

AIS, *reloaded*: A digital dialect atlas of Italy and southern Switzerland*

1. Introduction

In this paper, we will introduce to the scientific community, for the first time in printed form, the output of the research project AIS, *reloaded* (AISr), which yielded the following achievements:

- (1) a. a searchable digital version of half of the maps of the *Sprach- und Sachatlas Italiens und der Südschweiz* (*Atlante Italo-Svizzero*, AIS, 1928-1940), whose data were collected in 407 survey points in the two countries;
- b. a comparable corpus for the Italo-Romance and Romansh dialects of southern Switzerland (18 points each), corresponding to the original datapoints, consisting of new data that have been collected using the same questionnaire in the same localities almost one hundred years later.

The paper is organized as follows. In §2, we provide basic information on the project, which gave rise to the interactive database freely searchable at the website <https://www.ais-reloaded.uzh.ch>, and present the structure and functionalities of the latter. In §3, we exemplify the ways in which the AISr database can be used to compare subsequent diachronic stages of the same dialects, first (§3.1) discussing some changes – mostly leading towards standardization – which are observed in western Surselva. We then consider changes in the areal distribution of some kinship terms (§3.2), the selection of a gender agreement marker in the dialect of Mesocco (§3.3) and the syntax of negation in Swiss-Lombard dialects (§3.4). Finally, §3.5 addresses (the lexical diffusion of) a sound change in Engadine. By way of conclusion, §4 outlines prospects for further research along this line.

* The present article is part of the research project “AIS, *reloaded*” (SNSF grant 100012-162482), carried out at the Romanisches Seminar of the University of Zurich. Funding by the Swiss National Science Foundation (SNSF) for the period 2016-2019 is gratefully acknowledged. Thanks to Daniel Wanitsch for designing the database, to Claudia Cathomas and Dominique Dosch for their cooperation in the fieldwork in the Grisons, as well as to the Centro di Dialettologia e di Etnografia (CDE) in Bellinzona and the Institut dal Dicziunari Rumantsch Grischun (IDRG) in Chur, directed by Franco Lurà, Paolo Ostinelli and Carli Tomaschett, respectively, for invaluable help with fieldwork planning. The paper was conceived and realized jointly. Nevertheless, for academic purposes, responsibility must be divided as follows: ML and SS §§1, 2 and 2.3.2, ML §4, CZ §§2.1-2.2, DP §3.3, GD and DP §3.4, SN §§3, 3.1-3.2, 3.5, GT §2.3.1.

2. From AIS to AISr

The 1,705 AIS maps provide a comprehensive picture of 407 distinct dialects as spoken between 1919 and 1928, when the fieldwork was carried out under the supervision of Karl Jaberg (1877-1958), professor of Romance linguistics in Bern, and Jakob Jud (1882-1952), his counterpart in Zurich, thanks to funding by the Stiftung für wissenschaftliche Forschung an der Universität Zürich. The fieldwork was conducted by Paul Scheuermeier (1888-1973), who was in charge of Switzerland as well as – from July 1920 on (see Kunz 2018, 32) – northern and central Italy; Gerhard Rohlfs (1892-1986), who was in charge of southern Italy; and Max Leopold Wagner (1880-1962), who surveyed Sardinia. The survey began in Ardez (Grisons, pt. 7) on 19-23 November 1919 and was concluded through Rohlfs's field inquiry in Polistena (Calabria, pt. 783) on 17-19 October 1928 (Jaberg / Jud 1928, 40, 129).

In the framework of AISr, which ran from 2016 to 2019, new data from the 36 AIS datapoints in southern Switzerland (18 for the Lombard varieties in Canton Tessin and in the Italian-speaking part of the Canton of Grisons, and 18 for the Romansh dialects in the Canton of Grisons) were collected, recorded and transcribed (about 61,000 entries in all, for a total of about 400 recording hours)¹. For each datapoint, an informant was chosen and interviewed using a selection from the approximately 2,000 entries of the AIS intermediate questionnaire (1,728 in the Canton Tessin and the Italian part of Grisons; 1,800 in Canton Grisons):

(2) AISr interviews (#36). Informants:

Age:	70 years (average)	Oldest informant 89 y/o	Youngest informant 36 y/o
Gender:	28 men	4 women	

Contrary to AIS, answers were recorded digitally, the difference obviously lying not in the digital format but in the audio recording itself: recording techniques were in their infancy in the 1920s, and mounting a phonographic recording campaign would not have been feasible with the AIS time resources (three days/one transcriber per point)².

¹ The fieldwork was carried out by the Zurich-based coauthors (though ML and SS took part only for a few points), with the cooperation of Claudia Cathomas, Dominique Dosch and Alberto Giudici for several Romansh datapoints, and Camilla Bernardasci for Olivone and Sonogno. The first point to be visited was Corticiasca, pt. 73 (by DP, GD and SN on 15-16 February 2017), while the last round of fieldwork was carried out by CZ, GD and SN in Scharans (pt. 16) on 19-20 August 2019.

² In fact, as supervisor on behalf of the Zurich Phonogram Archives, Jud was involved in the phonographic dialect recordings of southern Switzerland dialects which took place in Bellinzona on 14-17 October 1929, exactly one year after the conclusion of the AIS fieldwork. This was the second recording campaign ever in southern Switzerland, after Carlo Salvioni's one in 1913 (see Loporcaro 2008, 53 n. 34, 82 n. 129). In the early 20th century, however, such recordings were extremely onerous – for one thing, they required the presence of an expert in the phonographic technique. In the specific case involving Jud in 1929, the colleague

The recorded data have now been made available (both IPA-transcribed and as audio files) in a freely searchable database whose contents can be downloaded. This database also hosts a searchable version of 50% of the transcriptions contained in the AIS (maps 1-880, or about 700,000 entries in all; see fn. 4 on these figures), which have been digitized in the original AIS transcriptions thanks to a collaboration with our research partner Dr. Graziano Tisato (Padua), who developed a customized OCR system to this effect, which was run on the digitized images already online at the NavigAIS website he completed in 2009 through a high-definition acquisition of the map images³. Furthermore, the transcribed materials from maps 1 to 880 have been converted automatically into IPA transcriptions, which can be downloaded along with the original ones⁴.

The limitation to 50% of the AIS maps ((1a)) and to southern Switzerland ((1b)) was imposed by resource and time constraints within the three-year funding period. For the same reasons, the digitization process targeted lemmas only, thus excluding the ethnographic information, which will be digitized in the continuation of the project (2021-2025; see §4).

2.1. *The AISr database*

In what follows, we will introduce the AISr database, which allows users to directly compare and search the two corpora for content with the limitations already mentioned: for the AIS data, apart from the obvious unavailability of audio recordings, as of now the corpus covers 50% of the maps (though see (19a) below), but, unlike the AISr data, which – for southern Switzerland – cover all maps, it also covers Italy. For southern Switzerland, the AISr data allow us to compare the two datasets collected at a 100-year interval and hence to observe any changes that occurred in the elapsed century.

operating the machine, Wilhelm Doegen, had to travel from Berlin (where he directed the *Lautabteilung* of the Preussische Staatsbibliothek); see Bernardasci/Schwarzenbach (2016, 24), Loporcaro (2016, 559).

³ The quantum leap in access to the AIS data, as compared to NavigAIS – where material can be searched by map and datapoint but not exported as text – is represented by the searchability of contents and their exportability in the form of phonetic transcriptions (see §2.2).

⁴ The attentive reader will have noticed an apparent contradiction between the figures just reported for the number of recorded items relative to the AIS and AISr data. The figure of approximately 61,000 entries for the latter corresponds to approximately 1,600 entries per point on average: the difference with respect to the number of questionnaire items (1,728 or 1,800) is due to lexical decay, as some answers were not given by informants. As for the AIS data, the figure of approximately 700,000, giving on average 3,440 entries per point, results from the fact that each item is available in the database in two different transcriptions (AIS vs. IPA), such that it is counted twice.

2.2. *The search, retrieve and display options*

The database offers multiple search options, which are commented on in §§2.2.1-2.2.6, with each heading corresponding to those appearing in the navigation bar.

2.2.1. Maps. – One possibility is to search by map, typing its number or name (e.g., 5 or *padre* ‘father’) in the “search” field or by ticking one box in the corresponding drop-down menu. It is also possible to select a range of maps by ticking more than one box in the drop-down menu. Alternatively, it is possible to type in the “search” field the numbers (e.g., 2; 6; 14) or the range (e.g., 5:20) corresponding to the desired maps. In the drop-down menu, uppercase strings (e.g., IL PADRE ‘the father’) identify regular maps; lowercase strings (e.g., *la gioventù* ‘youth’ or *trovare* ‘to find’) identify further materials (legends alongside a map, verb conjugations in vol. 8, etc.). Visualization of the output data can be selected either in table or in map format, the latter only when one single map is selected (§2.2.6).

