. 1. Ladin as spoken in the Dolomites, henceforth *Central Ladin*, can be distinguished into a number of local varieties, roughly corresponding to the main valleys: Badiot (Val Badia), Fascian (Val di Fassa), Fodom (Livinallongo), Ampëzzan (Ampezzo), Gherdëina (Val Gardena - Grontal).

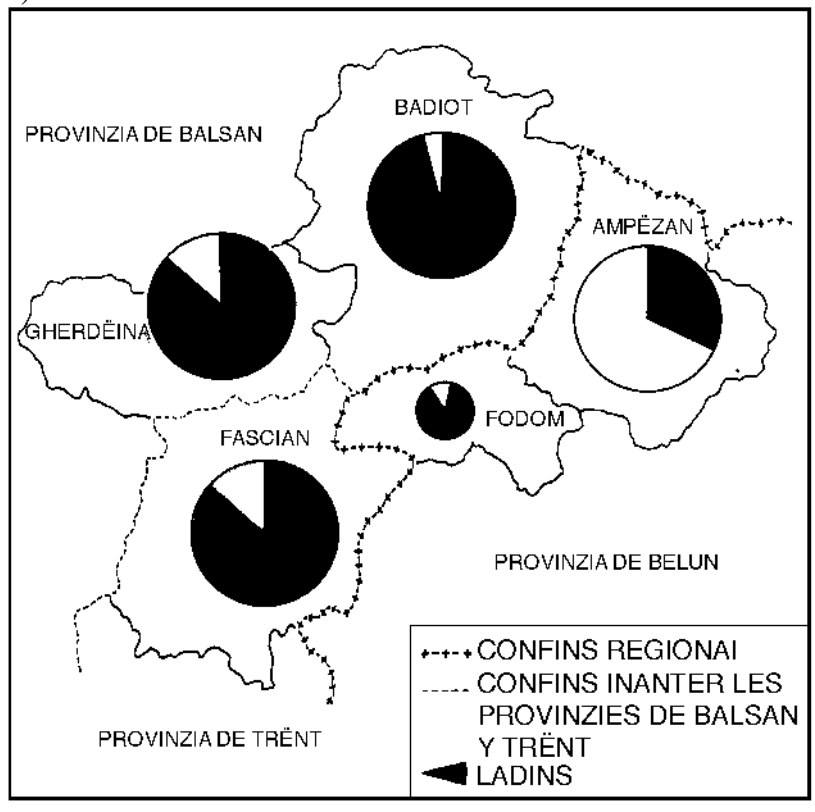

fig. 2

Such local varieties show important linguistic differences, which do not favour the use of any of them beyond its natural, geographical boundaries.

The present situation is the product of a number of geographical and geo-political factors, such as:

- a narrow community of speakers (approx. 30.000);
- the lack of a chief town for the whole area;
- a hilly zone with most built-up areas above 1200 asl;
- difficult communications due to the orographic structure.

From the socio-linguistic point of view, the following properties are noteworth:

- lack of a predominant variety;
- lack of a literary *koiné*;
- lack of extensive contacts with the other Ladin varieties (Romancio and Friulan), the only contacts being limited to few members of the elites;
- the strong competition of the two confining languages, German and Italian, which produce different and diverging effects on valley varieties.

