

Project report



German-Romance Language Contact in the Italian Alps: documentation, explanation, participation

Progetto di ricerca di rilevante interesse nazionale

Bando PRIN 2020, Decreto 1628 del 16/10/2020

Codice CUP B37G22000160006

Decreto di finanziamento 03/03/2022

Data inizio progetto 01/06/2022

Data chiusura progetto 01/06/2025

Project no. 2020SYSYBS

Principal Investigator: Stefan Rabanus

Contents

Project description	4
Team members	4
Work packages	5
Project timeline.....	6
WP1. Management and coordination	11
WP1.1 Team communication	11
WP1.2 Public communication	12
WP1.3 Contracts.....	12
WP1.4 Project partners.....	12
WP2. Platform creation and maintenance	13
WP2.1 Word Press: Website	13
WP2.2 Phonic: Data collection.....	14
WP2.3 Interactive map	15
WP2.4 Linguistic Landscape	16
WP3. Data Management Plan and personal data protection.....	17
WP3.1 Data Management Plan (DMP)	17
WP3.2 Data Processing Agreement (DPA).....	17
WP3.3 AlpiLinK repository	17
WP3.4 Recovering and restructuring of the data of the previous AThEME and VinKo projects and creation of the AThEME and VinKo repositories	18
WP4: Development of tools for data extraction, management and analysis	19
WP4.1 Half-automated Annotation Kit for Interlinguistic Material (HAKIM)	19
WP4.2 Minority Language Grapheme-To-Phoneme Conversion Tool (MinorityG2P).....	20
WP4.3 Automated Audio Acceptability Assessment Tool (AAAA)	20
WP5: Data collection.....	21
WP5.1 Questionnaire.....	21
WP5.2 Data collection.....	23
WP5.3 Audio quality control.....	26
WP6: Data elaboration and analysis	27
WP6.1 Data elaboration.....	27
WP6.2 Data analysis.....	28
WP6.3 Evaluation of methodology and project.....	30
WP6.3.1 Analysis of methodology	30
WP6.3.2 Participant appreciation of the project.....	35
WP7. Documentation and publication of results	38
WP7.1 Publications	38
WP7.2 Conferences	41

WP7.3 Advanced academic teaching	46
WP8: Public outreach: speech-community participation	48
WP8.1 VinKiamo	48
WP8.2 Public relations	52
WP8.3 Community support	53
Appendix 2.1 Website pages	54
Appendix 2.2 How to use Phonic	56
Appendix 2.3 Map, included stimuli	61
Appendix 2.4 Linguistic Landscape workflow.....	66
Appendix 3.1 Data Management Plan.....	69
Appendix 3.3 Data export: from Phonic to AlpiLinK repository	71
Appendix 5.1 Questionnaire.....	75
Appendix 5.3 Audio quality control workflow	83

The AlpiLinK report was written by Stefan Rabanus, using a structure and many contents prepared by Anne Kruijt, with contributions from Ilaria Driussi.

Last update of the AlpiLinK report: 31 July 2025.

For any questions and comments write to: vinko@ateneo.univr.it

Project description

The AlpiLinK project, official title “German-Romance Language Contact in the Italian Alps: documentation, explanation, participation”, short title “Alpine Languages in Contact”, developed in collaboration between the Universities of Verona, Trento, Bozen-Bolzano, Turin, and Aosta Valley, aimed at promoting and investigating the Germanic, Romance and Slavic minority languages and dialects spoken across the Alpine regions of Italy: Piedmont, Aosta Valley, Lombardy, Veneto, Trentino-South Tyrol, Friuli-Venezia Giulia. The project was funded by the Italian Research Ministry as ‘project of relevant national interest’ (see report title page for details). AlpiLinK documented cross-linguistically comparable data from the non-standard language varieties, e.g. dialects and minority languages, spoken in the alpine regions of Italy.

Team members

During the official three-years period the following researchers participated in five units.

University of Verona (UniVR)

	Period
Stefan Rabanus (PI)	06/2022–06/2025
Anne Kruijt	06/2022–08/2024
Alessandra Tomaselli	06/2022–06/2025
Andrea Padovan	06/2022–06/2025
Sabrina Bertollo	06/2022–06/2025
Barbara Vogt (University of Aquila)	06/2022–06/2025
Marta Tagliani	06/2022–06/2025
Fabrizio Chiarello	06/2022–06/2025
Ilaria Driussi	11/2024–06/2025

Free University of Bozen-Bolzano (UniBZ)

	Period
Birgit Alber (Unit leader)	06/2022–06/2025
Silvia dal Negro	06/2022–06/2025
Ruth Videsott	06/2022–06/2025
Alessandro Vietti	06/2022–06/2025
Alexander Glück	06/2022–06/2025
Joachim Kokkelmans	06/2022–06/2025
Angelica Bonelli	02/2024–05/2025

University of Trento (UniTN)

	Period
Ermenegildo Bidese (Unit leader)	06/2022–06/2025
Patrizia Cordin	06/2022–06/2025
Serena Bissolo	07/2024–06/2025
Michele Cosentino (University of Messina)	06/2022–03/2024
Jan Casalicchio (University of Siena)	06/2022–06/2025

University of Turin (UniTO)

	Period
Livio Gaeta (Unit leader)	06/2022–06/2025
Andriano Murelli	06/2022–06/2025
Matteo Rivoira	06/2022–06/2025
Raffaele Cioffi (University of Naples)	06/2022–06/2025
Caterina Saracco	06/2022–06/2025
Dario Capelli	01/2024–06/2025

University of Aosta Valley (UniVDA)	Period
Gianmario Raimondi (Unit leader)	06/2022–06/2025
Paolo Benedetto Mas	06/2022–06/2025
Aline Pons	06/2022–06/2025
Sara Erriu	10/2024–12/2024

Work packages

The project proposal divided the work for the project into eight work packages. See below a list and brief description of each work package and the unit(s) responsible for the work detailed in the work package.

- **WP1: Management and coordination**
This work package contains the general management of the project including initial and final operations of launch and closure of the project and the coordination of the research units.
Responsible unit: UniVR
- **WP2: Platform creation and maintenance**
This work package contains three tasks: (i) creation of the platform featuring the basic functions for data collection, storage and accessibility (on the basis of the already running VinKo platform); (ii) import of already collected VinKo data (by the end of 2021 presumably 50,000-100,000 geocoded audio recordings and connected sociolinguistic data); (iii) maintenance for guaranteeing permanent accessibility including security updates.
Responsible unit: UniVR
- **WP3: Data Management Plan and personal data protection**
In this work package we take care of two tasks: (i) elaboration of a Data Management Plan consistent with the guidelines of the European Commission for storage of the data in project-independent archive which guarantees long-term accessibility; (ii) constant surveillance of personal data protection, preparation and constant updating of privacy agreement conditions.
Responsible unit: UniVR
- **WP4: Development of tools for data extraction, management and analysis**
The action of this work package will consist of four steps: (i) evaluation of freely available resources for database management (for transcription, segmentation and annotation e.g. WebMaus services; for data management the EMU Speech Database Management System (EMU-SDMS) or the tools developed in the SPADE project SPeech Across Dialects of English such as Polyglot DB and ISCAN Python API); (ii) data pre-processing (orthographic transcription and annotation activities); (iii) organization of audio and textual information in a corpus; (iv) development of procedures that allow the extraction of linguistic features and their automatic analysis for scientific purposes.
Responsible unit: UniBZ
- **WP5: Data collection**
The work package features two tasks. In Phase 1 the questionnaires will be prepared. The elicited linguistic variables will be those of relevance for the investigation of contact induced change. They concern all structural levels of analysis. At the phonetic/phonological level we aim to elicit, among other things, syllable structure. For morphology, among other things, articles, pronouns. For syntax, among other things, determiners and proper names and spatial deixis. In Phases 3 and 4 the incoming data will be constantly controlled in order to put only quality-checked data in the open-access Geographic Information System.
Responsible unit: all units
- **WP6: Data elaboration and analysis**
This work package constitutes the core of the linguistic module of the project: data is prepared for linguistic analyses (e.g., phonetic transcriptions, morphosyntactic glossing), then the analyses are conducted. Critical evaluation of crowdsourcing methodology also forms part of this work package.
Responsible unit: all units
- **WP7: Documentation and publication of results**
The action consists in presentation of results in workshops and conferences, and in beginning to work on publications.
Responsible unit: all units

- **WP8: Public outreach: speech-community participation**

At the beginning of the project (Phase 2) we will conduct activities favoring the participation of speech-community members at data crowdsourcing, in concert with local stakeholders. In the main phase the aspect of restitution of the data to the speech communities will constantly become more important. We will promote the usage of resources prepared with and for the speech-community members to increase the visibility of dialects and minority languages and the awareness of multilingualism (see section 4 for details on social impact). In this work package all the research units will contribute with the amount of person months that corresponds to the number of varieties they are responsible for, considering also the size of the speech communities.

Responsible unit: all units

Project timeline

The original timeline as proposed by the project funding proposal was stated as such: The project will be organized in four phases through which we will achieve our objectives as analytically described in the work packages. These phases are briefly summarized; the hypothesized timeline is depicted in Figure 1.

- Phase 1: In the kick-off phase, covering the first six months, the internet platform will be activated, recreating on University of Verona's virtual machines the system running on the VinKo platform. VinKo data will be imported. In parallel the data collection will be prepared by revising the VinKo questionnaires and designing new stimuli, including the free speech production task.
- Phase 2: In the second phase data collection will be promoted in collaboration with the local and regional stakeholders. In parallel (i) the Data Management Plan will be developed and action to secure the protection of the incoming personal data will be taken, (ii) the development of tools for data extraction, management and analysis will start.
- Phase 3: The main phase focuses on linguistic analysis and the beginning of the dissemination of scientific results (mainly conference participation). Furthermore, interaction with local speech communities shifts from data collection to data restitution. Development of tools will go on.
- Phase 4: While analysis and dissemination will continue, the final phase will feature special action to make sure that the GIS platform continues to work after the end of the project funding and that the collected data will be accessible in general repositories, independently from our platform. An event organized together with the representatives of the speech communities will conclude the project.

The project has been conducted using the timeline detailed above. In Figures 2, 3, 4 please see the timeline for how it was realized.

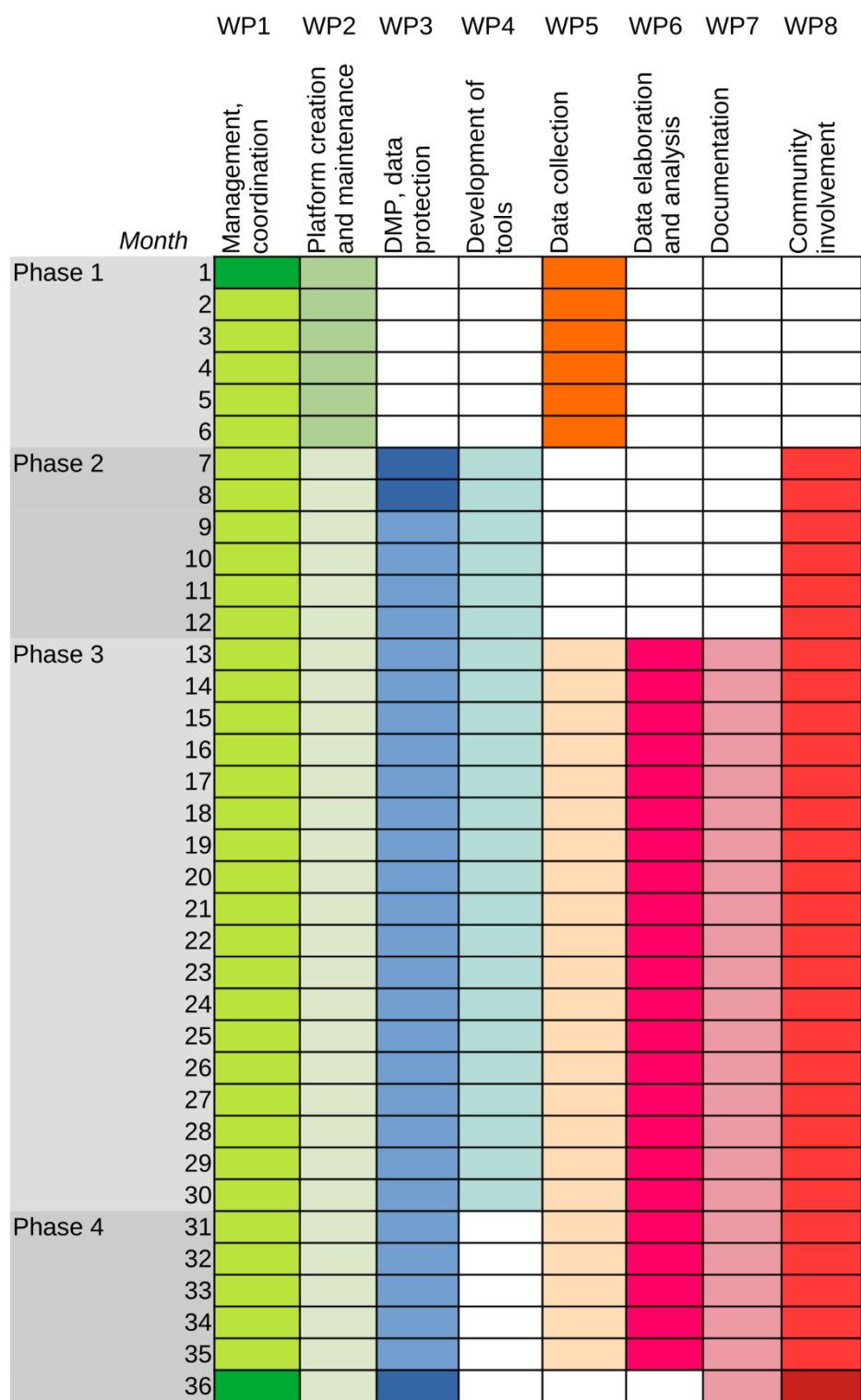


Figure 1: PRIN proposal timeline.