2.2.2. Locations. – Another possibility is to query the database by selecting a locality. Once again, users can select a datapoint by typing its number or name (e.g., 1 or Brigels) in the “search” field or by ticking one box in the corresponding drop-down menu. Users can also select all datapoints by ticking the “select all” option in the drop-down menu. In addition, it is possible to select a subset of datapoints by ticking more than one box in the drop-down menu or by ticking the first and the last datapoint of a desired range while holding down the “shift” key. Alternatively, it is possible to type into the “search” box the numbers (e.g., 3; 7; 14) or the range (e.g., 5:20) corresponding to the desired datapoints.

2.2.3. Period. – The database also allows users to separately query either the original AIS data by selecting the option “past” or the newly collected AISr data by selecting the option “present”. Selecting the option “all” results in querying both datasets at once. Users can also visualize the AIS data only in their original transcription by selecting the “AIS” option or only in the current standard transcription by selecting the “IPA” option. By default, the two types of transcription are displayed synoptically in the downloadable table (but not on the map; see §2.2.6) that summarizes the query results. As for the newly collected AISr data, only IPA phonetic transcriptions are available.

2.2.4. Phonetic. – After selecting the previous fields (“Maps”, “Locations”, “Period”), it is possible to further restrict a query to only those entries that contain a certain phone or string of phones by specifying the context of occurrence of the desired phone(s) in the entry by using the underscore symbol “_”: at the beginning of an entry (e.g., *la_* in *lana* ‘wool’), at the end of an entry (e.g., *_l*), in the body of an entry (e.g., *_l_*). It is worth noting that the level targeted by the search programme is not the word level, but the questionnaire entry level (namely, the whole phonetic string). Therefore, whenever the entry consists of a phrase, rather than a single word, a query like *l_* also returns entries such as *la mia cugina* ‘my (female) cousin’.

Users can also specify the exact position occupied by a given phone in the entry by typing it preceded by as many hash symbols “#” as the elements that precede it (e.g., #*a* to retrieve all occurrences of the phone [a] whenever preceded by just another phone, or ##*a* to retrieve all occurrences of the phone [a] when it is preceded by two phones, etc.). Note that this search targets all the elements that precede the searched phone and, therefore, not only other phones but also diacritics such as stress (') and length (:) marks, as well as the space between words when the entry is a phrase. For example, a query can target the tonic vowel /a/ in the second position by typing ##*a* in the case of entries consisting of a single word such as /b'a:p/ ‘father’, or #####*a* in the case of entries made of a phrase such as /il b'a:p/.

2.2.5. Download table. – Query results are immediately displayed, depending on the option chosen, either as a table, as shown in Figure 1, or as a map (§2.2.6).

ID	transcription	IPA	period	location	map details
1	4391 al ddt	la-to	el d et		156 153 12 4 M
3	4602 al pdy	el p n	/service/mepast		167 163 12 13 M
4	4404 al kalgp	el kalgp	/service/mepast		169 165 12 15 M
5	4406 al flit	el fl a:t	/service/mepast		171 167 9 14 M
6	4652 al flvli	el fr avl	/service/mepast		217 213 84 13 M
7	4682 al tasy	el t asj	/service/mepast		456 436 53 12 M
8	4699 al rdp	el r dp	/service/mepast		476 455 54 8 M
9	4724 al balon	el balon	/service/mepast		507 485 55 9 M
10	4738 al kuku	el k uku	/service/mepast		529 505 58 13 M
11	4752 al kacardir	el kacardir	/service/mepast		544 518 52 12 M
12	354919	il flat	/service/mepresent		171 167 9 14 M
13	355289	il dci le det	/service/mepresent		156 153 12 4 M
14	355502	il paj	/service/mepresent		167 163 12 13 M
15	355535	il kart logt	/service/mepresent		169 165 12 15 M
16	365093	il kart dte	/service/mepresent		544 518 52 12 M
17	365297	il tajs	/service/mepresent		456 436 53 12 M
18	365448	il rde	/service/mepresent		476 455 54 8 M
19	365793	il fe tagn, ts	/service/mepresent		507 485 55 9 M
20	366585	il kuku	/service/mepresent		529 505 58 13 M
21	375531	il tvari	/service/mepresent		217 213 84 13 M

Figure 1. Results of the query “153 – IL DITO; LE DITA / 163 – IL PIEDE; I PIEDI / 165 – IL CALCAGNO / 167 – IL FIATO / 213 – IL FABBRO / 436 – IL TASSO / 455 – IL ROSPO / 485 – IL RAGNO; I RAGNI / 505 – IL CUCULO / 518 – IL CACCIATORE” for pt. “10 – CAMISCHOLAS (TAVETSCH)” (table display)

A percent “%” symbol may appear at the end of a transcription of newly collected data. This indicates that the informant, while lacking active competence of the relevant word, nevertheless recognizes it as part of his/her dialect if the interviewer suggests it (as *extrema ratio*) based on the original AIS data collected *in loco*.

AIS	IPA	Period	Location	Map
—	or 'fegato %	present	71 – Breno	139 – IL FEGATO
—	la 'bile %	present	71 – Breno	140 – IL FIELE
—	l om'beliko %	present	71 – Breno	130 – L'OMBELLICO (BOTTONE)

Figure 2. The “%” symbol indicating passive competence

Whenever an answer has been recorded, the corresponding audio files (.wav) are automatically included in the results. They can be listened to by clicking on the turquoise “play” symbol.

The searched results can be saved and exported by clicking on the “Download table” option: a folder is generated containing both the audio files (.wav) and the transcriptions (.csv).

2.2.6. Switch to map view. – Whenever a single map is selected, users can perform queries and then also visualize the query hits on a geo-referenced map by clicking on the top left “Switch to map view” option. In this case, if the option “Period – all” is selected, the AIS data appear in black type on the right-hand side (in the transcription selected), while the AISr responses appear in turquoise, accompanied by the “play” button allowing the user to retrieve and hear the recording. In the geo-referenced map (a detail is shown in Figure 3), the string consisting of transcribed data is always preceded by a red number, which identifies the corresponding datapoint (e.g., 1 = Brigels).

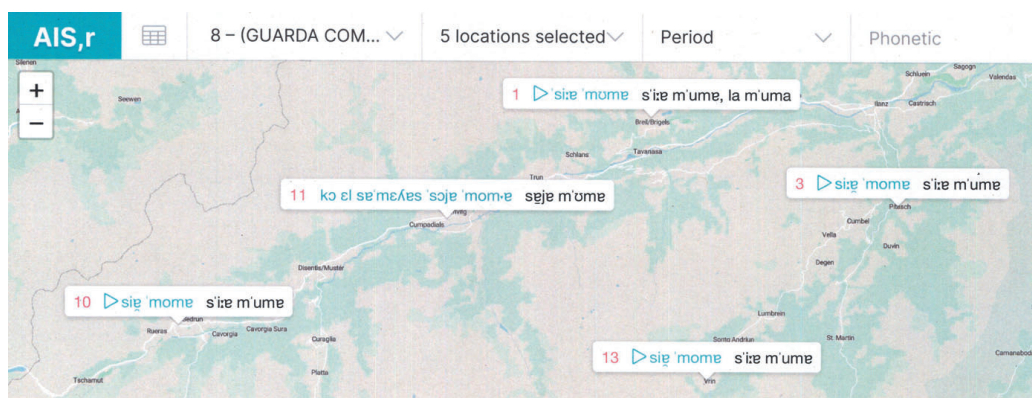


Figure 3. Map view of the query results

On the map, real data appear only if the user has zoomed in enough manually, whereas by default the map visualization starts out with the whole Italian peninsula and a series of turquoise circles on the different areas, each with a number in it indicating the sum total of points contained in the relevant area. Thus, at this default scale, southern Switzerland is lumped together with Lombardy and parts of the adjacent Alpine areas of Piedmont and Trentino, such that the corresponding circle bears the number “93” and no individual data are visualized. The latter appear, as explained, when zooming in.

2.3. *Behind the database: Digitization and phonetic transcriptions*

2.3.1. Digitizing the AIS transcriptions was all but an easy task, given their peculiarities, which preclude the use of common OCR (optical character recognition) engines. Firstly, the glyphs are handwritten by different people⁵, producing inter- and intraindividual graphic variation, and the italic font used, with a 70° inclination, causes neighbouring characters to overlap, reducing OCR performance. Secondly, transcriptions contain a large number of low-frequency symbols (~300) and of diacritics (25), which were devised to adequately express variation in the dialects considered. This makes the OCR approach with neural networks, deep learning, etc. extremely time-consuming due to the need to prepare thousands of samples for the required training. Thirdly, the dialects lack two facilities that common OCRs rely on in order to discriminate among possible candidates: a) a dictionary, listing all the legal words of the language(s) to be recognized; b) a language model, which could give the probability of the character chain (bigram, trigram) and word concatenation. The inevitable consequence of these AIS graphic features is that the performance of any probabilistic engine-based OCR will be diminished.