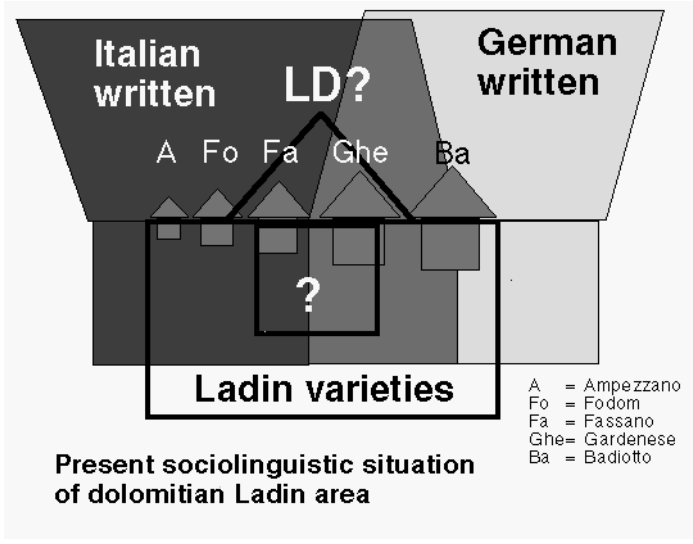

fig. 3

Such a situation is made even more complex by the administrative setup. The Ladin areas are included within three provinces (Trento, Bolzano and Belluno) and two regions (Trentino - Süd Tirol, and Veneto). As a consequence we have:

- a substantial dishomogeity of the giuridical status of Ladin language, which is protected in the Provinica of Bolzano, less so in the Provincia of Trento, and not protected at all in Veneto;
- a further loosening of inter-valley communications;
- social and institutional relationships shaped and oriented towards differing models.

On the other hand, the giuridical status of Ladin in the two provinces of Bolzano and Trento acknowledges Ladins the status of third linguistic group, on a substantial pair with the two other major groups, Italians and Germans. As a consequence, Central Ladin is present in a number of socially and administratively relevant situations:

- Ladin teaching is now a reality in the schools of Val Gardena and Val Badia, though with a limited number of hours. Furthermore, it is giuridically possible in Val di Fassa;
- concerning media, there is a weekly magazine ("La Usc di Ladins", 3.000 copies), and a weekly insert in three local newspapers. Furthermore, television programmes in Ladin are regularly produced by RAI (the Italian public broadcasting company), though with a highly restricted quantity of hours (two hours per month, approx.);
- there is a substantial increase in the use of Ladin in masses and other religious services and cerimonies, especially in Val Badia;
- in the public administration, Ladin is an official language in Val Gardena and Val Badia since 1990. Recently, it has been introduced in Val di Fassa too.

To further improve the use of Central Ladin, it is clearly necessary to create an acknowledged standard for the language, and to provide linguistic resources that can support the increasing needs of the local communities and administrations in all the relevant sectors.

In these respects, the situation by the end of the 80's can be characterised as follows:

- the absence of a widely acknowledged standard for Ladin;
- a yet to be completed process of ortographical unification;
- a too restricted lexicon to cope with the new contexts of use, registers etc..;
- the lack of institutions capable of producing the relevant standards and the relevant linguistic resources (lexica, corpora, terminologies, etc..).

According to Kloss' proposal (1976), the situation of written Ladin by the end of the 80's could be described as that of a language covering the first three degres of development (Ladin and general cultural subjects, natural sciences) at an elementary school level. Works addressing higher levels were, and in part still are rare.

F	=	=	=
G	4	(5)	(6)
V	1	2	3
	E	K	N

Application fields:

E = Ladin culture
K = general culture
N = natural sciences

Degree of development:

V = elementary school
G = highschool
F = scientific prose

The purpose of this paper is to describe a number of action that had been taken towards the production of a uniform standard for Central Ladin, and towards the production of a series of linguistic resources for both the new standard and the existing varieties.

2. The standardisation process.

The problem of providing a unifying standard for Central Ladin could be solved by following either of two different methodologies:

a) that adopted for Friulan, based on the choice of an existing literary variety (**Central Friulan**);
b) that followed with Romancio, which aimed at building a new communicative code from the existing local varieties.