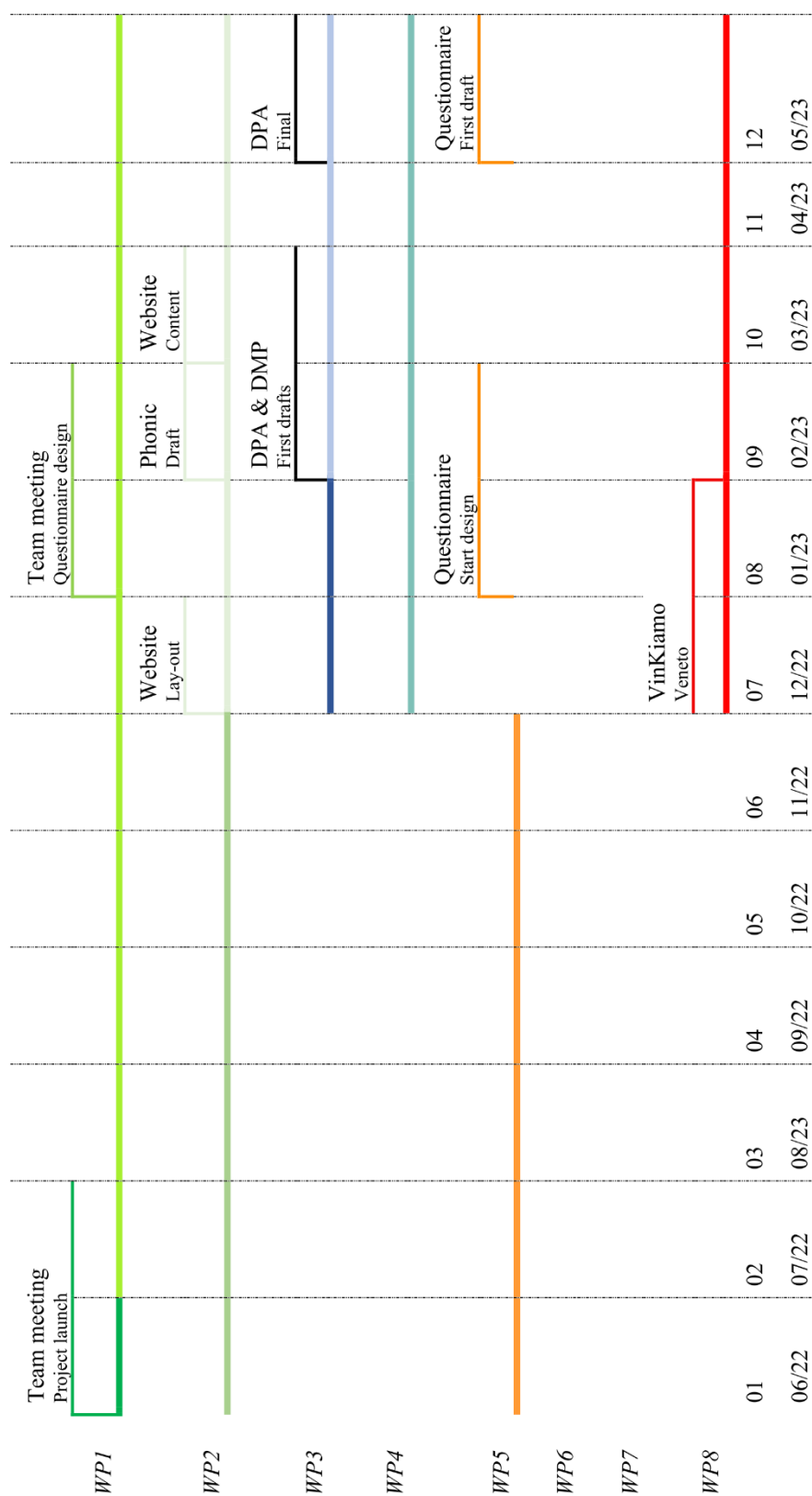


Figure 2: Realized timeline, Phases 1 & 2.

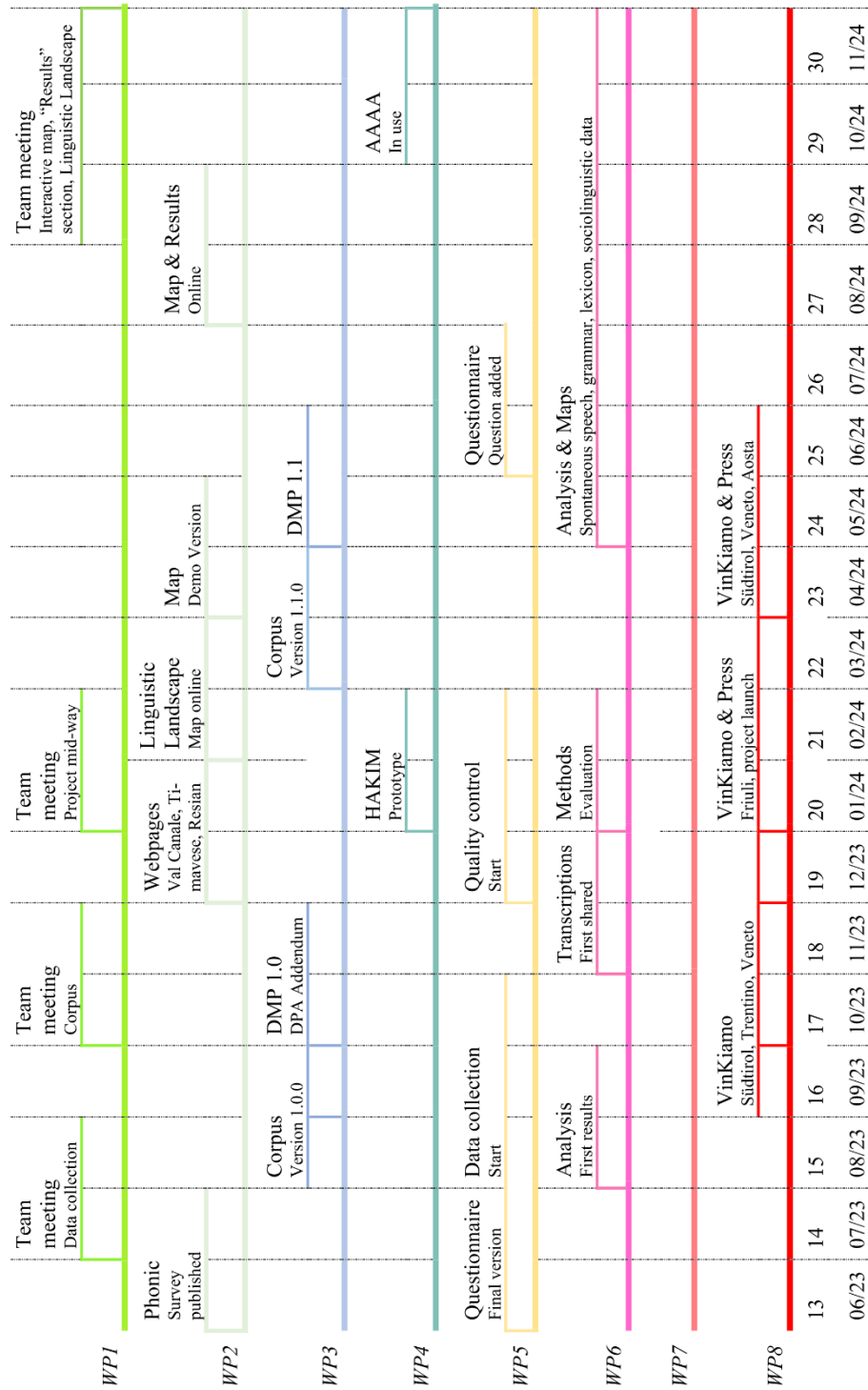


Figure 3: Realized timeline, Phase 3.

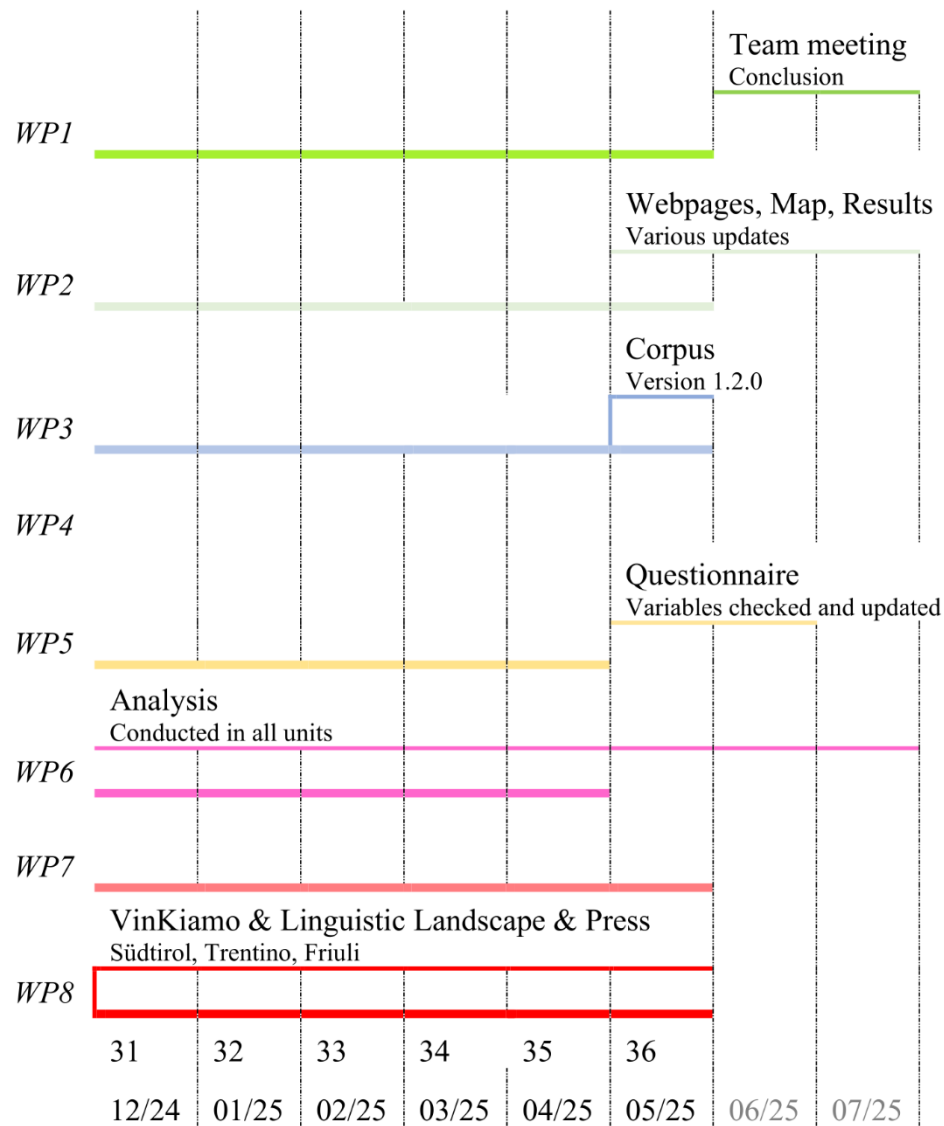


Figure 4: Realized timeline, Phase 4.

WP1. Management and coordination

This work package contains the general management of the project including initial and final operations of launch and closure of the project and the coordination of the research units. To guarantee optimal interchange and collaboration among research units and work packages we recruited a highly qualified young researcher (PhD in the subject area) for a “RTD-A” position. Since the RTD-A left her position in 2024, we recruited a specialist in language policies and interaction with the schools on a research-contract position (“AdR”), especially for taking care of *WP8.1 VinKiamo*.

Team	Period	Unit
Stefan Rabanus	06/2022–06/2025	UniVR
Anne Kruijt (RTD-A)	06/2022–08/2024	UniVR
Ilaria Driussi (AdR)	11/2024–06/2025	UniVR

Much of the work in this work package is reflected in the outcomes of WP2 and WP8.2, especially in the creation, coordination and maintenance of contracts, financial streams, and subscriptions to the various external companies hired for parts of the work (e.g., Blum, Phonic, Chabrea d’Oc).

WP1.1 Team communication

Communication and dissemination within the team was done via team meetings (either online or in-person) and a newsletter (three times a year) to update everyone on progress in the various aspects of the project and to coordinate and agree on future developments. The table below provides a very brief overview of the meetings held and the topics discussed.

Meetings	Topic	Place
27 June 2022	Launch	Verona
31 January 2023	Questionnaire design	Verona
3 July 2023	Launch data collection: website and Phonic	Online
19 October 2023	AlpiLinK Corpus and how to use it	Online
30 January 2024	Project progress, Press release	Verona
17 July 2024	Team meeting of the Verona unit	Verona
26 November 2024	Project progress and update, new team members, audio data collection, presentation of: new interactive map, “Results” section on website, “Third Mission” results and “Linguistic Landscape” project	Online
16 June 2025	Conclusive meeting: results of the project so far and perspectives for continuation	Verona

The internal newsletter went out once every 4 months, below a list of the newsletters.

	Publication date	Period
Newsletter 1	18 July 2023	01/06/2023–18/07/2023
Newsletter 2	23 November 2023	19/07/2023–15/11/2023
Newsletter 3	26 March 2024	16/11/2023–15/03/2024
Newsletter 4	25 July 2024	16/03/2024–15/07/2024
Newsletter 5	17 November 2024	16/07/2024–15/11/2024
Newsletter 6	17 March 2025	16/11/2024–15/03/2025
Newsletter 7	18 July 2025	16/03/2025–15/07/2025

The information in the newsletter was supplied by each unit, the collection of information was coordinated by Andrea Padovan, the release of the newsletters was taken care of by Stefan Rabanus. The information collected in the newsletters provided important information for this report.

WP1.2 Public communication

Direct communication with the public has been handled via the project email address: vinko@ateneo.univr.it. Any queries, comments or feedback from outside come in and are answered via this email address. This email address was (and continues to be) managed by Stefan Rabanus.

After the project launch and press release in January 2024, the email was used to reply to the feedback and to keep in touch with the cultural and linguistic associations in the area (e.g., Istituto di Cultura e Lingua Resiana, Magnifica Comunità di Fiemme, Federazione Ladina del Veneto, Presidenza della Provincia di Belluno, Agenzia Regionale per la lingua friulana, Wikimedia Italia) and with interested students, outside researchers and minority language activists.

The public-relations management was taken care of by Blum, see *WP8.2 Public relations*.

WP1.3 Contracts

For certain parts of the project the services of external companies were employed. Before entering into contracts, cost estimates/quotes for the work are requested from at least 3 different companies. Based on specialization and cost estimate the most suitable company was selected and retained.