To successfully recognize a character, the OCR needs an algorithmic description of each glyph (e.g., by a polygonal approximation), independent of variation in size and shape, and of noisy and degraded conditions. The extracted feature vector sets the position of the unknown character in an n -dimensional space and its distance from all the training prototypes, identifying it as the closest template.

Omitting the details of OCR methodologies/technologies (involving intelligent character recognition, machine learning, pattern recognition, deep learning, etc.), let us just mention that the state of the art as of 2016, when the *AISr* project was started, is aptly represented by Vithlani / Kumbharana (2015). They report the results of a recognition test on a string of 36 characters (English, alphanumeric, capital, isolated), repeated ten times, handwritten by seven people of various ages. The string was processed by six stand-alone and online OCRs. The Custom OCR Online performed best, with a recognition accuracy of about 44%, or a CER (character error rate) of 56%, very far from the near 100% rate obtained on printed characters. Performance worsens significantly if common online OCRs are run on *AIS* maps, mainly because the AIS uses diacritics on seven distinct levels: this prevents current OCRs from correctly segmenting the text lines and results in the omission of parts of the glyph itself, or the inclusion of parts of the glyphs belonging to neighbouring data-points.

Thus, the task had to be broken into more manageable steps. In NavigAIS, the noisy original AIS maps were subjected to a preliminary six-step processing to improve visualization (rotating and cropping the maps, enhancing contrast, elimi-

⁵ The scholars involved were assisted by a number of other people. For instance, Jakob Jud was helped by his family; see the vivid description of copying work done by the whole family on Sunday afternoons at Jud's home in Zollikon, as provided by Schorta (1952, 12).

nating background noise, etc.) and, crucially, to separate the text (black) from the background (orange) into two distinct images. The text image, stripped of the geographic borders, is a suitable input to the final OCR step. The OCR engine used in AISr implements a decision tree. The AIS symbols are split into the base glyph, easily recognizable by a modified, trained version of the free Tesseract OCR, and into the diacritics, classified through a dedicated ruleset, according to their geometric properties, relative position, and possible glyph-diacritics correlation and compatibility.

The recognition procedure is programmed in order to sequentially (or manually) move the cursor at the coordinates on the selected datapoints. The software identifies the area in the vicinity of each datapoint containing the sequence to be recognized, removes the character lines touched by the area borders, and inputs the cleaned image to the OCR for the segmentation and recognition steps. In order to prevent overlapping, the character space is increased after segmentation. Human supervision is needed to resize, when necessary, the area identified, to check the OCR output sequence, correcting any errors. Finally, the software saves the sequence in an .xls spreadsheet and moves on to the next point.

An acquisition test on real data from 14 complete AIS maps (100,000 characters) obtained an average CER of 3.65% (ranging from 1.35% to 5.87%) based on Levenshtein distance (Levenshtein 1966), including any insertion, omission and replacement of characters and diacritics. The average acquisition time per datapoint was < 20 s (ranging from 13.63 to 25.31 s according to the number of characters), a much lower value compared to an estimated time of > 30 s for manual transcription. The acquisition time per map is about two hours, depending on map conditions and number of characters in the answers. This yields a net realization time of about 3,400 hours for the acquisition of the 1,705 maps.

2.3.2. Moving on from graphical to structural properties of phonetic transcriptions, one has to mention that the peculiarities of the AIS system imposed some conventions that would allow a unique digitization and subsequent automatic conversion into IPA characters:

- AIS characters written on top of other characters have been digitized as superscript characters in brackets (e.g., $\overset{u}{a} \rightarrow a^{(u)}$); in the downloadable .csv file, the overwritten and underwritten characters are marked respectively as ^{SYMBOL} and _{SYMBOL};
- the AIS symbol indicating weak nasalization (a curled tilde) was digitized using the diacritic \tilde{a} overwritten to a character (e.g., \tilde{a});
- palatal plosives transcribed in the AIS as \tilde{c} , \tilde{g} have been digitized here as \tilde{c} , \tilde{g} ;
- the underwritten diacritics $\underset{cc}{a}$ that marked “very open” vowels in the AIS have been digitized by means of a double vertical stroke under the character (e.g., $\underset{cc}{a}$).

As mentioned above, AIS data can also be retrieved from the database in the form of an automatically generated IPA transcription, which is the output of a conversion process whose design and implementation imposed a serious challenge on

the research team. Again, the problem is in part one of structure, in part one of performance. As far as structure is concerned, it will come as no surprise that the two transcription systems are not perfectly isomorphic. In addition, they had to be applied online by different researchers, without the aid of sound recording that could help objectify the result. The output is an implementation which, while reasonably uniform, contains some variation and noise, as recognized by Jaberg / Jud (1928, 24-36) themselves, whose introduction to phonetic transcription contains a section entitled “Gelegenheits- und Verlegenheitszeichen” (pages 29-32), which can be translated as ‘(Transcription) signs employed occasionally and in case of embarrassment (= Verlegenheit)’, and a further one under the title “Inkonsequenzen, Unsicherheiten und Irrtümer in der Transkription” (‘Inconsistencies, uncertainties and errors in the transcription’, pages 33-35)⁶.

Such occasional uncertainty and variability forced us to make debatable choices, in order to proceed to IPA conversion. These choices are made explicit in the form of a conversion table that can be consulted on the database website.

Among the limitations imposed by the automated conversion process, some are rather harmless. For instance, such conversion made it impossible to mark stress at the beginning of the stressed syllable, as required by the IPA standards. For this reason, stress is marked immediately before the stressed vowel in the IPA transcriptions resulting from conversion, in keeping with an alternative convention that is occasionally encountered in the literature⁷.

In general, when consulting the database one must always keep in mind that the IPA transcription is a service provided to the user, subject to limits imposed by the original transcription. In particular, the automatic conversion mechanism has made it impossible to intervene to selectively correct transcription imperfections, which, tolerated in systems prevalent in Romance dialectology such as the AIS one, stand out as real mistakes in IPA transcriptions instead. This is the case, e.g., for the rendering of the palatal lateral with the symbol *l*, which the AIS collectors always wrote as singleton regardless of actual length, obviously applying the same principle of economy as in Italian spelling (since the palatal lateral approximant does not present any length contrast). This has the effect of obscuring the difference between singleton (in the north) and geminate realizations (from Tuscany southwards). Thus, on map 1.9, *mio figlio* ‘my son’ is transcribed *ɛ mi fiŋʎl* for pt. 511 (Campori, a hamlet of Castiglione Garfagnana), where the realization is actually not geminate; the IPA version is therefore faithful to the actual pronunciation here: [ɛ m_i i fiʎ^ho:l]. A transcription with singleton *l*, however, is also to be read for Florence (pt. 523: *i mmí fiŋʎlɔ*), although here the realization is actually geminate, as in Standard Italian: for Florence the IPA counterpart is therefore unfortunately incorrect [i m^h:i fiʎ^ho:lɔ] (instead of

⁶ Alongside Jaberg / Jud (1928, 24-36), the reader is referred to Jaberg / Jud (1927) and Canepari (1978) for description and discussion of the AIS transcription system.

⁷ By contrast, in the newly transcribed AISr data, the stress mark precedes the entire stressed syllable.

[i 'm:i fi'k:ɔlo], as it should be), given the impossibility of intervening by selectively correcting consonant length only for the points that needed it. The same applies to other non-contrastive geminates: for the palatal nasal, for example, map 3.541 *il legno* ‘the wood’ displays *ñ* almost everywhere, in Tuscany – including Florence – or in Lazio as well as in the north: the symbol is therefore converted into IPA as [ɲ] throughout (e.g., [l'eɲɔ] in Incisa, province of Florence, pt. 534; [l'eɲɔ] in Radda in Chianti, province of Siena, pt. 543; [l'e:ɲɔ] in Vinci, province of Florence, pt. 522), thereby obscuring the distinction between Tuscan [l'eɲ:ɔ] and Venetan [l'e:ɲɔ]. As apparent from these examples, vowels too may be in need of some adjustment. In fact, the AIS uses symbols without diacritics for mid vowels of intermediate height, while for IPA low vs. high mid cardinal vowels are transcribed with different symbols, and diacritics must be added to indicate intermediate height. Thus, for AIS *e o* we have chosen arbitrarily a lowered mid high [ɔ ɛ], as seen for pt. 543, whose AIS transcription is *lěño*, whereas for pts. 534 and 522 Scheuermeier notated a lower mid vowel, with diacritic: *lěño* and *il lěño* respectively. The latter transcription, furthermore, has a dubious macron which is respected in the IPA transliteration [e:] in spite of the fact that it is implausible for a stressed vowel to be allophonically lengthened before [ɲ:] in central Tuscany.

Such further adjustments of the IPA output, in these and similar cases, are left to the user.

3. What is the AISr database good for? A few examples

The AISr data make it possible to observe the reshaping over time of the spatial distribution of competing variables in all structural components. The following subsections will discuss a few selected examples.