As observed, no existing variety is presently available with an acknoledged prestige which can be used to provide the core of the standard. Therefore, the first option had to be discarded. On the other hand, the second option had already been successfully adopted for Romancio in a situation which was similar to that described for Central Ladin. We therefore proceeded to elaborate the basic principles for the construction process (Schmid, 1994). The resulting criteria for the lexicon can be summarised as follows:

a) frequency: preference is to be given to the most frequent forms among the varieties (cf. the forms *frut* and *neif* in fig. 4);
b) sistematicity: forms are given preference which enhance the regularity and coherence of the whole system (cf. the forms *peir* and *ceina* in fig. 4);
c) transparency: preference is given to "full" forms, more readly comprehensible than shortened ones (cf. the forms *plovei* and *scrive* in fig. 4);
d) typicality: forms are chosen which distinguish Ladin from competing languages (cf. the forms *gial* and *blanch* in fig. 4).

	lat.	mar./bad.	grd.	fas.	fod.	amp.	**LD**
	FRUCTUM NIVEM	früt nei/nëi	frut nëif	frut neif	frut nei	fruto gnée	**frut** **neif**
	PIRUM CENA	per/për cena/cëna	pëir cëina	peir cena	peier cëna	pero zena	**peir** **ceina**
	°PLOVERE SCRIBERE	ploei/plovëi scrí/scrí	pluëi scrì	piever scriver	pluove scrive	pioe scrie	**plovei** **scrive**
	GALLUM °BLANCUM	ial/gial blanch	gial blanch	gial, ial bianch	gial blânch	jal bianco	**gial** **blanch**

fig. 4

It is important to remark that our project aimed at providing a further variety of Ladin (called *Ladin Dolomitan*, LD), which is to be obtained by appropriately assembling material from presently spoken idioms; furthermore, it does so by aiming at a higher degree of coherence and unity of the overall system. For these reasons, Ladin Dolomitan can play the role of a standard, higher variety by

- functioning as a reference for ortographic norms and for the production of terminologies, even in local varieties;
- functioning as the prevailing media in the written exchanges between valleys (for the normal speaker, LD is more comprehensible than many of the existing varieties);
- being capable of promoting the prestige of Ladin as a whole;
- promoting the self-identification of the community in its relationships with the external world, especially at the administrative level;
- being the media used in commercial exchanges, addvertisement, etc.., and, more generally, in all the cases of functional communication where the existence of different varieties of limited diffusion hampers the use of Ladin.

In other words, Ladin Dolomitan is not meant to supplant the existing varieties but to complement them in all the functions, situations, etc.. which they cannot cope with.

3. The SPELL project.

Once the basic theoretical principles for the elaboration of the standards had been developed, the problem had to be addressed of creating suitable infrastructures capable of supporting the actual development of the standard, the production of the necessary linguistic resources, the theoretical elaboration and update of the basic paradigm, and of functioning as the coordinating agency of the whole process. To this end, in 1993 the major cultural institution devoted to Central Ladin launched a common project called **SPELL - Servizio di Pianificazione ed Elaborazione della Lingua Ladina** (Planning and Elaboration Service for the Ladin Language), see Chiocchetti et al. (1994). The SPELL working group, articulated in a number of small centers scattered on the territory, started in November 1994.

Among the goals of the SPELL working group was the design and construction of a data base collecting all the linguistic resources available for the different Ladin varieties. The purposes were:

- the construction of tools for encoding local varieties;
- the construction of basic dictionaries and grammars for the standard LD;
- planning and implementing a project for terminologies.

Given the purposes of the standard Ladin Dolomitan, it seemed important that the data base management system:

- provided tools for a fruitfull, bidirectional exchange between the "living" varieties and the standard, thereby securing a progressive and constant enrichment of the latter by drawing on the lexical resources of the former, and, conversely, improving the expressive capabilities of local varieties by resorting to the standard lexica of LD;
- aimed at enhancing the impact of resources on local institutions by providing links to/from Italian and German lexical entries;
- featured a user-friendly interface;
- supported distributed, network job execution.

For these reasons, the SPEEL working group chose a HyperCard based system (Chiocchetti & Vögeli, 1997) which had already been widely and successfully exploited for the similar project concerning Romancio (Bernardi et al., 1994).