- Immagine più (2023), theme of the Wordpress website, see *WP2.1 Word Press: Website*.
- Officina Grafica Eco (2023), tote bags for project promotion, see *WP8.1 VinKiamo*.
- Phonic (2023–2025), online survey software, see *WP2.2 Phonic: Data collection*.
- Blum (2023–2025), public communication and press releases, see *WP8.2 Public relations*.
- VRpromo (2024), roll-up for conference participation, workshops on AlpiLinK topics and VinKiamo activities, see *WP8.1 VinKiamo*.
- Chabmbra d'Oc (2024–2025), development of the interactive map, see *WP2.3 Interactive map*.

WP1.4 Project partners

AlpiLinK collaborates with the following research projects.

	Contact	Area
VerbaAlpina	Thomas Krefeld & Stephan Lücke	Linguistic maps; alpine area; Romance and Germanic lexicon
Regionalsprache.de	Alfred Lameli	Linguistic maps; incorporation of AlpiLinK audio in REDE
Lingscape	Christoph Purschke	Linguistic landscaping; map Linguistic Landscape
CliMALp	Livio Gaeta	Piemonte, Valle d'Aosta
Eurac Research	Greta Franzini	Dialect-speech recognition for South Tyrol

WP2. Platform creation and maintenance

This work package contained three tasks: (i) creation of the platform featuring the basic functions for data collection, storage and accessibility (departing from the platform of the previous VinKo project); (ii) import of the data collected by the VinKo project and stored in the interactive map on the VinKo platform (65.415 geocoded audio recordings); (iii) maintenance for guaranteeing permanent accessibility including security updates.

The tasks of this work package were incorporated into the four following elements:

- Wordpress website: the website is the main platform for dissemination and access to the linguistic questionnaire and its results. The website has been online since May 2023.
- Phonic survey: the software used to collect linguistic data using an online questionnaire. The “Participate” section of the AlpiLinK website, that redirected to the Phonic survey, was open for data collection from Juli 2023 to June 2025.
- Interactive map: the map was released in August 2024 in the “Listen & Explore” section of the website. It gives access to more than 110,000 audio recordings, roughly half of them from the AlpiLinK project, the other half from the previous VinKo project.
- Linguistic landscaping: data on linguistic landscaping is collected via a form on the AlpiLinK website and stored and displayed via the Lingscape map of the University of Luxembourg which is embedded in the “Linguistic Landscape” section.

WP2.1 Word Press: Website

From consultation with the technical staff (Marco Rospocher, Fabrizio Chiarello) at the University of Verona, it was clear that the best option for the realization of a website that was easy to maintain, and edit was using the University’s Wordpress server. As a result, this software was adopted for the creation of the public website. It was and partly is still used for the project communication with the general public, recruitment of new participants, showcasing subprojects, e.g. VinKiamo (see *WP8.1*), and access to the collected data via the interactive map (see *WP2.3 Interactive map*).

In May 2023 the public website was officially launched and published via <https://alpilink.it/>. Responsibility for the creation and maintenance of the website lies primarily with the Verona unit. The following list outlines the main contributors and their tasks.

Team	Responsibilities	Unit
Stefan Rabanus	Content, technical additions, German, English, Italian texts	UniVR
Anne Kruijt	Content, technical additions, English texts	UniVR
Fabrizio Chiarello	Platform maintenance, technical changes	UniVR
Marta Tagliani	Correction of Italian texts	UniVR
Joachim Kokkelmans	Translation of texts into French	UniBZ
Ilaria Driussi	Translation of texts into German, English, French	UniVR

For the creation of the content for the sections “Our varieties”, all units have been involved. Please see *Appendix 2.1 Website* pages for a full list.

The creation of the lay-out of the website was outsourced to an external company (Immagine più). They delivered the branding of the AlpiLinK in the form of a project logo (see title page), project colors, font and website lay-out in the form of a personalized Impreza Wordpress plug-in. The project was delivered in December 2022 in the form of a pdf file with high-definition images of the logo. Subsequently the Impreza plug-in was activated on Wordpress.

Unit	Cost
UniVR	€ 3,129 (VAT included)

Note on images used on the website: All images displayed on the website are either taken by a member of the team or licensed in the public domain or under a creative commons license. Details are present in the media section of Wordpress in the title of each image.

The majority of the AlpiLinK pages are static pages, which are, however, continuously updated (see *Appendix 2.1*). They are available in Italian, German, English and French. The “Results” section features posts (Italian: *articoli*), automatically added to the “carousel” for all languages it has a version in. At the time of writing (July 2025) there are three posts and one draft which has not been published for political reasons, see below.

Title	Languages	Author	Date	State
Let (it) snow: expletives with weather verbs	Italian, German, English	Anne Kruijt	29.July 2024	Published
Given Names and Expletive Articles in Trentino-South Tyrol and Veneto	Italian, German, English	Stefan Rabanus	1 August 2024	Published
Two years of data collection	Italian, German, English, French	Stefan Rabanus	30 May 2025	Published
Dialect in Google Translate	Italian	Anne Kruijt	29.July 2024	Draft

WP2.2 Phonic: Data collection

Phonic is a software specialized in the collection of online questionnaires including audiovisual data operated by an American-based company (Infillion). Based on advice from the IT department (Fabrizio Chiarello and Marco Rospocher) of the Lingue department of the University of Verona, the company was selected as a cheaper alternative to in-house software development.

The company works with subscriptions, with different plans that determine the available features, the number of people that need access to the account and the number of credits per month. One credit equals a single questionnaire being filled out. Academic users get a 50 % discount on their plans. The questionnaire was developed in the Starter plan which has a limited number of credits. Once it was published, we switched to a Premium plan with 500 credits a month, with only one account having access to the questionnaire.

Period	Plan	Credits per month	Costs
14 February 2023–14 February 2024	Starter	20	\$ 241
12 July 2023–12 July 2024	Premium (Academic)	500	\$ 818.57 (total cost of \$ 1,014 minus \$ 195.43 remaining value of the Starter plan)
12 July 2024–12 July 2025 (<i>de facto</i> end of service: 30 June 2025, see below)	Premium (Academic)	500	\$ 1,014

In compliance with national regulations (Law 136/2010, Art. 3, traceability of financial flows), a traceability form has been signed. In compliance with the Data Processing of the EU GDPR, a Data Processing Addendum has been signed between Infillion and the University of Verona (see *WP3. Data Management Plan and personal data protection*).

In February 2023 the first draft surveys were made and in June 2023 the final version of the survey was launched on the “Partecipate” section of the AlpiLinK website by embedding the Phonic questionnaire. During the project the survey has been continually checked and updated for typos, explanations have been clarified, new language varieties have been added (Timavese, Val Canale German/Slovenian, Resian), technical issues regarding the survey’s internal syntax has been fixed. Data collection ended on 30 June 2025 because Phonic closed the service on this day for all customers.

For a detailed account of how Phonic was used for checking incoming questionnaires, the creation and adaptation of the questionnaire and exporting the collected data, please see *Appendix 2.2 How to use Phonic*.

WP2.3 Interactive map

To make the AlpiLinK data easily available to both the general public and outside researchers, an interactive map was created, following the VinKo example. It allows users to navigate a large part of the AlpiLinK data (26 May 2025: 47,699 audio recordings) via a dropdown menu with stimuli. Any audio data connected to the selected stimulus is then plotted on a geographical map of the research area. Additionally, the map features the VinKo data (65,415 audio recordings) which were previously available at the (now disabled) VinKo map.

Team	Description	Unit
Anne Kruijt	Coordination, VinKo data adaptation	UniVR
Stefan Rabanus	Coordination	UniVR
Carlo Zoli	Coordination, conception map	UniBZ

The development of the map is being done via *Chambra d'Oc*, a non-profit organization specializing in minority languages and the development of linguistic resources for endangered languages (contact: Silvia Randaccio). The costs for the development of the map are presented below.

Unit	Costs
UniVR	€ 4,472 (VAT included)
UniBZ	€ 3,432 (VAT included)

A demo version had been created by April 2024. The final version of the map was published on the AlpiLinK website on 21 August 2024 (with small changes made subsequently, see table below for the exact timeline). The map is hosted on the servers of the data centre of the Free University of Bozen-Bolzano dedicated to linguistic minorities. This server already hosts similar projects and potential databases that could be connected, and Carlo Zoli is already a collaborator of them. This reduces the bureaucracy and technical issues that would have to be resolved if the map was moved to the servers of the University of Verona. On the AlpiLinK website, the map is embedded in a Wordpress iframe.

Meetings	Participants	Topic
26 April 2024	Silvia Randaccio, Carlo Zoli, Anne Kruijt	Showcase first demo version of the map
18 July 2024	Silvia Randaccio, Carlo Zoli, Anne Kruijt	Progress on the map, final version programmed before Ferragosto
20 August 2024	Stefan Rabanus, Carlo Zoli, Anne Kruijt	Small adjustments final version map

Categories in dropdown menus, with in total 96 selection options

The map features only the stimuli for which audio recordings for all varieties of the area were collected.

- Frasi/Sätze: 45 options (43 stimuli, 28 from AlpiLinK and 15 from VinKo).
- Storie/Geschichten: 34 options (34 stimuli, all from VinKo).
- Suoni/Laute: 10 selections (99 stimuli, all from VinKo).
- Immagini/Bilder: 7 options (7 stimuli, all from AlpiLinK).

For full overview of which stimuli are present in which part of the menu, please see *Appendix 2.3 Map, included stimuli*.

Choices made regarding the map formatting

- Satellite layer (showcases geographical features nicely).
- Outline of municipalities for which there is at least one audio file.
- When hovering on the outline, colored in with variety color, and name of municipality.
- Secondary filter with language variety.

Color codes for the map

	IT	DE	EN
#fee64f	Cimbro	Zimbrisch	Cimbrian
#4292c6	Francoprovenzale	Frankoprovençalisch	Francoprovençal
#deebf7	Friulano	Friulanisch	Friulian
#08519c	Ladino	Ladinisch	Ladin
#9ecae1	Lombardo	Lombardisch	Lombard
#fee64f	Mòcheno	Fersentalerisch	Mòcheno
#2171b5	Occitano	Okzitanisch	Occitan
#6baed6	Piemontese	Piemontesisch	Piedmontese
#238b45	Resiano	Resianisch	Resian
#fee64f	Sappadino	Plodarisch	Sappadino
#fee64f	Saurano	Sauranisch	Saurano
#238b45	Sloveno della Val Canale	Kanaltal-Slowenisch	Timavese
#fee64f	Tedesco della Val Canale	Kanaltal-Deutsch	Tyrolean
#fee64f	Timavese	Tischelwangerisch	Trentino
#fff7bc	Tirolese	Tiroler Dialekt	Val Canale German
#08306b	Trentino	Trentinisch	Val Canale Slovenian
#c6dbef	Veneto	Venetisch	Venetan
#fee64f	Walser	Walserdeutsch	Walser German

WP2.4 Linguistic Landscape

The section “Linguistic Landscape” is used to capture the expression of local language that are not auditory in nature, but visual, e.g. street signs, posters.

For this part of the project, we collaborate with the Lingscape (Citizen science meets linguistic landscaping; <https://lingscape.uni.lu/>) team at the University of Luxembourg (coordinator: Christoph Purschke). They have extensive experience with the capture of linguistic landscapes and have developed digital tools for documentation, storage, and representation. All materials collected in the AlpiLinK project therefore are subject to the terms and conditions of Lingscape:

<https://lingscape-app.uni.lu/mobile/en/terms.html>

Data collection is done via a contact form on the AlpiLinK page, which emails the filled-out information and image to an email address (vinko@ateneo.univr.it). The collected data is put into the Lingscape archive and plotted on a project-specific Lingscape map. This map is embedded in the AlpiLinK section “Linguistic Landscape” since February 2024. In May 2025 the Lingscape map underwent a complete technical update.

See *Appendix 2.4 Linguistic Landscape* for the full workflow applied to this part of the project.

WP3. Data Management Plan and personal data protection

In this work package we took care of two tasks: (i) elaboration of a Data Management Plan (DMP) consistent with the guidelines of the European Commission for storage of the data in a project-independent archive which guarantees long-term accessibility; (ii) creation of a Data Processing Agreement (DPA), constant surveillance of personal data protection, preparation and constant updating of privacy agreement conditions.

Team	Tasks	Period	Unit
Anne Kruijt	Structure & upload repository, creation & upkeep DMP	12/2022–08/2024	UniVR
Stefan Rabanus	Creation DPA	12/2022–06/2025	UniVR

WP3.1 Data Management Plan (DMP)

For the project, a DMP was developed that ensures the proper storage and processing of the collected data in line with the guidelines of the European Commission for Data Protection and Privacy regulations. It also outlines the processing of the data to make our data as FAIR (Findable, Accessible, Interoperable and Reuseable) as possible.

The first version of the DMP was delivered on 10 October 2023 and it was updated into version 1.1 on 28 May 2024. Please find the DMP in *Appendix 3.1 Data Management Plan*.

WP3.2 Data Processing Agreement (DPA)

The first phase of the project was used to design the DPA for participants which was needed for the start of the data collection.

In February 2023, the first draft was sent to the Privacy Office of the University of Verona for approval. In May 2023 the final version of the DPA was finished:

https://alpilink.it/wp-content/uploads/2023/05/TrattamentoDati_Datenverarbeitung_it-de.pdf

In October 2023 an additional Addendum for the DPA was signed with Phonic.

For the processing and sharing the data we used the Creative Commons license Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA). The “noncommercial” specification turned out to be too restrictive, even for sharing the data with scientific project partners or making the audio recordings available on Wikipedia pages. Hence, we intend to use a simple CC BY license in future projects.

WP3.3 AlpiLink repository

As specified in the DMP, all data collected in the AlpiLink project is openly published. The data was published on a monthly or bimonthly basis on Zenodo (<https://about.zenodo.org>). All versions can be accessed and cited using the following link: <https://zenodo.org/doi/10.5281/zenodo.8360169>.