3.1. Convergence and variation in western Surselva

Let us start by focusing on western Surselva, where pts. 10 (Camischolas, Val Tujetsch) and 11 (Surrhein/Sumvitg) represent the westernmost fringe of the Romansh speaking territory. The corresponding local dialects are the Sursilvan varieties which differ most from the literary standard (on Tuatschin, see Maurer Cecchini, forthcoming, with previous references). The AIS data attest to their divergence from the rest of the Surselva in several lexical and grammatical features which, as shown by the AISr data, have since become at least partly homogenized. This becomes apparent, for instance, on map 3.480 *la farfalla* ‘the butterfly’ (Figure 4)⁸, where in the AIS one finds, e.g., [lɛ 'tʃitɛ] in Brigels/Breil (pt. 1) vs. [lɛ 'biwɛ] in Surrhein/Sumvitg (pt. 11) and [lɛ 'belɛ] in Camischolas (pt. 10), while the AISr informant for pt. 11 has changed to [lɛ 'tʃitɛ].

⁸ AIS maps have a trilingual title (in this case *La farfalla – Schmetterling – Papillon*), which is meant to illustrate the concept corresponding to the questionnaire item in the three official and national languages of the Swiss Confederation at the time publication started. In what follows, we will cite only the Italian name, since it corresponds to the roofing language



Figure 4. Results of the query “480 – LA FARFALLA” (map display, showing the 9 points in the Surselva and Sutselva)

Likewise, the form of the m.sg definite article [ɐl] (Figure 5), which is characteristic for the local dialect (being a stereotype in the Labovian sense) and used to appear at pts. 10-11 in the western Surselva on the AIS maps, now competes with common Sursilvan [il] in the AISr data, whose answers display either forms: e.g., at pt. 10 [il 'bap] ‘the father’ 1.5, [il 'ciɹɹp] ‘the body’ 1.87, as opposed to [ɐl] in the AIS, which however persists in other questionnaire answers such as [ɐl 'viŋtɹ] ‘the belly’ 1.128, [ɐl 'dir] ‘the liver’ 1.139, etc. There is widespread instability in the area for this article form, as at pt. 11 we collected, e.g., [ɐl 'bap] as opposed to AIS [il 'ba:p].

AIS,r		Switch to map view	4 maps selected	10 – Camischolas (...)	Period	Phonetic	
AIS	IPA	Period	Location	Map	Download table		
al báp	el b'ap	past	10 – Camischolas (Tavetsch) 5 – IL PADRE				
al ciarp	el c'ierp	past	10 – Camischolas (Tavetsch) 87 – IL CORPO				
al viántar	el v'anter	past	10 – Camischolas (Tavetsch) 128 – IL VENTRE				
al d'ir	el d'ir	past	10 – Camischolas (Tavetsch) 139 – IL FEGATO				
—	il bap	present	10 – Camischolas (Tavetsch) 5 – IL PADRE				▶ Play
—	il ciɹɹp	present	10 – Camischolas (Tavetsch) 87 – IL CORPO				▶ Play
—	el 'viŋtɹ	present	10 – Camischolas (Tavetsch) 128 – IL VENTRE				▶ Play
—	el dir	present	10 – Camischolas (Tavetsch) 139 – IL FEGATO				▶ Play

Figure 5. Results of the query “5 – IL PADRE / 87 – IL CORPO / 128 – IL VENTRE / 139 – IL FEGATO” for pt. “10 – CAMISCHOLAS (TAVETSCH)” (table display)

The AISr data also hold some surprises in store relative to what is independently known from reference sources. Thus, map 2.264 *la sarta* ‘the (female) tailor’ shows the overwhelming prevalence of the Swiss German loanword [lɐ ʃnɛdr'i:nɐ] (the form is taken from Camischolas), with just one point in the whole Romansh area preserving

of one of the target dialect continua, though not the one relevant for the Surselva. (Recall that Romansh was declared a national language in 1938 – with a popular vote responding to Mussolini’s threat – and that the overarching standard for the five Romansh dialect subdivisions, viz. *Romansh Grischun*, was only established in the 1980s, by Schmid 1982.)

the original Romance name as an alternative: [lə kuz¹untʂə, ʃnidr¹eɲə] in Reams/Riom (pt. 25, in Oberhalbstein/Sursés). While the situation in Reams has not changed, the overall AISr picture shows a much broader diffusion of the Romance word ([lə kuz¹ontʂə] in Brigels, [lə ku¹zontʂə] in Camischolas, while in many other places the informants gave both answers).

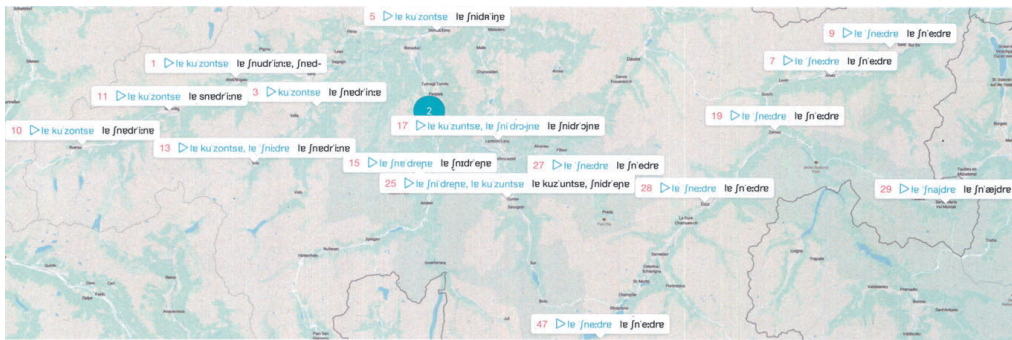


Figure 6. Results of the query “264 – SARTA” (map display, showing 16 (of 18) points in the Rumantschia)

This makes it possible to correct the unfavourable diagnosis contained in the DRG 4.607⁹:

cusunz, -a vermochte sich trotz Propagierung durch Wb. und Gramm. (Cahannes, *Gramm.* 180) insbesondere auch wegen seiner starken semantischen Bindung an ‘Näher, -in’ in der Bed[eutung] ‘Schneider, -in’ neben → *schneider*, -dra nicht durchzusetzen (cf. AIS. 2.259 und 264) [*cusunz*, -a was not able to assert itself alongside → *schneider*, -dra, despite propagation through dictionaries, especially because of its strong semantic link to the agent noun from ‘to sew’ in the meaning ‘tailor’ (see AIS 2.259 and 264)].

Of course, a caveat is in order here. Our comparison, in fact, is subject to the well-known limitations of any language atlas, which consists of a collection of *actes de parole* (Pop 1950, 131, 570-571): this means that, in principle, one might suspect that our informants’ responses to the translation task may be due to normative prescription, favouring indigenous *cusunza* over the German loan competitor¹⁰. A one-informant dialect atlas is not in a position to resolve such a doubt.

⁹ This is included under the entry *cusir*, written by Jachen Curdin Arquin for the fourth volume, which appeared in 1968.

¹⁰ As to this specific word (see Decurtins 1993, 212-216), we know that until the early 1930s the German loan was the most used: the dictionary by Pallioppi / Pallioppi (1895, 218) has the entry *cusunza*, an indigenous word documented as early as 1658, but is quick to add: “gewöhnlicher: *schnedra*” [‘more usually: *schnedra*’]; and Cahannes (1924, 180), discussing Germanisms, says: “Per ‘*schneider e schnadrina*’ vein nus ils buns plaids: *cusunz e cusunza*” [‘For tailor (male and female), we have got the good words *cusunz* and *cusunza*’]. These prescribed “good words” gained momentum especially in the 1930s, when the political situation hinted at in fn. 8 prompted a reaction, which also included a fight against non-indigenous words; see, e.g., the schoolbook by Steier (1933, 118 and 162). Dictionaries vacillate to this day: thus, the reader finds only *cusunz*, -a in Darms *et al.* (1989, 79) but both *cusunz* and *schneider* in Decurtins (2001, 249, 925).

3.2. The reshaping of lexical areas for some kinship terms

Noteworthy displacements in lexical distribution areas are also observable in core lexical domains, such as kinship terms. Moving from Surselva to Engadine, one can see an interesting case of lexical substitution due to German pressure. In order to describe this, it is first necessary to briefly retrace the history of the terms for ‘aunt’ in Lower Engadine. In a first stage, the whole region probably shared the type of onomatopoeic or infantile origin *jaja*; the DRG 10.162 still reports it only for the marginal areas north of Scuol towards the Austrian border and in Val Müstair. In the central area of Lower Engadine, instead, *jaja* is flanked by the inherited type *anda* (< Latin *AMĪTAM*), which is widespread, as shown by both DRG 1.263 and AIS 1.20 ‘their aunt, their aunts’, also in the rest of the Raeto-Romance territory. A more careful look at the AIS data (in black in the picture below) shows, however, that already in the early 1920s the Upper-Engadinian varieties (Fex-Platta pt. 47 and Zuoz pt. 28) presented alongside *anda*, reported as archaic, also the German loan *tanta*.