4. The SPELL Data Base

At present, the SPELL Data Base consists of a complete collection of all the Ladin lexicographical material available both in printed form or from handwritten sources. It amounts to approximatively 200.000 entries, corresponding to approximatively 80.000 lemmas, organised at various levels.

Given the sources of our material, and their importance to reconstruct the history of each single lemma, the whole data base has been conceived to provide appropriate information and pointers to the basic documentation it was built from. This also means that the SPELL group had to define and produce appropriate fonts, with diacritics, etc.. to allow the comparison among items belonging to different local varieties, often available only through historically different writing systems.

At present, the SPELL Data Base consists of:

I. Data concerning local varieties (6 data bases, 85.700 entries)

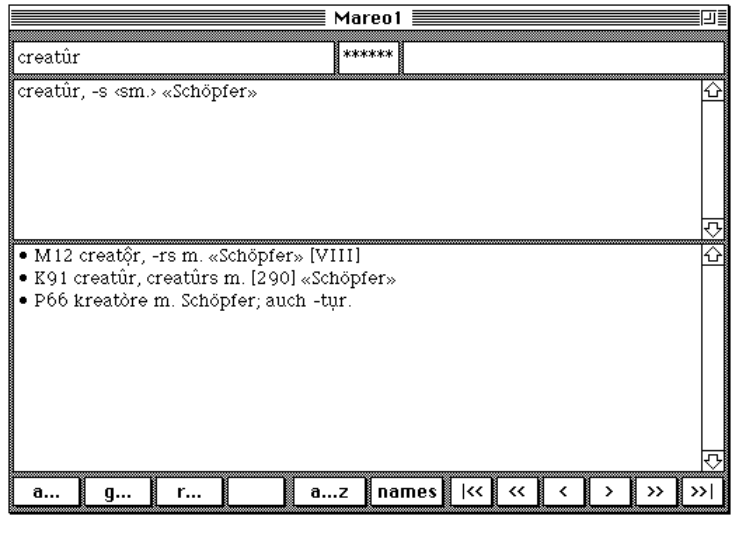

fig. 5

Each *data base I* is devoted to a local Central Ladin variety. The entries are mainly meant to provide information about the sources, which are collected in an appropriate field (the lower field in the figure above).

II. Temporary data bases concerning local varieties (4 archivi: 50.000 voci)

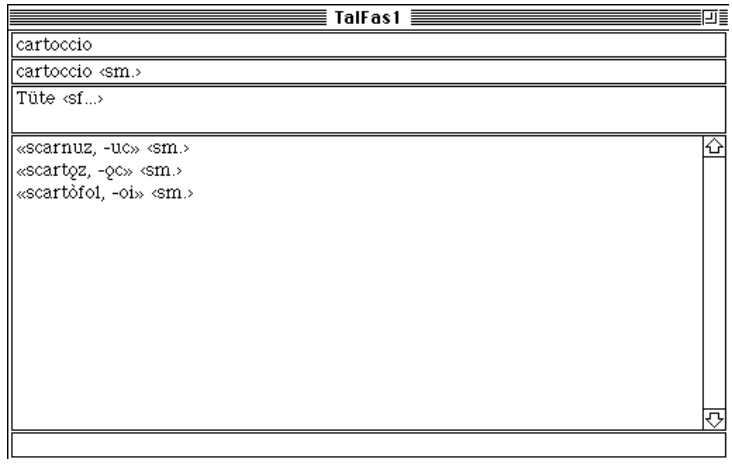

fig. 6

This data bases are drawn from available lexicographical sources. Each entry, besides reproducing the lemma for a local variety, is indexed by means of the corresponding Italian and/or German lemma. This information is used to produce normative lexica for local varieties and can be compared and/or integrated by material drown from *data bases I* .