Date	Version	Link	Participants	Audio
30 September 2023	v1.0.0	https://doi.org/10.5281/zenodo.8360170	58	2,056
30 October 2023	v1.0.1	https://doi.org/10.5281/zenodo.10054120	80	2,862
30 November 2023	v1.0.2	https://doi.org/10.5281/zenodo.10224351	240	8,025
22 December 2023	v1.0.3	https://doi.org/10.5281/zenodo.10418858	336	12,316
29 January 2024	v1.0.4	https://doi.org/10.5281/zenodo.10533936	424	15,538
28 February 2024	v1.0.5	https://doi.org/10.5281/zenodo.10715916	723	25,840
28 March 2024	v1.1.0	https://doi.org/10.5281/zenodo.10852939	813	29,153
29 April 2024	v1.1.1	https://doi.org/10.5281/zenodo.11083948	942	33,951
17 June 2024	v1.1.2	https://doi.org/10.5281/zenodo.11352290	1,003	35,502

31 July 2024	v1.1.3	https://doi.org/10.5281/zenodo.12697578	1,090	37,777
29 August 2024	v1.1.4	https://doi.org/10.5281/zenodo.13144330	1,174	39,076
4 November 2024	v1.1.5	https://doi.org/10.5281/zenodo.13488507	1,233	40,315
15 February 2025	v1.1.6	https://doi.org/10.5281/zenodo.14033457	1,474	47,080
3 April 2025	v1.1.7	https://doi.org/10.5281/zenodo.14870649	1,627	51,198
26 May 2025	v1.2.0	https://doi.org/10.5281/zenodo.15129710	1,781	57,109

Detailed instructions on how the corpus was updated can be found in *Appendix 3.3 Data export: from Phonic to AlpiLinK repository*.

Team members in charge (“corpus creators”)	Period	Unit
Anne Kruijt	09/2023–08/2024	UniVR
Angelica Bonelli	09/2024–05/2025	UniBZ

WP3.4 Recovering and restructuring of the data of the previous AThEME and VinKo projects and creation of the AThEME and VinKo repositories

In parallel with the development and implementation of the AlpiLinK DMP and DPA and the preparation of the first version of the AlpiLinK Corpus, the data from the preceding language-contact study projects AThEME and VinKo, which had no DMP at their start, were restructured and stored in repositories. The main issues in the creation of the databases for AThEME and VinKo were the lack of structure in the data and the boundaries of the consent forms signed by participants in the initial data collection. VinKo’s data was already present in a digital format and in a single sentence format, so it could be quite easily restructured by renaming files with a consistent naming convention and by adding metadata. For the AThEME project this was not the case and therefore the data preparation was very labour intensive. Long audio files had to be cut on a sentence and word-level, labelled using a consistent naming convention, the written data had to be transcribed, and metadata had to be added. The second issue was informed consent of participants for this new use of the data. VinKo’s consent form did permit the reuse of the collected audio data for the creation of a scientific database and therefore posed no challenges. The consent form of the AThEME project expressly forbid the reuse of data outside of the AThEME project. Participants had to therefore be contacted with the request to sign a new consent form which would allow for this reuse. For some participants consent could not be gotten (either due to being unable to contact a participant or because a participant was deceased), and their data was therefore excluded from the corpus.

VinKo (Varieties in Contact) Corpus

Date	Version	Link	Participants	Audio
July 2021	v1.0	http://hdl.handle.net/20.500.12124/32	394	37,806
May 2022	v1.1	http://hdl.handle.net/20.500.12124/46	573	63,863
August 2023	v1.2	http://hdl.handle.net/20.500.12124/74	1,429	189,679

AThEME Verona-Trento Corpus

Date	Version	Link	Participants	Audio
December 2022	v1.0	http://hdl.handle.net/20.500.12124/53	37	7,544

The work was supported by AdR funding from the PRIN 2017 “Models of language variation and change” whose UniVR unit was coordinated by Alessandra Tomaselli.

Corpus creator	Period	Unit
Anne Kruijt	06/2022–08/2023	UniVR

WP4: Development of tools for data extraction, management and analysis

The action of this work package consisted of: (i) evaluation of freely available resources for database management (for transcription, segmentation and annotation e.g. WebMaus services; for data management the EMU Speech Database Management System [EMU-SDMS] or the tools developed in the SPADE project SPEECH Across Dialects of English such as Polyglot DB and ISCAN Python API); (ii) data pre-processing (orthographic transcription and annotation activities); (iii) organization of audio and textual information in a corpus; (iv) development of procedures that allow the extraction of linguistic features and their automatic analysis for scientific purposes.

Team	Description	Period	Unit
Joachim Kokkelmans	Coordination, conceptualization, programming	12/2022–11/2024	UniBZ
Birgit Alber	Consulting, testing,	12/2022–11/2024	UniBZ
Angelica Bonelli	Testing	10/2024–05/2025	UniBZ
Alessandro Vietti	Consulting, testing	12/2022–10/2023	UniBZ
Barbara Vogt	Testing	12/2022–10/2023	UniVR

All tools are available to the team members in the shared AlpiLinK folder.

WP4.1 Half-automated Annotation Kit for Interlinguistic Material (HAKIM)

This tool, written in the Praat and tcsh programming languages, was designed to let researchers transcribe and annotate the AlpiLinK (and VinKo) audio data as quickly and efficiently as possible. So-called “pretranscriptions” are first formulated, representing the orthographic form corresponding to the utterance expected to occur most frequently in the data for a given stimulus in a given linguistic variety. Researchers then hear each audio file and adapt the pretranscription if necessary to match the audio. HAKIM converts the orthography into IPA/SAMPA and uploads the transcriptions to the MAUS server: <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSGeneral>.

The server returns, for each audio file, a time-aligned phonetic transcription that can be read by Praat or ELAN. All or a filtered set of transcriptions can be exported to an Excel file and/or shared with other researchers, who can import them into HAKIM on their own computer.

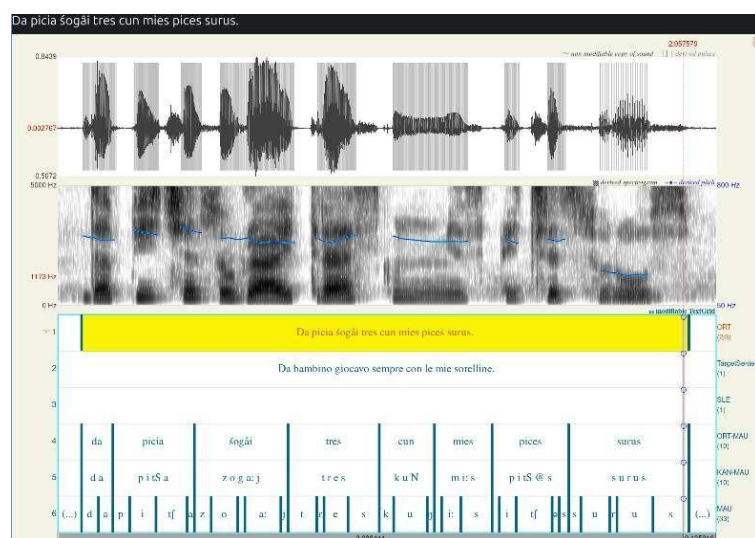


Figure 5: HAKIM tool.

The prototype of HAKIM was presented at the team meeting in Verona on 30 January 2024 and made available to all team members. Since then, the tool has been used and continuously updated.

WP4.2 Minority Language Grapheme-To-Phoneme Conversion Tool (MinorityG2P)

Some of the linguistic varieties in the AlpiLinK project (e.g., Tyrolean) use writing conventions that closely match those of a well-described standard language (e.g. German; see <https://doi.org/10.13092/lo.127.11087>). In this case, HAKIM uses the online G2P service offered by the developers of MAUS to convert orthographic input into IPA/SAMPA.

<https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/Grapheme2Phoneme>

Other languages in the AlpiLinK data, however, have their own separate writing conventions (e.g. Val Badia Ladin). In that case, HAKIM uses the MinorityG2P tool. This tool, written in Javascript and HTML, allows to define these writing conventions as a series of substitutions that automatically transform orthography into IPA and SAMPA. Lexical exceptions can be listed, references to word boundaries and phonological classes can be made (e.g. “<n> becomes velar [ŋ] at the end of a word after a vowel”) and numbers are converted to full words (e.g. <31> becomes [θɜ:tiwʌn]). As of today, the substitutions have been defined for Val Badia Ladin and Mòcheno.

WP4.3 Automated Audio Acceptability Assessment Tool (AAAA)

This Praat-based tool makes the quality check of the audio files in the AlpiLinK corpus faster and more efficient. For each audio file, it measures its total duration, average intensity and intensity peak difference. It then calculates by how much these three measurements deviate from the “prototypical” audio file of this category (formulated as a z-score in statistics) and ranks all audio files to be verified from largest to lowest deviation (i.e. by “suspiciousness”). The researcher then hears each audio fragment and decides in a Praat interface whether to validate this audio file as such, to trim the long silences it contains or to remove it (if unsuited for the corpus). The researcher can export the list of validated, trimmed or deleted audios to Excel.

AAAA is used for the quality check of the audio files since November 2024 (AlpiLinK Corpus 1.1.5).

WP5: Data collection

The work package featured two tasks. In Phase 1 the questionnaires were prepared. The elicited linguistic variables were those of relevance for the investigation of contact-induced change. In Phases 3 and 4 the incoming data were constantly controlled in order to put only quality-checked data in the open-access Geographic Information System. In this work package, all research units work in concert for the shared parts. For the variety-specific aspects responsibility is held by the research units assigned to these varieties.

WP5.1 Questionnaire

In the first phase of the project, the questionnaire and the task design were developed. Each unit contributed to the questionnaire and the task design. The complete questionnaire is stored in the AlpiLinK Corpus in the Zenodo repository. Please see also *Appendix 5.1 Questionnaire* which features stimulus IDs, target sentences and the specific features the stimuli were created to elicit.

The structure of the questionnaire was discussed and decided during the team meeting in Verona on 31 January 2023. The pilot of the questionnaire was done in May 2023. The completed version of the questionnaire came online in June 2023. The last content update of the questionnaire was on 21 June 2024 when the question about how people found the questionnaire was added. In May 2025 Stefan Rabanus and Andrea Padovan performed a systematic check of the labels of the variables of the linguistic questionnaire, corrected incoherences and prepared the questionnaire file for the publication of the AlpiLinK Corpus 1.2.0.

In the general and sociolinguistic part of the questionnaire speakers are asked to supply the following information:

- Linguistic variety (multiple choice, open field).
- Location: province, municipality, and if they still live there.
- Age, gender.
- If they speak their variety well, if they speak it often, if they speak it with friends and family, if they write their variety (and if so, where) and which other languages they know.
- Where/how they came across the AlpiLinK questionnaire (added in June 2024).

The linguistic questionnaire was created from input given by the various research units as presented in the following table.

Unit	Contribution
UniBZ	<ul style="list-style-type: none">▪ Focus varieties: Ladin and German varieties▪ Word formation in Tyrolean of iterative/collective nouns (G01-G10), <i>Ge-...-erei</i>, use and phonology. Task design: sentence completion, audio.▪ Personal-name truncation, variables gender and age (N01-N19). Task design: forced choice, no audio.▪ Free speech (M01). Task design: free speech, written prompt, audio.
UniTN	<ul style="list-style-type: none">▪ Focus varieties: Romance and Germanic varieties in the regions of Veneto, Trentino-South Tyrol and Friuli Venezia Giulia.▪ Phrasal verbs, presence or absence of particle (I01-I05, I07). Task design: image description, audio.▪ Deixis, demonstratives, proximal and distal (S02, S05-S08, T03). Task design: translation and tense transformation, audio.▪ Evaluative morphology (S15-S16, S22). Task design: translation, audio.▪ Indefiniteness (S09, S19-S21, S28-S29, T04). Task design: translation, audio.▪ Possessives (S06-S08, S12, S15-S18, S22, T01). Task design: translation, audio.▪ Comparatives (S05, S11, S13). Task design: translation, audio.▪ Subordinate clauses (S11-S12, S14). Task design: translation, audio.

UniVR	<ul style="list-style-type: none"> ▪ Focus varieties: Romance and Germanic varieties in the regions of Veneto and Trentino-South Tyrol. ▪ Coordination: combining and checking of stimuli, adding structures and IDs. ▪ Methodology: task design, limiting linguistic transfer via translation effects. ▪ Phrasal verbs, presence or absence of particle (I01-I05, I07). Task design: image description, audio. ▪ Personal names, presence or absence of definite article (S01, S08, S10-S12, S14, S23-S24, T02, T05-T06). Task design: translation and tense transformation, audio. ▪ Personal pronouns (S01-S30, I01-I07, T01-T06), subject, object and weather expletives. Task design: translation, image description, and tense transformation, audio. ▪ Possessives (S06-S08, S12, S15-S18, S22, T01). Task design: translation, audio. ▪ Verb complex (T01-T02). Task design: tense transformation, audio. ▪ Auxiliary selection weather verbs (S30, T03). Task design: translation and tense transformation, audio. ▪ Negative concord, negation (S28-S29, T04). Task design: translation and tense transformation, audio.
UniTO	<ul style="list-style-type: none"> ▪ Focus varieties: Romance varieties in Piemonte and Valle d'Aosta and Walser German. ▪ Pilot version of questionnaire conducted during field work (15-17/05/2023). ▪ Deverbal nouns (S25-S26). Task design: translation, audio. ▪ Progressives (I06, S24). Task design: image description task and translation, audio. ▪ Presentative (S23). Task design: translation, audio. ▪ Deixis, demonstratives, proximal and distal (S02, S05-S08, T03). Task design: translation and tense transformation, audio. ▪ Comparatives (S05, S11, S13). Task design: translation, audio. ▪ Subordinate clauses (S11-S12, S14). Task design: translation, audio. ▪ Negative concord, negation (S28-S29, T04). Task design: translation and tense transformation, audio. ▪ Interrogative clauses (S03-S04, S10, S12, S23, S27). Task design: translation, audio. ▪ Causative (S26). Task design: translation, audio. ▪ Verb clause and inflection (expression of need, imperative, infinitive/participle opposition) (S02-S04, S09). Task design: translation, audio. ▪ Left-dislocation (S03-S04). Task design: translation, audio.
UniVDA	<ul style="list-style-type: none"> ▪ Focus varieties: Romance varieties in Piemonte and Valle d'Aosta and Walser German. ▪ Deixis, demonstratives, proximal and distal (S02, S05-S08, T03). Task design: translation and tense transformation, audio. ▪ Comparatives (S05, S11, S13). Task design: translation, audio. ▪ Subordinate clauses (S11, S14). Task design: translation, audio. ▪ Negative concord, negation (S28-S29, T04). Task design: translation and tense transformation, audio. ▪ Interrogative clauses (S03-S04, S10, S12, S23, S27). Task design: translation, audio. ▪ Verb clause and inflection (expression of need, imperative, infinitive/participle opposition) (S02-S04, S09). Task design: translation, audio. ▪ Left-dislocation (S03-S04). Task design: translation, audio.