Figure 7. Results of the query “20 – LA LORO ZIA; LE LORO ZIE” (map display, showing the 6 points in the Engadine)

Comparison with AISr data clearly shows that during the last century the Germanism *tanta* (ultimately of French origin in German) has also expanded in Lower Engadine, supplanting *anda* in the whole region except, once again, the marginal area of Val Müstair (Santa Maria, pt. 29), where onomatopoeic *jaja* resists.

The example just illustrated for the Engadine region is interesting as well because the same dynamics of change can also be observed in the Italo-Romance area. In fact, also for the Lombard dialects of the Canton Tessin and the Italian-speaking part of Grisons, the AIS map 2.20 *la loro zia; le loro zie* ‘their aunt; their aunts’ attests to the presence of two distinct types, inherited *anda* (< AMĪTAM) alongside the Italian loan *zia*. Some varieties had both lexical types:

(3) AIS 1.20 (1920s)

<i>anda</i>	both types	<i>zia</i>
pt. 22 Olivone pt. 35 Bivio pt. 42 Sonogno pt. 45 Soglio pt. 46 Coltura pt. 50 Campo pt. 51 Vergeletto	pt. 31 Osco pt. 32 Chironico pt. 44 Mesocco pt. 52 Aurigeno pt. 70 Indemini	pt. 41 Caveragno pt. 53 Prosito pt. 58 Poschiavo pt. 71 Breno pt. 73 Corticiasca pt. 93 Ligornetto

In this case, too, comparison with AISr data shows a spread of borrowed *zia* to the detriment of autochthonous *anda*, now preserved only in isolated datapoints, such as Bivio (pt. 35), Olivone (pt. 22) and Caveragno (pt. 41), as shown in (4) and in Figure 8:

(4) AISr 1.20 (late 2010s)

<i>anda</i>	both types	<i>zia</i>
pt. 35 Bivio	pt. 22 Olivone pt. 41 Caveragno	pt. 31 Osco pt. 32 Chironico pt. 42 Sonogno pt. 44 Mesocco pt. 45 Soglio pt. 46 Coltura pt. 50 Campo pt. 51 Vergeletto pt. 52 Aurigeno pt. 53 Prosito pt. 58 Poschiavo pt. 70 Indemini pt. 71 Breno pt. 93 Ligornetto



Figure 8. Results of the query “20 – LA LORO ZIA; LE LORO ZIE” (map display, showing a selection of 9 points in the Canton Tessin)

A further example of the influence of the roofing language is visible in the case of *génar* ‘son-in-law’. In fact, the LSI 2.650 entry *génar*² reports almost exclusively *gendar* < GĚNĚRUM (REW 3730), with epenthetic *-d-* introduced after the syncope of *-e-*: only the urban centres of Bellinzona and Locarno, the region of Centovalli, and the localities of Rossa, Muggio, Palagnedra, Brissago and Rivera attest an “Italian” form without epenthesis of the dental consonant. According to the LSI, the *gendar* type is widespread in the rest of the territory of Canton Tessin and the Italian Grisons. A similar picture emerges from the data in AIS 1.33 *il genero* ‘the son-in-law’: among the 18 investigated varieties, 16 reported a translation with epenthetic *-d-*, whereas only the informants from Ligornetto (pt. 93), in the extreme south of the Canton Tessin, and Poschiavo (pt. 58), in the valley of the same name, gave the forms [ul 'dʒenɐr] and [ɛl 'dʒenɐr] respectively. Once again, the AISr survey has highlighted a change: as can be seen in (5), the influence of the roofing language has turned the picture upside down, with as many as 14 varieties out of 18 displaying an outcome without epenthetic *-d-*:

(5)

<i>genar</i>		<i>gendar</i>	
[ɔl 'ʒendra]	pt. 22 Olivone	[ul 'dʒe:ndru]	pt. 31 Osco
[ɔl 'dʒe:nɐr]	pt. 32 Chironico	[ɛl mɛ 'dʒendɐr]	pt. 44 Mesocco
[ɛl 'dʒe:nɐr]	pt. 35 Bivio	[ɛl 'ʒe:ndru]	pt. 51 Vergeletto
[lu 'dʒe:nɐr]	pt. 41 Cavigno	[ɛl 'ʒe:ndro]	pt. 53 Prosito
[ɛl 'dʒe:nɐr]	pt. 42 Sonogno		
[ɛl 'dʒenɐr]	pt. 45 Soglio		
[ɛl mɛ 'dʒe:nɐr]	pt. 46 Coltura		
[ɛl 'dʒe:nɐr]	pt. 50 Campo		
[al 'dʒe:nɐr]	pt. 52 Aurigeno		
[al 'dʒe:nɐr]	pt. 58 Poschiavo		
[u 'dʒe:nɐr]	pt. 70 Indemini		
[o 'dʒe:nɐr]	pt. 71 Breno		
[ɛl 'dʒe:nɐr]	pt. 73 Corticiasca		
[ul 'dʒenɐr]	pt. 93 Ligornetto		

3.3. The morphosyntax of gender agreement in the dialect of Mesocco

The diachronic comparison between AIS and AISr data can shed light on the evolution of inflectional systems, showing that change does not always tend towards

standardization under the pressure of the roofing language, as observed in §3.1 with respect to the morphology of definite articles in Romansh.

To illustrate this point, let us focus on agreement morphology in the dialect of Mesocco, one of the Italo-Romance varieties spoken in the Grisons (see Salvioni 1902, 139; Sganzi 1932; Loporcario 2006; Manzini / Savoia 2019; Pescarini 2020). In Mesocco – and, more generally, in Mesolcinese dialects – the verb in the third person plural agrees in gender, with the suffix *-n* occurring if the subject or the (clitic) object is feminine:

- (6) a. la vaka l= e mōn'dʒud-a AISr 6.1195, pt. 44
the cow 3F= be.3.SG milked
'The cow is milked.'
- b. la vak-en l= e:-n mōn'dʒud-en
the cow-F.PL 3F= be-3.F.PL milked-F.PL
'The cows are milked.'

By contrast, with masculine subjects the third person plural is syncretic with the singular: e.g., *i 'diz* vs. *la 'diz-en* 'they.M/-F say' (Salvioni 1902, 911). Comparison of AISr and AIS data shows that the gender agreement system has been extended to verb tenses that, in the 1920s, did not exhibit gender agreement. According to the data contained in the AIS, gender agreement was found in the present indicative, but not in the future, which exhibited the inflection *-n* with both masculine and feminine subjects (as a leftover of the original distribution in which *-n* marked 3PL person/number rather than gender):

- (7) i= foràn kel k i= voràn AIS 8.1661, pt. 44
3M.PL= do.FUT.3PL that REL 3M.PL= want.FUT.3PL
'they will do what they will want'

Conversely, nowadays the ending *-n* is selected just by feminine subjects, in the future as well as in the present, thus marking gender agreement across tenses:

- (8) a. i= fa'ra(*n) kel k i= vo:(*ran) AISr 8.1661, pt. 44
3M.PL= do.FUT.3PL that REL 3M.PL= want.FUT.3PL
- b. la= fa'ra-n kel ke la= vo'ra-n
3F= do.FUT.3-F.PL that REL 3F= want.FUT.3-F.PL

- (10) a. nu 'mɔ:v=ɐt stage I AISr 8.1647, pt. 41
 NEG move.PRS.2SG=REFL
- b. nu 'mɔv=ɐt 'mi:a stage II
 NEG move.PRS.2SG=REFL NEG
- c. 'mɔv=ɐt 'mia stage III
 move.PRS.2SG=REFL NEG
 'do not move'

Thus, based on AISr data, the present-day distribution of sentential negation in Swiss-Lombard dialects can be summarized as follow:

- (11) a. Stage I: Soglio and Coltura (pts. 45, 46)
 b. Stage II: Sonogno, Aurigeno, Indemini and Ligornetto (pts. 42, 52, 70, 93)
 c. Stage III: Breno, Corticiasca, Olivone, Osco, Chironico, Bivio, Mesocco, Cimalmotto, Vergeletto, Prosito, Poschiavo (pts. 71, 73, 22, 31, 32, 35, 44, 50, 52, 53, 58)
 d. Stages I to III: Caverigno (pt. 41).