III. The data base of Ladin Dolimitan *(10.000 voci)*

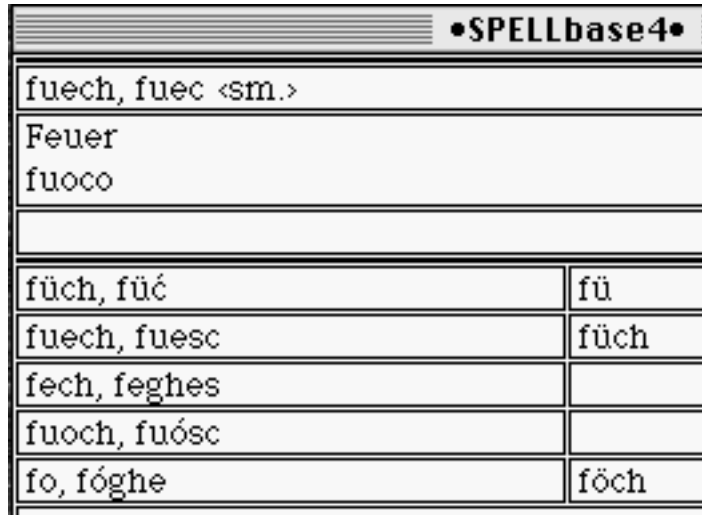

fig. 7

This is the major data base. It supports the comparison among lemmas from local varieties (drawn from *data bases I* and *II*) for the purposes of the standardisation process. Hence, each

entry alligns local variety lemmas, and lemmas for the two reference languages (Italian and German). The main field contains the lemma for the standard LD, obtained by applying the metholodgies and criteria described in §2. Separate fields deal with reverse indexing and cross-references.

The fields for local varieties are automatically filled by procedures accessing *data bases I* and *II*. The lexicographer can postedit the results.

IV. Special data bases for LD (4 data bases: 16.200 entries)

fig. 8

They consist of three terminological data bases (alpinism, administration, commercial law; see fig. 8) for LD, and one data base for phrasal idioms (fig. 9). The entries of the terminological data bases allign lemmas in a number of european language, lemmas for local varieties of Ladin (when available) and the LD lemma. We are currently planning a new biological-medical terminology.

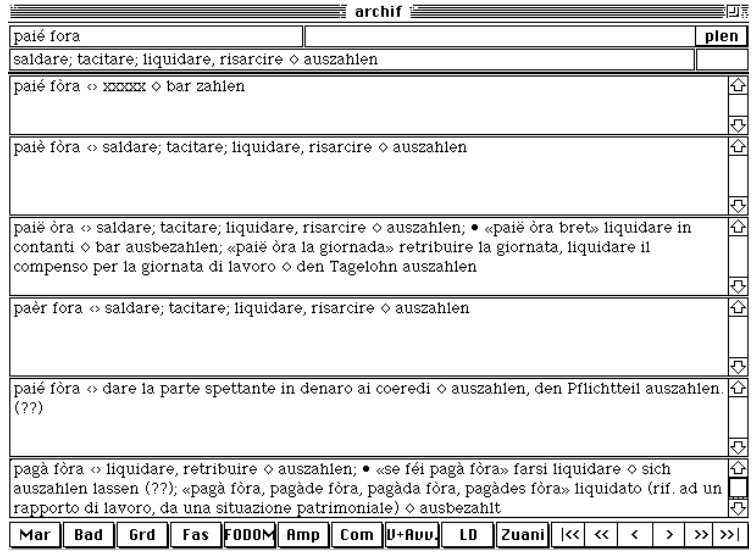

fig. 9

V. Reference data base for Italian and German (27.500 entries)

It is used to build the LD lexicon (*data base III* above). It is also used to produce small terminological dictionaries for the local varieties of Ladin, which are then made available to local institutions (schools, newspapers, administrations, etc..)

The advantages of the whole architecture are:

- simple and intuitive query system;
- user-friendly interactions;
- simple software tools can be easily implemented to manage automatic updating, consistency checking, the production (in printed form) of small terminologies etc..;
- easy maintainance and reduced costs;

The disadvantages are:

- strong limitation in the amount of data that each data base can handle (no more than 6000/8000 entries). This requires a fragmentation of the data bases that produces further problem with retrival, access time, ecc..
- as a data bases management system HyperTalk has very limited capabilities. Thus, it is impossible to submit real queries, form dynamic links between different entries (even within the same data base) and so on.