WP5.2 Data collection

The questionnaire collected the first data in July 2023; data collection ended on 30 June 2025. The results of data collection as per 26 May 2025 (AlpiLinK Corpus 1.2.0) are provided below. The final version of the AlpiLinK Corpus will be published by the end of 2026, considering the quality control of audio recordings and informant data which is conducted during the ongoing analysis of the data.

	Participants	Average age	M %	F %	O %
Cimbrian	42	50.5	54.76 %	45.24 %	0.00 %
Francoprovençal	130	51.8	50.00 %	50.00 %	0.00 %
Friulian	104	47.5	39.42 %	60.58 %	0.00 %
Ladin	160	41.1	41.88 %	58.13 %	0.00 %
Lombard	189	48.4	62.96 %	36.51 %	0.53 %
Mòcheno	25	46	32.00 %	68.00 %	0.00 %
Occitan	110	60.2	55.45 %	43.64 %	0.91 %
Piedmontese	201	54.8	56.22 %	42.79 %	0.00 %
Resian	12	66.5	16.67 %	83.33 %	0.00 %
Sappadino	9	55.2	33.33 %	66.67 %	0.00 %
Saurano	1	73		100.00 %	0.00 %
Timavese	10	69.1	20.00 %	80.00 %	0.00 %
Tyrolean	262	43.6	40.46 %	56.49 %	3.05 %
Trentino	285	46.8	49.82 %	49.12 %	1.05 %
Val Canale German	6	59.3	50.00 %	50.00 %	0.00 %
Val Canale Slovenian	1	69		100.00 %	0.00 %
Venetan	212	46.8	56.13 %	43.40 %	0.47 %
Walser German	22	74.2	36.36 %	63.64 %	0.00 %
Overall	1,781	48.9	49.52 %	49.58 %	0.90 %

From a look at the distribution of data coming in, the majority of the data collected was directly correlated to outreach activities, such as VinKiamo or press releases. In the months that activities were running (e.g., November, January), there was a relatively large increase in data, whereas in the summer months there was much less coming in.

The manner of data collection varied by variety and region. It included VinKiamo activities (see *WP8.1*), fieldwork done by team members or student assistants, personal contacts, press releases (see *WP8.2*), etc. This section only lists fieldwork and VinKiamo activities, as their quantitative effects on data collection can be monitored more easily. The effects of press releases, e.g. articles or interviews, are not as easily measurable, so they are not represented here. A full list of press related activities can be found in *WP8.2 Public relations*.

Fieldwork

Fieldworker(s)	Period	Variety	Location(s)	Participants	Unit(s)
Livio Gaeta	10/2022 02/2023 05/2023	Walser German	Gressoney Saint-Jean, Gressoney La Trinité	n/a	UniTO
Michele Cosentino	07/2023	Cimbrian	Lusérn	10	UniTN
Silvia Dal Negro	08/2023	Walser German	Formazza	n/a	UniBZ
Flavia Guerini	12/2023– 08/2024	Lombard	various locations	> 42	UniVR
Riccardo Ferracin	12/2023– 06/2025	Cimbrian	Ljetzan/Giazza	7	UniVR
Livio Gaeta	03/2024– 07/2024	Walser	Formazza, Issime	4	UniTO

Serena Bissolo	09/2024, 03/2025	Cimbrian	Lusérn	16	UniTN
Matteo Rivoira, Luca Poetto, Rosella Pellerino, Tiziana Gallian, Agnès Garrone, Daniele Dalmaso	12/2024– 02/2025	Piedmontese, Occitan	various locations	38	UniTO
Serena Bissolo, Angelica Bonelli, Istituto Culturale Mòcheno	02/2025, 03/2025	Mòcheno	Garait, Vlarotz, Palai en Bersntol	14	UniBZ UniTN
Ilaria Driussi	04/2025– 05/2025	Val Canale German	Malborghetto, Tarvisio	4	UniVR
Livio Gaeta	05/2025	Walser German	Gressoney	n/a	UniTO

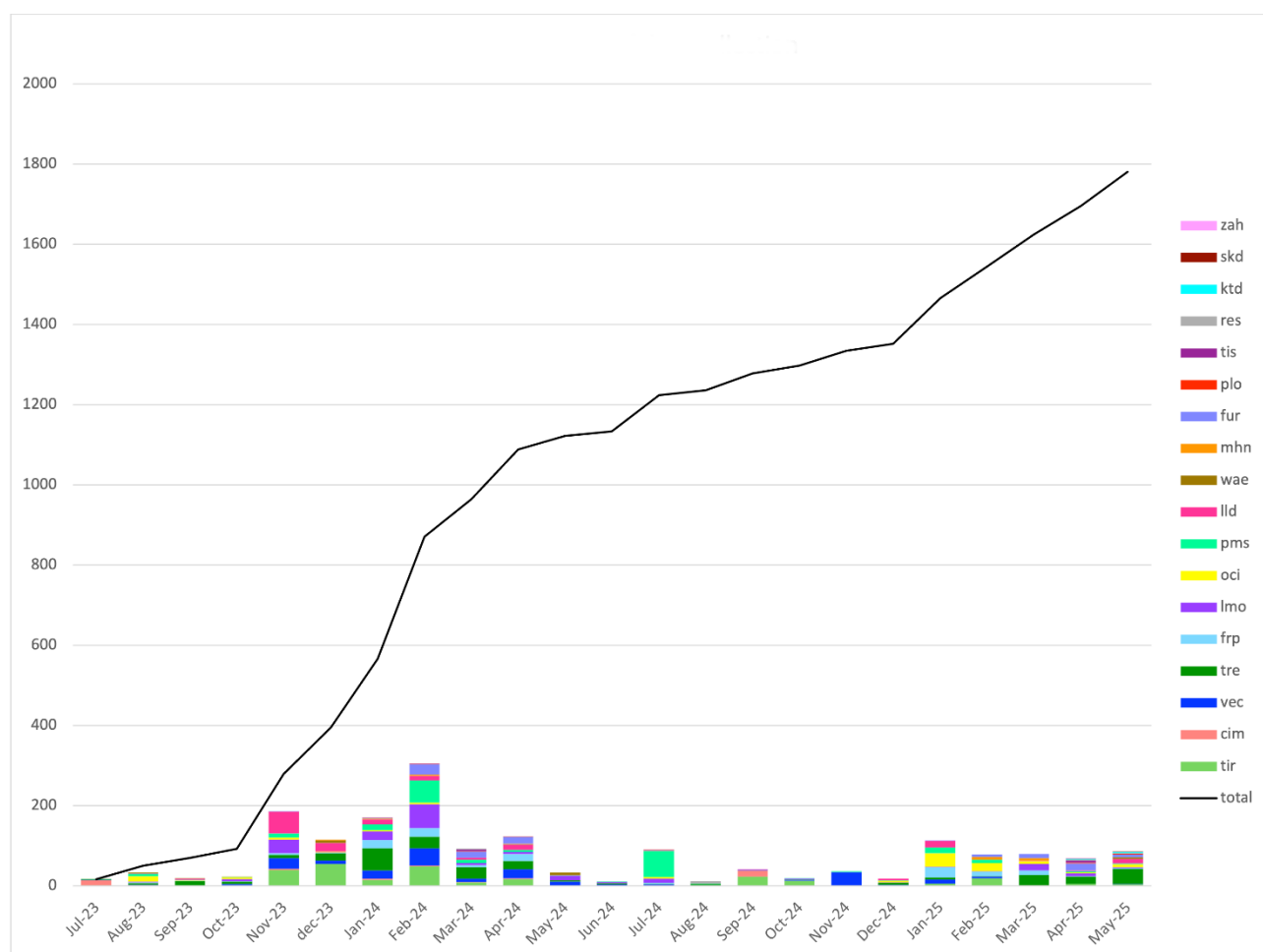


Figure 6: AlpiLinK data collection (July 2023 to May 2025).

VinKiamo activities

Unit	Period	Activity	Variety	Location(s)	Participants
UniVR	10/2022– 01/2023	VinKiamo.3	Venetan, Trentino, Tyrolean, Saurano, Sappadino, Ladin	Verona	120
UniVR	03/2023	Linguistic autobiography	Venetan, languages of immigration	Verona, Vicenza, Treviso	120
UniBZ	10/2023	Tag der histori- schen Mehrspra- chigkeit/De dl multilinguism storich, 1 st edition	Tyrolean, Ladin	Brixen/ Bressanone	170
UniTN	10/2023– 05/2024	VinKiamo Trentino, 1 st edition	Trentino, Ladin (Fassan), Mòcheno, Cimbrian (Lusérn)	Trento, Mezzo- lombardo	48
UniVR	01/2024– 05/2024	VinKiamo.4 Friuli	Friulian, Sappadino, Timavese	Tolmezzo	25
UniVDA	01/2024– 06/2024	VinKiamo Valle d'Aosta	Francoprovençal	Aosta	56
UniBZ	03/2024	Tag der histori- schen Mehrspra- chigkeit/De dl multilinguism storich, 2 nd edition	Tyrolean, Ladin	Brixen/ Bressanone	+/- 25
UniBZ	10/2024– 01/2024	Tag der histori- schen Mehrspra- chigkeit/De dl multilinguism storich, 3 rd edition	Tyrolean, Ladin	Brixen/ Bressanone, St. Ulrich/ Ortisei	+/- 60
UniVR	01/2025– 05/2025	VinKiamo.5 Friuli	Friulian, Sappadino, Timavese, Saurano, Val Canale German and Slovenian, Resian	Tolmezzo	+/- 40
UniTN	01/2025– 05/2025	VinKiamo Trentino, 2 nd edition	Trentino, Ladin	San Michele all'Adige	4
UniBZ	03/2025	Tag der histori- schen Mehrspra- chigkeit/De dl multilinguism storich, 4 th edition	Tyrolean, Ladin	St. Ulrich/ Ortisei	33
UniTN	03/2025– 05/2025	VinKiamo Trentino, 3 rd edition	Trentino, Ladin	Cles	32
UniVR	03/2025– 05/2025	AlpiLinK at “eco musei”	Friulian, Venetan	Maniago, Gemona del Friuli	25

The activities listed in the table were concentrated on the promotion of the collection of audio recordings. See detailed information on all activities below in *WP8.1 VinKiamo*.

WP5.3 Audio quality control

The incoming data was controlled for quality and suitability. The process ensured that no empty files or inappropriate audio were put onto the public linguistic map. Where necessary audio files were either trimmed down or deleted entirely. The following personnel were involved in the check of the audio.

Team	Languages	Unit	Period
Anne Kruijt	Germanic minority languages, Francoprovensal, Ladin, Piedmontese, Occitan	UniVR	12/2023–08/2024
Angelica Bonelli	Tyrolean, Trentino, Friulan, Lombard, Ladin	UniBZ	03/2024–05/2025
Stefan Rabanus	Resian	UniVR	04/2024–06/2025
Ilaria Driussi	Germanic minority languages, Friulian, Venetan	UniVR	06/2025

Assistants ("150 oristi")

Marco Pettinà	All Romance varieties	University of Verona	12/2023–03/2024
Srdjan Sankovic	Venetan	University of Verona	05/2023–03/2025
Carlo Alberto Berton	Venetan	University of Verona	06/2025

The following table shows the number of files checked, deleted and trimmed per version until AlpiLinK Corpus 1.7.0. The total amounts of audio files are by approximation: since checks were done per variety, they do not strictly adhere to corpus version numbers. At the time of writing (July 2025) the quality check is still ongoing. The final version of the corpus with all audio recordings checked will be published by the end of 2026. For an overview of the workflow used in the quality checks see *Appendix 5.3 Quality control workflow*.

Version	Deleted		Trimmed		Total	Note
	#	%	#	%		
1.0.0 (12/2023) U0001-U0057	12	0.6 %	16	0.8 %	2,056	
1.0.2 (02/2023) U0058-U0240	154	2.5 %	16	0.3 %	6,051	High percentage deleted due to deletion of all audio files of 2 speakers
1.0.5 (04/2023) U0241-U0725	230	1.8 %	563	4.3 %	13,000	Controllers were more attentive on extra noises or pauses
1.1.1 (06/2023) U0725-U0943	97	1.2 %	399	5.0 %	8,000	
1.1.2 (06/2024) U0944-U1010	46	1.33 %	269	7.8 %	3,460	
1.1.3 (07/2024) U1011-U1099	23	1.05 %	215	9.8 %	2,186	
1.1.4 (08/2024) U1100-U1183	42	1.07 %	270	6.9 %	3,892	
1.1.5 (11/2024) U1184-U1233	49	1 %	517	10.5 %	4,920	
1.1.6 (02/2025) U1234-U1474	102	1.6 %	1422	23 %	6,177	High percentage trimmed due to the high number of audios
1.1.7 (04/2025) U1475-U1636	108	1.82%	301	5.09%	5,910	

WP6: Data elaboration and analysis

This work package constituted the core of the linguistic module of the project: data was prepared for linguistic analyses (e.g., phonetic transcriptions, morphosyntactic glossing), then the analyses were conducted. Critical evaluation of crowdsourcing methodology formed also part of this work package.

Only a small percentage of the data collected between July 2023 and June 2025 has been analysed by June 2025. At the time of writing (July 2025) data analysis is ongoing in all units.