The case of Caverigno lends itself to an observation of method. The AISr data in (10) reflect a survey which, although conducted with a methodology comparable to that of AIS, aimed at documenting variation, especially for syntactic aspects such as the one under scrutiny here. This is also true, in principle, for AIS, whose openness towards the dynamics of variation is well-known. And yet in this specific case the same answer on the AIS map presents only one of the three options, namely discontinuous negation: *nu mɔvad miɑ* (AIS 8.1647, pt. 41). Taken globally, however, the AIS data show that the same variation between the three stages in (10) was already observed in the early 20th century:

- (12) a. Stage I: *ù n az movéva* AIS 8.1665 'he wasn't moving' Caverigno AIS, pt. 41
 b. Stage II: *k ù n al tròḗ mĩɑ* AIS 8.1661 'that you.PL don't find him'
 c. Stage III: *a dǎrmi prǒpi mĩɑ* AIS 4.653 'I don't sleep at all', *a kapĩsi mĩɑ* AIS 4.653 'I don't understand'

Given what we know in general about Jespersen's cycle, the opposite would be odd, since the discontinuous negation presupposes the preverbal one. Consequently, that the latter was present in AISr ((10a)) but not in the AIS data – as inspection of map 8.1647 alone would seem to indicate – would be a priori implausible. Therefore, although one individual piece of data sometimes, as in this case, does not give a complete picture, taken globally AIS data are confirmed to be a reliable resource, including for the study of syntax. Consequently, they are also a valuable basis of comparison, in diachrony, with respect to the data collected today¹¹.

¹¹ Obviously, this also applies to comparison with previously collected data. For Caverigno, in fact, one can resort to the late 19th-century texts published by Salvioni (1905) and their linguistic commentary, edited posthumously by C. Merlo (Salvioni 1937, 475f.). Those data show that the three options already coexisted then and that therefore the variation in (10) has been present for well over a century, albeit with a quantitative shift, in accordance with the predictions offered by Jespersen's cycle. Indeed, preverbal negation occurs more frequently in Salvioni's texts than postverbal negation, which was still rare at the time (Zeli 1968, 297).

The phenomena falling under the rubric “Jespersen’s cycle” intersect with so-called negative concord. This is observed, in dialects with preverbal negation, when the latter co-occurs with negative aspectual adverbs (‘(n)ever’ and ‘anymore’) and postverbal negative quantifiers (‘nobody’ and ‘nothing’). As shown in (13), negative concord in dialects with (residual) preverbal negation is not mandatory:

- (13) a. lú nu kúr mǎy AIS 8.1605, pt. 46
 3SG NEG run.PRS.3SG never
 b. ly ɐ= l= kor ma i AISr
 3SG CL= CL3SG.M= run.PRS.3SG never
 ‘he never runs’

Now consider negative quantifiers such as *nient*, *na(g)ota* ‘nothing’. In the AIS Swiss-Lombard datapoints, preverbal negators can double negative polarity items, whereas postverbal negators such as *mia* are in complementary distribution with negative items, as foreseen by Parry’s (1997) generalization. Hence, negative concord is attested only in dialects with preverbal ((14a)) or discontinuous negation ((14b))¹², but not in stage III varieties where negation has become entirely postverbal ((14c)):

- (14) a. nʉ vɛ́la nagóta AIS 4.829 ‘non valgono niente’ Soglio, pt. 45¹³
 b. i ne vār naóta Sonogno, pt. 42
 c. i vɛ́r niént Olivone, pt. 22¹⁴
 3PL NEG are_worth nothing
 ‘They aren’t worth anything’

We observe, however, that the negative quantifier within a prepositional phrase can co-occur with a postverbal negator:

- (15) a. à= n= trov míga da neşúna párt AIS 8.1597, pt. 73
 CL= NEG= find.PRS.ISG NEG at no_one place (Corticiasca)
 b. a= l= trǒvia mínga i násũn sít AIS 8.1597, pt. 22
 CL= 3SG.ACC= find.PRS.1SG NEG in no_one place (Olivone)
 ‘I cannot find him anywhere’

Thus, with respect to prepositional phrases containing a negative quantifier, four patterns are found in the AIS/AISr dataset; each type is illustrated in (16) with data from AISr:

- (16) AISr 8.1597 ‘non lo trovo in nessun luogo’ (‘I cannot find it anywhere’)

¹² Notice that the postverbal negative marker is in complementary distribution with postverbal negative words, which entails the neutralization of the contrast between preverbal and discontinuous negation.

¹³ Coltura, pt. 46, exhibits the same pattern: *nʉ vɛ́lan nagót*.

¹⁴ Osco, pt. 31, exhibits the same pattern: *u vɛ́ran nõta*.

- a. Type 1: preverbal negation without negative concord¹⁵, e.g. *a ne l'trɔ:v in pa:rt* (Sonogno, pt. 42)
- b. Type 2: preverbal negation with negative concord, e.g. *e no l trɔ:f də 'negyɲ lɔ:k* (Soglio, pt. 45)
- c. Type 3: postverbal negation with negative concord, e.g. *el 'trɔvi 'miga in ni'sum sit* (Mesocco, pt. 44)
- d. Type 4: postverbal negation without negative concord, e.g. *a l 'trovjə ɪ nɐ'ɕ:ɲ ɕit* (Olivone, pt. 22)

Here, too, comparison of the two datasets shows – to put it in merely descriptive terms – that, if change has occurred, this happens overwhelmingly (except for Bivio, pt. 35) in the form of a rightward shift of the negative element¹⁶:

(17) Types of negative concord (see (16)) with prepositional phrases: AIS vs. AISr 8.1597

	AIS		AISr
42 – Sonogno	1	=	1
41 – Cavergho	1	>	2
45 – Soglio	2	=	2
46 – Coltura (Stampa)	2	=	2
71 – Breno	2	=	2
73 – Corticiasca	2	>	4
44 – Mesocco	3	=	3
52 – Aurigeno	3	=	3
50 – Campo	3	>	3

	AIS		AISr
53 – Prosito (Lodrino)	3	=	3
22 – Olivone	3	>	4
51 – Vergeletto	3	>	4
70 – Indemini	3	>	4
93 – Ligornetto	3	>	4
31 – Osco	4	=	4
32 – Chironico	4	=	4
58 – Poschiavo	4	=	4
35 – Bivio – Beiva	4	>	3

3.5. Sound change

While we have so far provided examples in the areas of morphology, syntax and the lexicon, the AISr database can also obviously be instrumental to investigating sound change, since it allows for an immediate comparison between parallel datasets at a one century interval. As an example, we will consider the outcomes of Latin *c^{v/i}*-, drawing on Negrinelli (2021, 286). Table (18) compares the AIS and AISr datasets for pt. 19 Zernez, on the border between the Puter and the Vallader linguistic areas (the greyed-out cells show the cases in which change has occurred):

¹⁵ Although the prepositional phrase *in part* (literally ‘in [a] place’) does not contain any negative adverb/quantifier, in some Lombard dialects it behaves like a negative polarity item (like Eng. *anywhere*).

¹⁶ The like sign in (17) indicates diachronic persistence. Comparison is subject to the usual caveat (see above): the AIS and, to a lesser extent, the AISr may contain a subset of the possible variants.

(18)	AIS map	AIS/1920s	AISr/late 2010s	
a.	[ʃ]/[tʃ] > [tʃ]	5.945 'dopo cena'	[ʃe̞inə] / [tʃe̞inə]	[tʃa̞inə]
		5.909 'la cera'	[ʃa̞ərə] / [tʃa̞ərə]	[tʃa̞ərə]
		7.1282 'la ciliegia'	[ʃi're:ʒe] / [tʃi're:ʒe]	[tʃe're:ʒe]
		1.94 'il cervello'	[ʃer've] / [tʃer've]	[tʃer've]
b.	[ʃ] > [tʃ]	5.930 'la cenere'	[ʃendrə]	[tʃendrə]
		7.1368 'la cipolla'	[ʃi'guələ]	[tʃi'guələ]
		6.1219 'la ricotta'	[ʃi'grun]	[tʃi'grun]
		3.507 'la civetta'	[ʃo'etə]	[tʃu'wetə]
		8.1564 'la cintura'	[ʃintə]	[tʃintə]
		3.636 'andare a cercare'	[ʃer'cær]	[tʃær'cær]
c.	[ʃ]	1.102 'le ciglia'	[ʃa̞iʎs]	[ʃa̞jəs]
		2.365 'la nebbia'	[ʃi'εərə]	[ʃi'əərə]
d.	[tʃ]	2.286 'cinque; sei'	[tʃinc]	[tʃe̞inc]
		2.304 'cento'	[tʃiɛnt]	[tʃjɛnt]
		2.310 'cinquanta'	[tʃiŋ'kwantə]	[tʃiŋ'kwantə]

The situation recorded in the 1920s (third column) was more variegated, with free variation between word-initial [tʃ] and [ʃ] reported for a few lexical types (*tscheina* 'dinner', *tschera* 'wax', *tscherescha* 'cherry' and *tschurvi* 'brain'), while on the whole the vast majority of lexemes were recorded with [ʃ] (from earlier *[tʃ] < c^{e/i}-). The same situation is widely described in the literature with reference to the first half of the 20th century¹⁷. Consider, for example, the following remarks by Walberg (1907, §107):

Davanti a *e, i c* si riduce a *tʃ* [...] In *flaríña*, anticamente *Tschlari(g)na* (*CELLARINA?), il gruppo consonantico, di pronunzia difficile, è stato ridotto del suo primo elemento. [...] Del resto la esplosiva iniziale è anche negli altri casi poco energica e sta per dileguarsi completamente. Ho sentito più di una volta forme come *fartçer, fintç, fiŋkænta* QUINQUAGINTA, *fiənt, fiént* CENTUM, *fúrma* 'ciurma', accanto a quelle con *tʃ*. ['Before *e, i c* changes to *tʃ* [...] In *flaríña*, formerly *Tschlari(g)na* (*CELLARINA?), the hard-to-pronounce consonant cluster has been stripped of its first element.¹⁸ [...] After all, the initial plosive is also in other cases not very energetic and is about to disappear completely. I have heard more than once forms like *fartçer, fintç, fiŋkænta* QUINQUAGINTA, *fiənt, fiént* CENTUM, *fúrma* 'crew', alongside those with *tʃ*.']