5. Conclusion.

In this paper we have described the project SPELL, whose aims are to elaborate a unifying standard for Central Ladin (LD) and provide for the linguistic resources supporting the circulation and diffusion of the LD among the local institutions.

In particular, we have described the theoretical and methodological principles underlying the standardisation process and the current status of the acitivities in the field of linguistic resources. Concerning the latter, the SPELL project has made available a complete collection of all lexicographical resources from known printed and hand-written sources, divided according to the local variety they belong to. From this, standard (LD) forms were produced which are now available to all the major local insistutions and administrations. Moreover, a number of termonological lexica have been provided which deal with domains of major interest, such as commercial law, alpinism, etc.. In the end, it has been possible to trigger a process in which the production of linguistic resources supports the development and the adoption of the standard LD, and viceversa.

As already observed, despite the achievements, linguistic resources still need to be improved, especially from the point of view of the information systems managing them. Furthermore, the need is arising of other kinds of linguistic resources, such as grammars, morphological data bases, etc... Finally, we can start addressing the problem of software resources exploiting linguistic ones.

Concerning the first two points we are planning a major revision not only of the software tools, but also of the whole conception of the various lexica, to improve their usability, flexibility, and richeness of content. We are also planning activities towards grammatical and morphological resources, drawing on the standardisation work done (Videsott, 1997; Valentini, 1997).

Concerning the third point, we are planning support systems for users in local insistutions (administrations, newspapers); in particular, we are working at a spelling corrector and towards a translation-support system meant to help the user in translating from each local variety to the standard.

Bibliography

Bernardi, R., A. Decurtinis, W. Eichenhofer, U. Saluz & M. Vögeli, (1994), *Handwörterbuch des Rätoromanischen*. Societa' Retorumantscha & Verein für Bündner Kulturförschung, Koira, Sw.

Chiocchetti F., R. Bernardi, N. Chiocchetti & P. Videsott, (1994), "*Language Planning" ed Elaborazione della Lingua - Un progetto per lo Sviluppo del Ladino delle Dolomiti*. Istitut Cultural Ladin, Vigo di Fassa, It.

Chiocchetti, N. & M. Vögeli, (1997), *Linguistiche Dataverarbeitung: Probleme und Lösungen*. SPELL/Istitut Cultural Ladin, Vigo di Fassa, It.

Kloss, H., (1976), 'Abstandsprachen und Ausbausprachen'. In J. Göschel, N. Nail, & G. Van der Elst (eds.), *Zur Theorie des Dialekts. Aufsätze aus 100 Jahren Forschung. Mit biographischen Annerkungen zu den Autoren*. ZDL, Beihefte, N.F., 16, Wiesbaden.

Schmid, H. , (1985), 'Rumantsch Grischun - eine Schriftsprache für ganz Romanischbünden'. In *Ladinia*, 9, 171-202.

Schmid, H., (1994), *Wegleitung für den Aufbau einer gemeinsamen Schriftsprache der Dolomitenladiner*. ms. SPELL/Istitut Cultural Ladin, Vigo di Fassa, It.

SPELL, (1997a), "*Language Planning" ed Elaborazione della Lingua - Un progetto per lo Sviluppo del Ladino delle Dolomiti- Seconda Fase 1997-1999*. Istitut Cultural Ladin, Vigo di Fassa, It.

Valentini, D., (1997), *Curs de Ladin Standard - Pruma Verscion*. SPELL/Istitut Cultural Ladin, Vigo di Fassa, It.

Videsott, P. (1997), *Gramatica Elementara dl Ladin Dolomitan - III Versione*. SPELL/Istitut Cultural Ladin, Vigo di Fassa, It.