WP6.1 Data elaboration

Period	Activity	Researcher(s)	Unit
10/2023–04/2024	Extraction and tagging of the data regarding name truncation for all varieties; task N01-N19.	Joachim Kokkelmans	UniBZ
11/2023–03/2024	Time-aligned non-IPA transcription (ELAN format) of audio recordings from the Walser speakers U0037 and U0038.	Silvia Dal Negro	UniBZ
02/2024–07/2024	Non-IPA transcription of audio recordings data from all Cimbrian speakers in the AlpiLinK corpus 1.1.3.	Riccardo Ferracin under the supervision of Alessandra Tomaselli	UniVR
05/2024–06/2024, 05/2025–ongoing	Extraction (with HAKIM) and tagging of the data regarding <i>Ge-...-e</i> nominalizations; Tyrolean varieties; task G01-G10.	Birgit Alber, Angelica Bonelli, Joachim Kokkelmans	UniBZ
05/2024–07/2024	Transcription of spontaneous speech (ELAN format): M01_lmo_U0113, M01_lmo_U0121, M01_lmo_U0154, M01_lmo_U0159, M01_lmo_U0166, M01_lmo_U0171, M01_tre_U0018, M01_tre_U0042, M01_vec_U0089, M01_wae_U0037, M01_wae_U0038.	Angelica Bonelli	UniBZ
10/2024–12/2024	Transcription (IPA, BREL) and morphosyntactic annotation of approx. 760 Francoprovençal audio recordings of S02-S08, S10-S11, S13-S15, S22.	Sara Erriu under the supervision of Gianmario Raimondi	UniVDA
11/2024–12/2024	Extraction of the data regarding the use of written dialect for Tyrolean varieties; sociolinguistic part of questionnaire.	Angelica Bonelli	UniBZ
05/2025–ongoing	Converting the 784 transcriptions made by Riccardo Ferracin and the 81 transcriptions made by Silvia Dal Negro to a HAKIM-compatible format with enriched tags and linguistic tiers; Cimbrian and Walser varieties, all tasks.	Joachim Kokkelmans	UniBZ

The results of the data elaboration are available to the team members in the shared AlpiLinK folder. External researchers are invited to contact the team at vinko@ateneo.univr.it to get access to the data.

WP6.2 Data analysis

Period	Data	Analysis	Researcher(s)	Unit
07/2023– 11/2023	Name Truncation: N01-N19	Analysis of the name truncation task (N01-N19). Advanced statistical analysis for all varieties (west-east, Germanic-Romance), published by Alber et al. 2024, see <i>WP7.1 Publications</i> .	Birgit Alber, Joachim Kokkelmans	UniBZ
08/2023	Image description: I01-I05, I07	Analysis of the Cimbrian data regarding the presence or absence of phrasal verbs, presented at CIDSIM 17, see <i>WP7.2 Conferences</i> .	Anne Kruijt	UniVR
07/2023– 11/2023	VinKo data	VinKo data has been analyzed and results used to create several linguistic maps for educational purposes, see <i>WP8.1 VinKiamo</i> .	Stefan Rabanus	UniVR
11/2023– 03/2024	VinKo data	Phonetic-acoustic analysis of various variables (publications in preparation).	Joachim. Kokkelmans, Birgit Alber, Alessandro Vietti, Barbara Vogt	UniBZ, UniVR
11/2023– 03/2024	Sociolinguistic data: U0003-U0014	Analysis of the sociolinguistic data of the Cimbrian speakers collected during fieldwork in July 2023, presented at CLAM2021, see <i>WP7.2 Conferences</i> .	Michele Cosentino	UniTN
11/2023– 03/2024	CliMAlp data	Written use of Walser German (publication in press).	Caterina Saracco	UniTO
05/2024– ongoing	Spontaneous speech: M01	Analysis ongoing of the contents of the spontaneous speech audio (M01) of speakers of Tyrolean, Trentino, Ladin and eastern Lombard.	Angelica Bonelli, Silvia Dal Negro, Alexander Glück, Ruth Videsott	UniBZ
04/2024– 07/2024	Word formation: G01-G10	Preliminary analysis of the data regarding <i>Ge-...-e</i> nominalizations (Tyrolean), see Alber, 6 June 2024 in <i>WP7.2 Conferences</i> .	Birgit Alber	UniBZ
06/2024	Translation task: S13, S22 Image description: I03	Creation of linguistic maps on the realization of subject pronouns/clitics (S13), diminutive suffixes (S22) and words for ‘girl’ (I03), presented at MiLES, see <i>WP7.2 Conferences</i> .	Stefan Rabanus	UniVR
07/2024– 11/2024	Sociolinguistic data	Preliminary analysis of sociolinguistic data, specifically regarding declared written usage.	Silvia Dal Negro	UniBZ
07/2024– 11/2024	Translation task: S18 Image description: I01	Creation of linguistic maps on the possessive pronoun ‘their’ (S18) and on words for ‘salami’ (I01).	Stefan Rabanus	UniVR

11/2024–ongoing	Image description: I01-I05, I07	Analysis of all data regarding all 18 dialects and minority languages with reference to the presence or absence of phrasal verbs and creation of preliminary linguistic maps (publications in preparation).	Stefan Rabanus, Ilaria Driussi, Angelica Bonelli, Alessandra Ferretto, Patrizia Cordin	UniVR, UniTN, UniBZ
11/2024–03/2025	Spontaneous speech: M01	Listening and categorization of spontaneous speech in order to extract information on glottonyms, particularly on Tyrolean, Trentino, Ladin.	Angelica Bonelli, Silvia Dal Negro, Alexander Glück	UniBZ
11/2024–03/2025	Translation task	Preliminary analysis of data related to the Walser communities of Issime and Gressoney, specifically focusing on morphosyntactic phenomena associated with verbal prefixation and the selection of lexemes from the Germanic and/or Romance contact lexical layers.	Livio Gaeta	UniTO
11/2024–03/2025	Translation task: S02-S08, S10-S11, S13-S15, S19, S22, S28 Tense transformation: T04	Preliminary analysis of the Francoprovençal data, specifically focusing on the transcription of sentences S02, S03, S04, S05, S06, S07, S08, S10, S11, S13, S14, S15, and S22 in BREL script and IPA. The sentences were annotated for the relevant morphosyntactic variables. Overall, the 'variables' 'wordformation' and 'deixis' were analyzed for all stimuli of the translation tasks. Additionally, sentences S19, S28, and T04 were annotated exclusively for variables related to 'negation'.	Sara Erriu, Gianmario Raimondi	UniVDA
12/2024–01/2025	Sociolinguistic data, Tyrolean varieties	Analysis of the data regarding the use of written dialect, see Alber 2025 in <i>WP7.1 Publications</i> .	Birgit Alber	UNiBZ
01/2025–ongoing	Translation task	Processing of Alpilink data with mapping of the distribution of 1 st , 2 nd and 3 rd person clitics and expletives for Trentino, Veneto, and Friulian.	Ermenegildo Bidese, Simone Barco, Arianna Celeste	UniTN

WP6.3 Evaluation of methodology and project

Period	Data	Analysis	Researcher(s)	Unit
12/2023–05/2024	AlpiLinK Corpus 1.1.1: Lombard, Venetan and Trentino speakers	Analysis of results of fieldwork (mainly in Lombardy) and evaluation of the performance and experience of participants, see <i>WP6.3.1 Analysis of methodology</i> .	Flavia Guerini under the supervision of Alessandra Tomaselli	UniVR
01/2023–08/2024	AlpiLinK Corpus 1.1.2	Analysis of the collected data to evaluate the methodology and to identify potential problems early, see <i>WP6.3.1 Analysis of methodology</i> .	Anne Kruijt	UniVR
10/2023–05/2024	VinKiamo Trentino (1 st edition) reports	VinKiamo participant feedback, see <i>WP6.3.2 Participant feedback</i> .	Anne Kruijt	UniVR
01/2024–05/2024	VinKiamo.4 Friuli reports	VinKiamo participant feedback, see <i>WP6.3.2 Participant feedback</i> .	Ilaria Driussi	UniVR
04/2025–06/2025	VinKo T0303 vs. AlpiLinK I03	Comparative evaluation of the results of translation and image-description tasks on the lexical level (words for ‘girl’), presented at MiLES, see <i>WP7.2 Conferences</i> .	Stefan Rabanus	UniVR
01/2025–05/2025	VinKiamo.5 Friuli reports	VinKiamo participant feedback, see <i>WP6.3.2 Participant feedback</i> .	Ilaria Driussi	UniVR
05/2025–ongoing	VinKo S0116, T0303 vs. AlpiLinK I03	Comparative evaluation of the results of translation and image-description tasks on the grammatical level (subject clitics and phrasal verbs) and on the lexical level (words for ‘girl’), see <i>WP6.3.1 Analysis of methodology</i> .	Stefan Rabanus	UniVR

WP6.3.1 Analysis of methodology

Kruijt analyzed how many participants skipped specific items of the questionnaire. Completely empty questionnaires were not counted, and varieties with less than 10 participants were not taken into consideration, because their results could skew the data significantly. Of the different methodologies, the translation task did slightly better (average upper 80s) than the image description and tense transformation tasks (both in the lower 80s). Participants always start with the translation task followed by the image description and tense transformation tasks. Since January 2024 incomplete questionnaires with some data were added to the corpus. It is therefore difficult to see if the difference in performance is due to the difference in task design or if this difference between tasks could be attributed at least partly to fatigue and the incomplete questionnaires. The lowest performing task is the section for spontaneous speech (M01) with an average of 68 %. This task was optional, at the very end of the questionnaire, and requires people to speak freely about personal experiences.

The translation task (S) performed well, with few items being skipped. The lowest performing items were S02 (77.50 %) and S26 (79.53 %). The rest of the items performed around the upper 80s. There is some variation between different varieties, especially in the less-preferred items (S02 with the lowest score of 59 % for Ladin and 92.31 % for Walser). It must be noted that the stimuli were presented in random order, so fatigue did not play a role in the performance of a stimulus. See the performance of each item for each variety plotted in Figure 7. Guerini noted that difficulties in tasks S and T often stem from the structure of the presented sentence, a known difficulty even in experienced speakers, which in some cases led to a restructuring of the sentence. In some sentences the lexical items presented issues. Items most often presenting issues were the indefinite pronouns (Italian variants *ognuno* in S17, *alcuni* in S20), modal verbs (*dovere* in T01, *potere* in T02) and some expressions or combination of words (*conto suo* in S17, *amici* in S21 and, specifically, the combination *pochi vecchi amici*). The most problematic term was *aiuola* in S02, which could be the reason for the low percentage of production of this sentence. There was also some confusion about whether the sentence was affirmative or

interrogative in S06 *Quello bianco è il tuo cane*, S08 *La casa di Sara è questa* and S27 *Mi siedo qui?* Lastly, Guerini noted some fatigue symptoms (forgetting some words, different verbs used) towards the end of the translation tasks S, the longest task block.

The image-description task items performed roughly the same across the board but with variation between language varieties, e.g. with Occitan performing overall lower (probably due to the relatively high percentage of incomplete questionnaires). In the case of Cimbrian, the underperformance in the image description task was the direct result of a technical malfunction of the questionnaire software in July 2023. See Figure 8. Guerini noted that age influenced the approach to this task. Speakers younger than 60 answered usually in a single short sentence, while speakers over 60 were more creative or more elaborate in their responses. The only image which posed some problems for interpretation was I05 *Dalla finestra viene dentro l'acqua*.

The tense-transformation task is not randomized, and stimuli are presented in the same order. This might be the cause for the slightly lower performance of T06 (78 %) in comparison to the other stimuli (around 81 %). The difference is, however, minimal and might be simply by chance. See Figure 9. Guerini indicated that for the name-truncation task for some speakers it was not clear whether the abbreviations requested were standard Italian or dialect. Regarding task M01, Guerini indicated that especially the participants over 60 years old were very happy to be able to share stories from their childhood and their relationships to the local languages.

Rabanus evaluated the validity of the results of the translation task and the image-description task comparing VinKo data (translation task) and AlpiLinK data (image-description task; AlpiLinK corpus 1.1.5) eliciting the same lexical and grammatical variables (publication in preparation). With respect to the lexicon the words for ‘girl’ in the dialects and minority languages of the regions of Trentino-South Tyrol and Veneto were compared. The most important difference concerns the frequency of the variants *ragazza* and *Mädchen*. These words belong to the standard languages: they do not form part of the lexicon of any of the dialects or minority languages of the area. In the VinKo translation task T0303 approx. 10 % of the Tyrolean speakers used the standard-German *Mädchen* instead of dialect words like *Madl* or *Gitsche*, and, even more, 22 % of the Trentino and Venetan speakers used standard-Italian *ragazza* instead of *butela*, *tosa*, *fiola* or other dialect words. In the AlpiLinK image-description task I03 no standard variants were produced by the Tyrolean speakers, and there was just one single occurrence of standard-Italian *ragazza* in Trentino. Hence, the relatively high frequency of standard-language variants in the translation task must be considered induced by the stimulus sentences in standard Italian and German. This interpretation is further corroborated by the absence of the German or Italian standard-language variants in the minority languages Ladin, Mòcheno and Cimbrian which do not acknowledge German or Italian as their standard languages.

As for grammar, the frequency of the phrasal verb ‘cut down’ with respect to the simple verb ‘cut’, analyzed for the Trentino dialects, also shows priming effects of the standard Italian stimulus sentences (S0116 *Lavano i piatti le ragazze*). Considering that the choice of the phrasal verb is optional, AlpiLinK’s image-description task I03 elicited notably more phrasal verb than VinKo’s translation task S0116: 59 % of the I03 audio recordings feature phrasal verbs but only 39 % of the S0116 audio recordings (the difference is statistically significant). Differently from that, a subject clitic is realized in 76.7 % of the image descriptions (I03) and 80.9 % of the sentence translations (S0116) although almost all target sentences feature a subject NP giving rise to “clitic doubling” (the difference is not statistically significant). This means that not all grammatical features are subject to the priming effects of standard-language structure in translation tasks. Core areas of grammar like subject clitics in Trentino dialects are faithfully produced by the dialect speakers independently of the data-elicitation method. However, in other areas of grammar and in the lexicon the standard-language stimulus influences the translation into the dialect or minority language.