¹⁷ See Eichenhofer (1999, §281 n. 1), who reviews earlier descriptions: Luzi (1904, §87) for Sutsilvan; Lutta (1923, §128) and Grisch (1939, §17) for Surmiran; Pult (1897, §179) for the Lower Engadine dialect of Sent.

¹⁸ Walberg's terminology is infelicitous, as it does not do justice to the fact that the "first element" in such a consonant cluster, prior to change, was the entire affricate, while what was lost is just its plosive gesture.

The situation displayed in the fourth column in (18) contradicts Walberg's speculation: the AISr data show that over time the palato-alveolar affricate [tʃ] has gained ground among the lexemes that showed variation in the 1920s ((18a)), and further spread also to many words that Scheuermeier recorded only with initial [j] in the AIS survey ((18b)), such as *tschendra* 'ash', *tschaguola* 'onion', *tschigrun* 'ricotta cheese', *tschuetta* 'owl', *tschenta* 'belt' and *tschercar* 'to search'. This development is one out of many which are probably due to the normalizing pressure of official spelling, which presents the <tsch> (= [tʃ]) grapheme in all the cases mentioned: over time, adherence to these spellings may have eradicated free variation between the two pronunciations in (18a) and directed the change in the others. In fact, the spelling is the same also for the words in (18c-d): *tschagl* 'eyebrow', *tschiera* 'fog', *tschinch* 'five', *tschient* 'hundred', *tschinqu(u)nta* 'fifty'.

4. Conclusion and further prospects

To conclude, the same limitations addressed above in §2, imposed by reasons of feasibility in the three-year funding period, also indicate two paths for further development:

- (19) a. digitization of the remaining 825 AIS maps in searchable form using the same technique;
- b. extension of fieldwork to Italy, Slovenia and Croatia, to cover the rest of the territory on which the AIS datapoints are scattered.

In 2020, a sequel four-year project was funded by SNSF, which will consist in the implementation of objective (19a). As for (19b), the project foresees a fieldwork campaign in Lombardy, which started on September, 2021, under harder circumstances than expected in the planning stage due to the Covid-19 pandemic¹⁹. Once this is completed, *deo iuvante et nobis viventibus*, the database will include present-day data for about one fifth of the 407 AIS datapoints²⁰. This will be a good starting point for

¹⁹ In the hope that the situation will have somehow normalized by then, one can't help but note that the very high lethality rate in Lombardy during spring 2020 massively impacted the population of eligible informants, since, as seen in (2), the overwhelming majority of them were aged people. This is due to the character of the original AIS questionnaire, with its focus on rural life and related traditional activities and practices, which, in the early 21st century, implies selecting elderly people as informants.

²⁰ Further, smaller-scale campaigns are currently being planned at the University of Zurich: the six datapoints in Istria and Dalmatia are the focus of Giudici/Negrinelli (forthcoming). In addition, the *Atlante Multimediale dei Dialetti Veneti* (AMDV) includes modern data from Veneto, recorded with a comparable procedure (and indeed served as a source of inspiration for the AISr project). However, full comparability is hindered by the fact that, due to funding limitations, AMDV covers 430 lexical entries, or about one fourth of the original AIS questionnaire (see Tisato/Vigolo 2011, 109).

further extension, with the ultimate goal of covering the remaining territory, region after region, possibly in cooperation with local teams²¹.

Should we succeed in the enterprise, history will repeat itself: Jaberg and Jud, too, started in Switzerland and first planned an extension to northern Italy only (see Jaberg/Jud 1928, 12-16), to eventually end up with a monumental atlas covering all of the Romance varieties of southern Switzerland and the whole of Italy.

Universität Zürich, Romanisches Seminar	Michele LOPORCARO
Universität Zürich, Phonetisches Laboratorium	Stephan SCHMID
Universität Zürich, Romanisches Seminar	Chiara ZANINI
CNRS, Université Côte d'Azur, BCL	Diego PESCARINI
Universität Zürich, Romanisches Seminar	Giulia DONZELLI
Universität Zürich, Romanisches Seminar	Stefano NEGRINELLI
ISTC – CNR, Padua	Graziano TISATO

References

- AIS = Jaberg, Karl/Jud, Jakob, 1928-1940. *Sprach- und Sachatlas Italiens und der Südschweiz*, Zofingen, Ringier, 8 vols.
- AISr = Loporcaro, Michele *et al.*, 2019. *AIS, reloaded*, University of Zurich, SNSF grant 100012-162482 (2016-2019), <<https://www.ais-reloaded.uzh.ch/>>.
- AISdt = Loporcaro, Michele *et al.*, forthcoming. *AIS, the digital turn*, University of Zurich, SNSF grant 100012_192259 (2021-2024).
- ALI = Bartoli, Matteo Giulio *et al.*, 1995-. *Atlante linguistico italiano*, Rome, Istituto Poligrafico e Zecca dello Stato/Libreria dello Stato.
- AMDV = Zamboni, Alberto/Tisato, Graziano/Vigolo, M. Teresa *et al.*, 2009-2015. *Atlante Multimediale dei Dialetti Veneti*, funded by Fondazione CARIPARO, Padua, University of Padua/ISTC – CNR.
- Bernardasci, Camilla/Schwarzenbach, Michael (ed.), 2016. *Stòri, stralüsc e stremizzi. Registra-zioni dialettali nella Svizzera italiana (1929)*, Bellinzona, Salvioni arti grafiche (*Quaderni del Bollettino Storico della Svizzera Italiana* 13).
- Cahannes, Gion, 1924. *Grammatica romontscha per Surselva e Sutselva*, Mustér, Lia Romontscha.
- Canepari, Luciano, 1978. «Raffronto tra il sistema IPA e altri sistemi di trascrizione fonetica», *Rivista Italiana di Dialettologia* 2/1, 102-113.
- Darms, Georges/Dazzi, Anna Alice/Gross, Manfred, 1989. *Langenscheidts Wörterbuch Rätoromanisch. Rätoromanisch-Deutsch, Deutsch-Rätoromanisch*, Zürich, Langenscheidt.

²¹ While, for comparability, it is essential that the new survey be completed within a reasonable time frame, the issue of uniformity in transcription has lost some importance, given that users themselves are now able to check transcriptions against the sound input.

- Decurtins, Alexi, 1993. «Zur Problematik von Neuschöpfungen im Bündnerromanischen (erweiterte Version nach der romanischen Originalfassung, erschienen in: *Annalas* 88, 9-51)», in: Decurtins, Alexi (ed.), *Rätoromanisch. Aufsätze zur Sprach-, Kulturgeschichte und zur Kulturpolitik*, Cuira, Società Retorumantscha, 193-233 (*Romanica Raetica* 8).
- Decurtins, Alexi, 2001. *Niev vocabulari romontsch: sursilvan-tudestg*, Chur, Legat Anton Cadonau/Societad retorumantscha/Societad per la perscrutaziun da la cultura grischuna.
- Donzelli, Giulia, 2020 [ms.]. *Le strutture interrogative nei dialetti lombardi della Svizzera Italiana*, PhD thesis, University of Zurich.
- DRG = von Planta, Robert *et al.*, 1939–. *Dicziunari rumantsch grischun*. Publichà da la Società retorumantscha cul agüd de la Confederacziun, dal Chantun Grischun e da la Lia rumantscha, Cuira, Bischofberger (then Cuira, Institut dal Dicziunari Rumantsch Grischun).
- Eichenhofer, Wolfgang, 1999. *Historische Lautlehre des Bündnerromanischen*, Tübingen/Basel, Francke.
- Giudici, Alberto/Negrinelli, Stefano, forthcoming. *Le inchieste AIS nell'Adriatico orientale a 100 anni di distanza*, ms., University of Zurich.
- Grisch, Mena, 1939. *Die Mundart von Surmeir (Ober- und Unterhalbstein). Beitrag zur Kenntnis einer rätoromanischen Sprachlandschaft*, Paris, Droz/Zurich-Leipzig, Max Niehans (*Romanica Helvetica* 12).
- Jaberg, Karl/Jud, Jakob, 1927. «Transkriptionsverfahren, Aussprache und Gehörschwankungen», *Zeitschrift für romanische Philologie* 47, 172-218.
- Jaberg, Karl/Jud, Jakob, 1928. *Der Sprachatlas als Forschungsinstrument*, Halle, Niemeyer (repr. Nendeln, Kraus, 1973).
- Jespersen, Otto, 1917. *Negation in English and Other Languages*, Copenhagen, A. F. Høst.
- Kunz, Aline, 2018. *Tra la polvere dei libri e della vita: il carteggio Jaberg-Scheuermeier, 1919-1925*, Alessandria, Edizioni dell'Orso.
- Levenshtein, Vladimir, 1966. «Binary codes capable of correcting deletions, insertions, and reversals», *Soviet Physics Doklady* 10, 707-710.
- Loporcaro, Michele, 2006. «Crossing form and function: first and second person plural imperatives in the dialect of Mesocco», *Folia linguistica* 40, 135-154.
- Loporcaro, Michele, 2008. «Carlo Salvioni linguista», in: Loporcaro, Michele/Pescia, Lorenza/Broggini, Romano/Vecchio, Paola (ed.), Carlo Salvioni, *Scritti linguistici*, 5 vols., Bellinzona, Edizioni dello Stato del Cantone Ticino, vol. 5, 45-113.
- Loporcaro, Michele, 2016. «Tra Zurigo, Bellinzona, Milano e Pisa: dialettologia italiana in Svizzera nella prima metà del Novecento (con una lettera di C. Merlo a J. Jud)», *Studi italiani di linguistica teorica e applicata* 45, 555-578.
- LSI = Lurà, Franco/Moretti, Michele, 2004. *Lessico dialettale della Svizzera italiana*, Bellinzona, Centro di dialettologia e di etnografia, 5 vols.
- Lutta, Conrad Martin, 1923. *Der Dialekt von Bergün und seine Stellung innerhalb der rätoromanischen Mundarten Graubündens*, Halle a. S., Niemeyer (*Beihefte ZRPh* 71).
- Luzi, Johann, 1904. *Lautlehre der Subselvischen Dialekte*, Dissertation Zürich, Erlangen, Junge & Sohn.
- Manzini, Maria Rita/Savoia, Leonardo Maria, 2005. *I dialetti italiani e romanci. Morfosintassi generativa*, Alessandria, Edizioni dell'Orso, 3 vols.
- Manzini, Maria Rita/Savoia, Leonardo Maria, 2019. «N morphology and its interpretation: Romance feminine/plural -a», in: Cruschina, Silvio/Ledgeway, Adam/Remberger, Eva Maria (ed.), *Italian Dialectology at the Interfaces*, Amsterdam, Benjamins, 257-293.

- Maurer Cecchini, Philipp, forthcoming. *A Grammar of Tuatschin*, Berlin, Language Science Press.
- NavigAIS = Tisato, Graziano G., 2009-2018. *NavigAIS – AIS Navigator*, ISTC – CNR, Padua, <<https://www3.pd.istc.cnr.it/navigais>>.
- Negrinelli, Stefano, 2021. *Gli esiti delle occlusive velari latine nelle varietà retoromanze grigionesi dell’AIS*, PhD thesis, University of Zurich.
- Pallioppi, Emil/Pallioppi, Zaccaria, 1895-1902. *Dizionari dels idioms romauntschs d’Engiadin’ ota e bassa, della Val Müstair, da Bravuogn e Filisur, con particulera consideraziun del idiom d’Engiadin’ ota/Wörterbuch der romanischen Mundarten des Ober- und Unterengadins, des Münsterthals, von Bergün und Filisur, mit besonderer Berücksichtigung der oberengadinischen Mundart*, Samedan, Tanner, 2 vols.
- Parry, Mair, 1997. «Negation», in: Maiden, Martin/Parry, Mair (eds.), *The dialects of Italy*, London, Routledge, 179-185.
- Parry, Mair, 1998. «On negation in the Ligurian Hinterland», *Quaderni di Lavoro ASIt* 2, 85-112.
- Penello, Nicoletta/Pescarini, Diego, 2008. «Osservazioni su mica in italiano e alcuni dialetti veneti», *Quaderni di Lavoro ASIt* 8, 43-56.
- Pescarini, Diego, 2005. «Mica nell’area metropolitana di Verona», in: Marcato, Gianna (ed.), *Dialetti in città*, Padova, Unipress, 283-288.
- Pescarini, Diego, 2020. «L’accordo asimmetrico nel Grigioni italiano. Convergenze morfologiche e divergenze sintattiche», *L’Italia Dialettale* 82, 211-241.
- Pescarini, Diego/Donzelli, Giulia, 2017. «La negazione nei dialetti della Svizzera italiana», *Vox Romanica* 76, 74-96.
- Pop, Sever, 1950. *La dialectologie. Aperçu historique et méthodes d’enquêtes linguistiques. Première partie: Dialectologie romane*, Louvain, the author/Gembloux, J. Duculot.
- Pult, Gaspard, 1897. *Le parler de Sent (Basse-Engadine)*, Dissertation Lausanne, Lausanne, F. Payot.
- REW = Meyer-Lübke, Wilhelm, 1935³. *Romanisches Etymologisches Wörterbuch*, Heidelberg, Winter.
- Salvioni, Carlo, 1902. «Del plurale femminile di 1^a declinazione esposto per -a ed -ǎn in qualche varietà alpina di Lombardia», *Rendiconti del Reale Istituto Lombardo di Scienze e Lettere* (s. 2) 35, 905-919. Repr. in: Loporcaro, Michele/Pescia, Lorenza/Broggini, Romano/Vecchio, Paola (ed.), 2008. Carlo Salvioni, *Scritti linguistici*, 5 vols., Bellinzona, Edizioni dello Stato del Cantone Ticino, vol. 1, 133-147.
- Salvioni, Carlo, 1905. «Poesie in dialetto di Caveragno (Valmaggia)», *AGI* 16, 549-590. Repr. in: Loporcaro, Michele/Pescia, Lorenza/Broggini, Romano/Vecchio, Paola (ed.), 2008. Carlo Salvioni, *Scritti linguistici*, 5 vols., Bellinzona, Edizioni dello Stato del Cantone Ticino, vol. 1, 375-416.
- Salvioni, Carlo, 1937. «Illustrazione dei testi di Caveragno (valle Maggia), edite, con aggiunte, da C. Merlo», *L’Italia Dialettale* 13, 1-55. Repr. in: Loporcaro, Michele/Pescia, Lorenza/Broggini, Romano/Vecchio, Paola (ed.), 2008. Carlo Salvioni, *Scritti linguistici*, 5 vols., Bellinzona, Edizioni dello Stato del Cantone Ticino, vol. 1, 465-519.
- Schmid, Heinrich, 1982. *Richtlinien für die Gestaltung einer gesamtbündnerromanischen Schriftsprache “Rumantsch Grischun”*, Chur, Lia Rumantscha.
- Schorta, Andrea, 1952. «Professer dr. Jakob Jud: 12.1.1882-19.6.1952», *Annalas da la Societad Retorumantscha* 66, 6-20.

- Sganzini, Silvio, 1932. «Di alcune forme verbali nella parlata di Mesocco», *L'Italia Dialettale* 8, 259-263. Repr. in Sganzini, Silvio, 1993. *Scritti dialettologici*, Basel/Tübingen, Francke, 57-61.
- Steier, Andrea, 1933. *Codesch da liger per las scolas primaras da Surmeir. Publitgia digl Cunsegl Pitschen*, Glion, Maggi.
- Tisato, Graziano/Vigolo, M. Teresa, 2011. «AMD. Atlante Multimediale dei Dialetti Veneti», in: Borgato, Gianluigi/Vanelli, Laura (ed.), *In ricordo di Alberto Zamboni. Atti del Convegno. Padova, 25 gennaio 2011*, Padova, Università di Padova – Dipartimento di Discipline Linguistiche, Comunicative e dello Spettacolo, 99-126.
- Vai, Massimo, 1996. «Per una storia della negazione in milanese in comparazione con altre varietà altoitaliane», *ACME* 49/1, 57-98.
- Vithlani, Purna/Kumbharana, C.K., 2015. «Comparative study of character recognition tools», *International Journal of Computer Applications* 118/9, 31-36.
- Walberg, Emmanuel, 1907. *Saggio sulla fonetica del parlare di Celerina-Cresta (Alta Engadina)*, Lund, Ohlsson.
- Zanuttini, Raffaella, 1987. «Negazione e concordanza negativa in italiano e in piemontese», *Rivista di Grammatica Generativa* 12, 153-172.
- Zeli, Rosanna, 1968. «Spunti per lo studio della negazione nei dialetti del Ticino e del Moesano», *Vox Romanica* 27, 289-298.