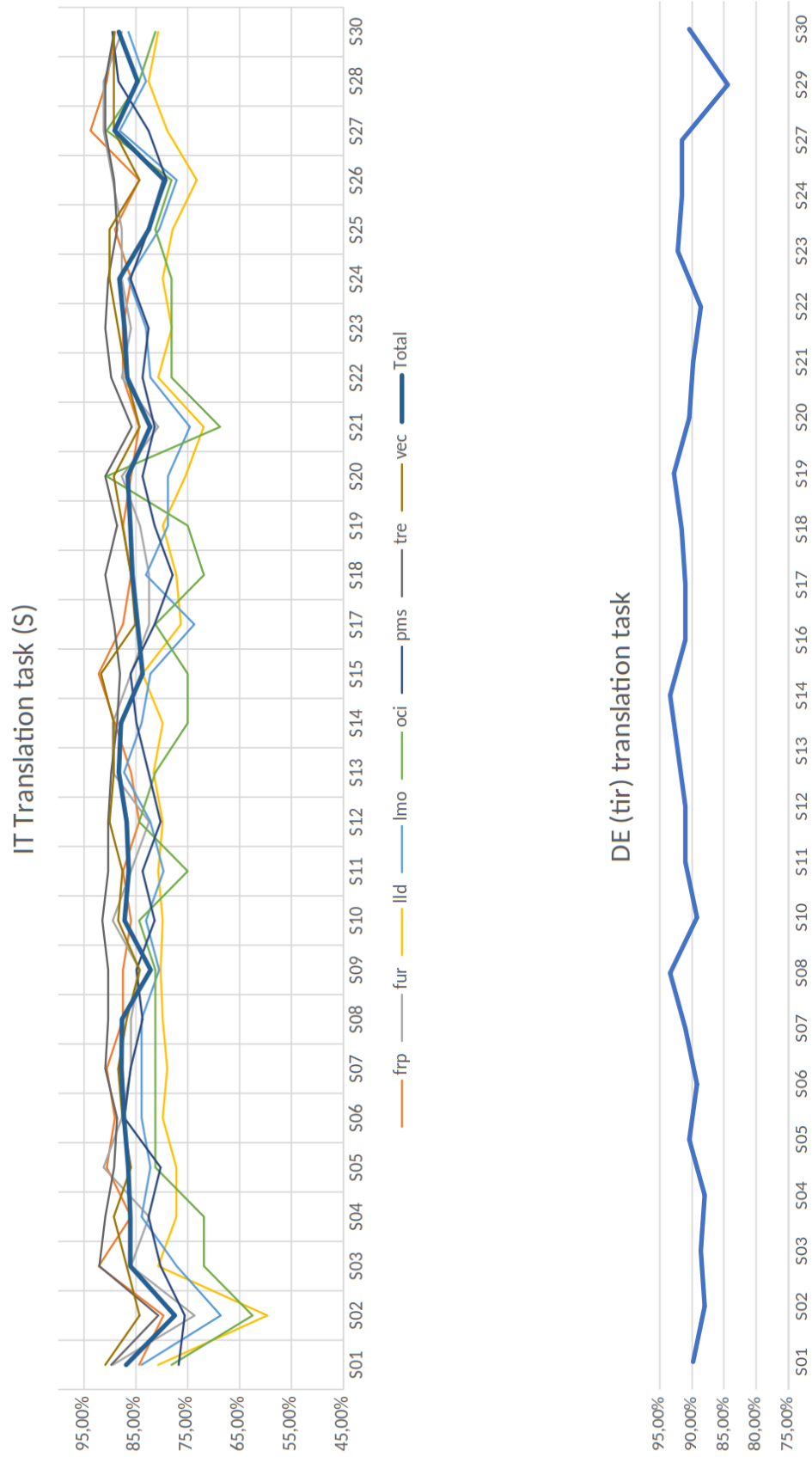


Figure 7: Stimuli performance per variety: on the left in the Italian survey language questionnaire, on the right for the German survey language questionnaire (AlpiLinK Corpus 1.1.2).

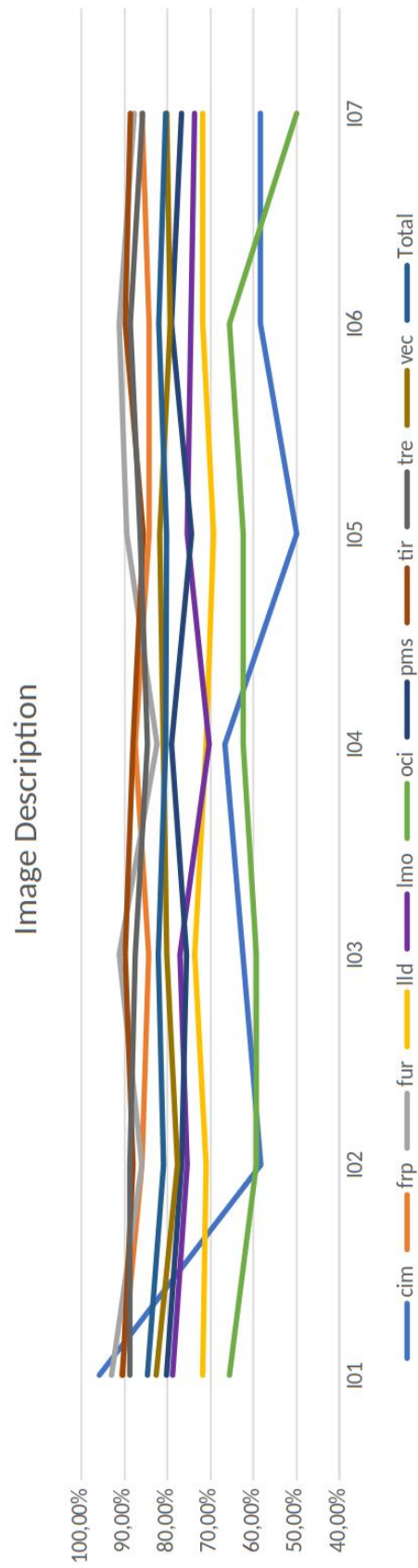


Figure 8: Image description task performance for all varieties (AlpiLink Corpus 1.1.2).

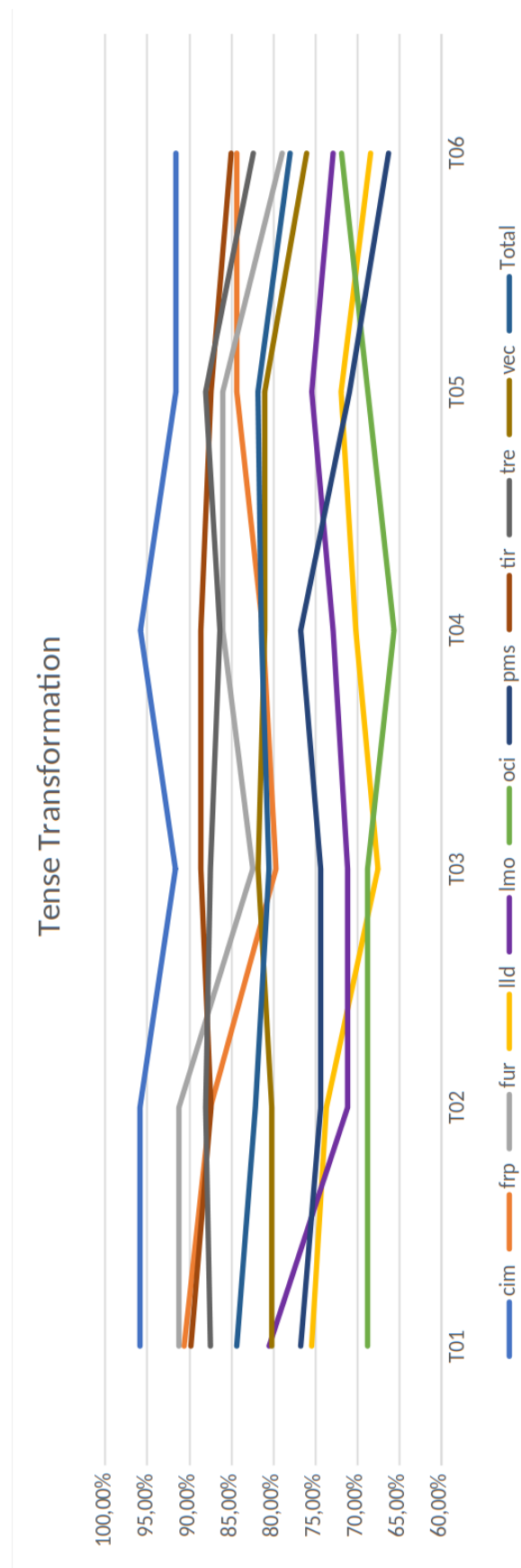


Figure 9: Tense transformation task performance for all varieties (AlpiLinK Corpus 1.1.2).

WP6.3.2 Participant appreciation of the project

This section includes feedback received by the high-school students involved in the VinKiamo projects conducted in Trentino and Friuli.

VinKiamo.5 Friuli reports (reports of 12 students)

The interviews conducted across more than 25 locations involved speakers of at least eight different dialects or minority languages, such as Friulian, Resian, Timavese, Saurano, Sappadino, and both German and Slovenian from Val Canale. Most students reported that finding participants was not difficult, as they relied on family members, friends, and word-of-mouth. Although most participants were cooperative and even enthusiastic, privacy concerns came up in seven reports, particularly regarding the anonymity of the questionnaire. These were always resolved by explaining the project's confidentiality, which reassured the participants.

Students observed a variety of linguistic differences between age groups and regions. Many reported that older speakers used more traditional and complex terms, while younger ones tended to "italianize" vocabulary. In some communities, younger speakers showed stronger influence from Italian or even German, as noted in Timau. Despite these differences, the shared cultural attachment to local language varieties was evident.

Tra i vari parlanti dei vari paesi ci sono state differenze a livello lessicale. (Student no. 12)

Sì, ho notato parecchie differenze. Per quanto riguarda gli anziani, vengono utilizzati termini appunto più datati e complicati mentre i giovani, parlando del friulano, tendono ad "italianizzare" i termini mantenendo con difficoltà la vera lingua. (Student no. 7)

Sì, ho notato delle differenze tra le persone più giovani e quelle più anziane. Ci sono diverse parole che i giovani non usano più al giorno d'oggi, ma che invece, ai loro tempi, erano molto comuni. (Student no. 8)

I parlanti più giovani sono più influenzati dalla lingua tedesca rispetto a quelli più anziani. (Student no. 13)

Ho deciso di intervistare un signore di 85 anni e una ragazza di 25 parlanti della stessa varietà linguistica in modo da poter vedere le differenze linguistiche. Quelle più evidenti sono la scelta delle parole usate e della costruzione della frase. (Student no. 14)

The questionnaire itself was generally well received. Ten students described it as clear and well-structured, and several students appreciated the simplicity of the sentences which made the tasks easier for elderly participants. At the same time, six students felt that the questionnaire was too long, particularly for older informants. Five students also pointed out that some images were hard to interpret, and three mentioned difficulties participants had with exercises involving verb tense transformations.

There were also constructive suggestions for improvement. Six students recommended updating or replacing the images, while five proposed putting the sentences in a suitable context. Some students suggested promoting the project more actively on social media or with flyers in community spaces, so that it could reach a wider audience beyond schools.

Secondo me i punti di debolezza del questionario è che ci sono molte domande riguardanti traduzioni da fare a voce che richiedono un po' di tempo a farlo, invece i punti di forza è che il questionario è ben strutturato e facile da capire. (Student no. 2)

Il questionario è stato per tutti gli intervistati molto semplice e ben strutturato, ma alcuni hanno trovato le immagini non molto chiare e di conseguenza ogni versione era molto differente. (Student no. 5)

Un punto di forza che ho riscontrato è il fatto che si ha diverse opzioni per mantenere viva la lingua, ma un punto di debolezza è la lunghezza del questionario. (Student no. 8)

Penso che uno dei punti deboli sia l'attività di cambio del verbo dal presente al passato. Questa è l'unica attività che ha creato difficoltà ai parlanti, ma penso che abbiano semplicemente fatto fatica a cambiare il tempo verbale. (Student no. 13)

Ritengo che il questionario sia ben strutturato sia per quanto riguarda la traduzione delle frasi, sia per quanto riguarda la trasformazione dei tempi verbali, su cui personalmente avrei lasciato più spazio. I punti di forza

è sicuramente l'ampia scelta di frasi da tradurre che includono molti vocaboli dell'uso quotidiano, mentre di punto di debolezza non ne riscontro. (Student no. 14)

I punti di forza sono sicuramente approfondire le differenze linguistiche anche semplicemente tra le vallate vicine e l'interesse da parte delle persone verso il progetto, mentre attualmente non mi viene in mente alcun punto di debolezza. (Student no. 15)

Many students reflected positively on the experience, describing it as meaningful and personally enriching. Several noted that the project helped them discover aspects of local culture and language they had not previously noticed, while others said it pushed them outside their comfort zones and improved their communication skills. In many cases, the act of conducting interviews and listening to stories made students more aware of the fragility of these minority languages and the importance of preserving them for future generations.

Punti di forza: è un bellissimo progetto e aiuta la salvaguardia delle lingue minoritarie Punti deboli: è alquanto lungo e certe volte difficile da svolgere in una sola sessione. (Student no. 3)

Nella zona in cui abito prevale l'Italiano soprattutto tra i giovani, mentre le persone adulte o/e anziane parlano anche lingue minoritarie come il tedesco e lo sloveno della valcanale o il friulano. La presenza di queste minoranze linguistiche mette in primo piano la storia del mio territorio, essendo frammentato da diverse culture e tradizioni. Questo multiculturalismo è appunto segnato anche dai semplici menu di un ristorante, che da me sono in tre lingue (italiano, tedesco e sloveno). (Student no. 3)

Sicuramente un punto di forza è il fatto che tramite questo questionario puoi scoprire nuove culture e parole diverse dalla tua variante di dialetto. (Student no. 7)

Secondo me è un buon modo per mantenere vivi i dialetti, in modo che le generazioni future li parlino. Per quanto riguarda il punto di debolezza è il numero di frasi. (Student no. 6)

Sono riuscita a svolgere interviste per tutte le lingue richieste (Friulano, Resiano, Tedesco della Val Canale, Sloveno della Val Canale, Sappadino, Saurano, Timavese) riscontrando tra le stesse qualche similitudine sia in termini di vocaboli sia nella difficoltà di traduzione. Essendo lingue antiche molti termini sono stati italianizzati in quanto non esistenti nella lingua stessa (aiuola, macchina, frutta). In tutti i posti in cui sono stata ho trovato nel paesaggio la lingua in forma scritta per esempio nei nomi delle vie o nei cartelli del paese. Questo unito a dizionari, volantini, calendari, libri scolastici rappresentano il tentativo di tramandare una lingua che essendo fondamentalmente orale rischia di andare persa. (Student no. 9)

VinKiamo.4 Friuli reports (reports of 18 students)

The students overall had no issues with finding participants, as many were happy to participate and enthusiastic about the project goals. Some informants raised concern about privacy issues (mentioned in seven reports), but these concerns were always easily resolved by the students explaining that participation was anonymous. Four students recommend further promotion of the project via more social media.

I miei intervistati si sono sentiti a loro agio e felici di dimostrare la loro cultura soprattutto quelli sui 70 anni. Il giovane che ho intervistato è stato fiero di parlare il suo dialetto. (Student no. 5)

All reports indicate that both the questionnaire and the website were easy to navigate and of interest to participants and themselves, especially the "Listen & Explore" section. Despite the clarity of the questionnaire, all students wrote that their presence was helpful for the successful completion of the questionnaire, either by providing additional explanations or encouraging participants to complete all tasks. Five students wrote that the questionnaire was a bit too long, while another six found the length adequate. Regarding specific tasks, it has been noted that the translation-task section was too long and that name-truncation task was disliked as not related to local varieties or confusing.

No technical difficulties have been noted other than a single person saying that the processing of the audio recordings per item took some time, and this increased significantly the overall questionnaire time. The most heard feedback was the inclusion of more subvarieties, e.g. *carnico, paularino, gjierean* (Students no. 1, 9, 16) on the website, and to include more information regarding local proverbs, cultural practices, words for flora or fauna etc.

A parte la piccola notazione che ho scritto precedentemente per quanto riguarda l'aggiunta di qualche sottogruppo di varietà linguistica in più (come il carnico) trovo che il sito sia completo; alla portata di tutti; incuriosisce le persone che non si sono mai avvicinate ad argomenti come le lingue minoritarie sia da un punto di vista contenutistico sia per quanto riguarda l'estetica del sito. (Student no. 16)

Nel sito aggiungerei un piccolo questionario di vocaboli riguardanti flora e fauna, oppure termini che concernono la gastronomia, che la nostra lingua ha completamente differenti rispetto alla figlia del latino. Anche molti verbi sono diversi relativamente all'italiano. (Student no. 6)

Students noted both phonological and lexical differences between the languages.

I miei intervistati vanno dai 15 ai 70 anni. Ho potuto comunque riflettere che nonostante il dialetto delle cinque persone da me intervistate sia il friulano, cambia molto in base al paese anche se sono tre km di distanza. Poi se il comune cambia, la cadenza e certe parole proprio cambiano totalmente. (Student no. 5)

Ho riscontrato alcune differenze tra le località, poiché ogni vallata ha la propria parlata (ad esempio nel comune di Sutrio le parole tendono a finire con la lettera "e", mentre a Enemonzo con la lettera "a"). (Student no. 9)

VinKiamo Trentino 1st edition

From the reports of two students emerge that the project was considered interesting and informative with respect to the linguistic variation of the area: the students appreciated the project.

Abbiamo acquisito la consapevolezza che non è necessario spostarsi all'estero o allontanarsi dal proprio luogo di residenza per immergersi nell'apprendimento di una lingua o di un dialetto. Nel contesto del Trentino, in particolare, abbiamo riconosciuto la straordinaria varietà di dialetti e minoranze linguistiche presenti sul territorio, ciascuno dei quali rappresenta una forma autentica e ricca di espressione culturale. (Students from Liceo Linguistico Sophie M. Scholl, Trento)

Ci ha molto colpito l'aspetto umano della ricerca. I parlanti, in particolare quelli più anziani, ci hanno sorpreso con un entusiasmo che non ci aspettavamo. Per loro, era un onore e un piacere poter far conoscere ad un pubblico più ampio la loro lingua o il loro dialetto e poter quindi allo stesso tempo portare avanti un'eredità secolare che sembra star scomparendo con il passare degli anni. (Students from Liceo Linguistico Sophie M. Scholl, Trento)

In the reflection on the linguistic variation, they often underline the influence of age on the use and fluency of dialect (a major factor for Trentino varieties, less present in Ladin communities). They also discuss the influence of standard Italian on dialect, with a stronger influence on younger speakers.

Invece, abbiamo notato una grande differenza generazionale nella modalità di parlare il dialetto, sia nell'ambito grammaticale che nell'ambito del vocabolario. Sotto questo punto di vista, più il parlante è giovane, maggiore sarà l'influenza dell'italiano nel dialetto. Per esempio, un anziano direbbe "magnava" per dire "mangiavo", mentre un parlante più giovane direbbe "magnavo"; oppure, mentre le generazioni più giovani utilizzano parole in italiano dialettizzate, per esempio "lucchetto" che diventa "lucheto", un anziano direbbe "merlos"; ad Oltrecastello, e per estensione a Trento, è trasparso in modo lampante come nel capoluogo e nei dintorni il dialetto si vada perdendo, venendo sostituito dall'italiano. I bambini e i giovani della città spesso non conoscono e non parlano il dialetto, mentre a Moena e Lavarone è prassi che anche i giovani parlino quasi più dialetto (o lingua ladina nel caso della Val di Fassa) che italiano, sia con i più anziani che tra di loro. (Students from Liceo Linguistico Sophie M. Scholl, Trento)

The students encountered very few problems with the questionnaire. The issues mentioned in the reports were the following. For some older speakers the task explanations being exclusively in standard Italian was a bit problematic and they had some issues understanding the tasks but still managed to successfully participate with the help of the students. Another issue was that some participants were 'shy' or afraid to make mistakes and some therefore were hesitant to supply a lot of information, especially in the open speech task (M01).

WP7. Documentation and publication of results

This work package consists of the presentation of results in workshops and conferences, and in beginning to work on publications. Relevant international conferences include “Methods in Dialectology”, “International Conference on Language Variation in Europe” and the yearly conference of the “Societas Linguistica Europaea”; additionally, in Italy, the conferences of the “Società di Linguistica Italiana” (SLI), in Germany the conferences of the “Internationale Gesellschaft für Dialektologie des Deutschen” (IGDD).

The following sections present lists of publications (WP7.1), conferences (WP7.2) and advanced academic teaching modules at MA and PhD level (WP7.3) based on AlpiLinK data or other aspects the AlpiLinK project and on data from preceding projects in the AlpiLinK area which team members participated at (e.g., VinKo, CliMAIp, APV). The table below gives an overview of the numbers (July 2025).

Typology	Period	Items
Publications	06/2022–06/2025	40
Presentations at scientific conferences	06/2022–06/2025	90
Academic teaching (complete seminars, cycles of classes, single lectures)	06/2022–06/2025	19

WP7.1 Publications

The list features only published papers and books. Considering the long processes in academic publishing and the fact that the AlpiLinK data collection started only in July 2023, at the time of writing (July 2025) many papers on AlpiLinK are “in press” or “in preparation”, hence, they are not mentioned here.

- Alber, Birgit. 2025. Sprachkontakt und sprachliche Vitalität im Südbairischen. Tiroler Dialekte, Mòcheno, Zimbrisch und Hutterisch. In: Butcher, John, Marta Penchini & Josef Prackwieser (eds.): Historische Mehrsprachigkeit im südlichen Tirol vom Frühmittelalter bis zum Ersten Weltkrieg. Milan/Udine: Mimenis, 33–51.
- Alber, Birgit, Sabine Arndt-Lappe & Joachim Kokkermans. 2025. The Predictability of Name Truncation: Factoring in Language Change. *Catalan Journal of Linguistics* 24 (1), 7–39.
<https://doi.org/10.5565/rev/catjl.467>.
- Benedetto Mas, Paolo & Gianmario Raimondi (eds). 2023. L’Atlas des Patois Valdôtains: sguardi incrociati/regards croisés, Alessandria: Edizioni dell’Orso.
- Benedetto Mas, Paolo. 2023. Rappresentazioni geolinguistiche della Valle d’Aosta. In: Benedetto Mas & Raimondi (eds), 35–50.
- Bertollo, Sabrina & Stefan Rabanus. 2023. VinKiamo: ein Citizen-Science-Projekt für Schulen zur Förderung von (sprach-)übergreifenden Kompetenzen. In *Alsic* 26 (1): 1-19.
<https://doi.org/10.4000/alsic.7076>.
- Bidese, Ermenegildo. 2023. Sprachkontakt generativ: Eine Untersuchung kontaktbedingten syntaktischen Wandels im Zimbrischen. Berlin: De Gruyter. <https://doi.org/10.1515/9783110765014>.
- Bidese, Ermenegildo. 2024. I Cimbri. In: Destro Bisol, Giovanni, Erica Autelli, Marco Capocasa & Marco Caria (eds.): Gli Italiani che non conosciamo. Lingue, DNA e percorsi delle comunità storiche minoritarie. Alghero: Edicions de l’Alguer, 67–77.
- Cordin, Patrizia, Vittorio Dell’Aquila, Fernando Ramallo & Sabrina Rasom. 2023. Guida per l’educazione al plurilinguismo con lingue locali. Trento: Erickson.
- Cordin, Patrizia. 2022. Iniziative dell’Ateneo trentino per la valorizzazione delle lingue cimbra e mòchena. In Fusco, Fabiana (ed.): Atti della prima Conferenza regionale sulla tutela delle minoranze di lingua tedesca del Friuli Venezia Giulia/Akten der ersten Regionalkonferenz über den Schutz der deutschsprachigen Minderheiten Friaul Julisch Venetiens. Udine: Forum, 57–70.

- Cordin, Patrizia. 2023. Analisi SWOT per il ladino in Val Badia e in Val Gardena. *Mondo Ladino* 47, 126–137.
- Cordin, Patrizia. 2023. La vitalità delle lingue di minoranza nel Trentino. Esiti della ricerca ‘CLaM (Cimbro Ladino Mòcheno) 2021’. *Studi Trentino: Storia* 102 (1), 211–228.
- Cordin, Patrizia. 2024. I verbi sintagmatici nei dizionari dialettali trentini. In *Linguistik Online* 125 (1), 77–96. <https://doi.org/10.13092/lo.125.10786>.
- Cosentino, Michele. 2023. Tre giorni di raccolta dati a Luserna: qualche considerazione (socio)linguistica. CLaM 2021 (Cimbro Ladino Mòcheno 2021). <https://cimbroladino-mocheno-2021.lett.unitn.it/raccolta-dati-luserna>.
- Favre, Saverio, Marina Fey & Gianmario Raimondi. 2023. L’Atlas des Patois Valdôtains, dalla ricerca sul campo alla restituzione. In: Benedetto Mas & Raimondi (eds.), 9–32.
- Gaeta, Livio. 2023. Phonology and conventionalization: Naturalness and beyond in German(ic) sonorants. In: Brunner, Jennifer, Petra Hödl, Veronika Mattes & Thomas Schwaiger (eds.): *Für Bernhard Hurch. Special issue of Grazer Linguistische Studien* 94, 121–134.
- Gaeta, Livio. 2024. Evolutionary steps for linguistic signs: The place of indexicality. In: Nielsen, Peter Juul & María Sol Sansiñena Pascual (eds.): *Indexicality. The role of indexing in language structure and language change*. Berlin: De Gruyter, 99–126. <https://doi.org/10.1515/9783110791433-003>.
- Gaeta, Livio. 2024. Iconicity Principles from an Evolutionary Perspective. In: Evans, Bethwyn, Maria Kristina Gallego & Luisa Miceli (eds.): *Historical Linguistics 2019. Selected Papers from the 24th International Conference on Historical Linguistics, Canberra 1.–5.7.2019*. Amsterdam/Philadelphia: Benjamins, 194–215.
- Gaeta, Livio. 2024. Intense Language Contact and Collapse of Lexical Strata: Verbs Ending with *-urun* in Issime. *Journal of Language Contact* 17, 642–663.
- Gaeta, Livio. 2024. The Observer’s Paradox Meets Corpus Linguistics: Written and Oral Sources for the Walser Linguistic Islands in Italy. In: Genesin, Monica, Gerhard Hempel & Thede Kahl (eds.): *Endangered Language Varieties in Italy and the Balkans/Bedrohte Sprachvarietäten in Italien und auf dem Balkan/Varietà linguistiche a rischio in Italia e nei Balcani*. Vienna: Austrian Academy of Sciences, 361–374.
- Gaeta, Livio. 2025. Word-formation in Linguistic Islands – Abstracts in Walser German. In: Matrisciano-Mayerhofer, Sara, Johannes Schnitzer & Elisabeth Peters (eds.): *Patterns, Variants, and Change: Through the Prism of Morphology. Studies in Honor of Franz Rainer*. Strasbourg: Éditions de Linguistique et de Philologie, 315–328.
- Gaeta, Livio, Marco Angster, Raffaele Cioffi & Marco Bellante. 2022. Corpus linguistics for low-density varieties. Minority languages and corpus-based investigations. In: Grossmann, Maria, Fabio Montermini, Diana Passino, Diego Pescarini & Florence Villoing (eds.): *Corpus et données en morphologie*. Special Issue of *Corpus* 23, 1–25. <https://journals.openedition.org/corpus/7345>.
- Gaeta, Livio, Saracco, Caterina & Cioffi, Raffaele. 2024. *Greschòneytisch. Nuovo Dizionario Titsch – Italiano / Tedesco*. Alessandria: Edizioni dell’Orso.
- Kruijt, Anne, Patrizia Cordin & Stefan Rabanus. 2023. On the validity of crowdsourced data. In: Pustka, Elissa, Carmen Quijada Van den Berghe & Verena Weiland (eds.): *Corpus Dialectology*. Amsterdam/Philadelphia: Benjamins, 10–33.
- Kruijt, Anne, Stefan Rabanus & Marta Tagliani. 2023. The VinKo-Corpus. Oral data from Romance and Germanic local varieties of Northern Italy. In: Kupietz, Marc & Thomas Schmidt (eds.): *Neue Entwicklungen in der Korpuslandschaft der Germanistik. Beiträge zur IDS-Methodenmesse 2022*. Tübingen: Narr, 203–212.
- Kruijt, Anne & Stefan Rabanus. 2025. From VinKo to AlpiLinK: Web-based Long-term Storage and Accessibility of Information. In: Krefeld, Thomas & Stephan Lücke (eds.): *Berichte aus der digitalen Geolinguistik (IV)*. Munich: <http://www.kit.gwi.uni-muenchen.de/?p=106395&v=1>.

- Pons, Aline. 2023. Sull'unità delle Valli Valdesi dal punto di vista linguistico. In: Bauer, Roland & Thomas Krefeld (eds.): *Lo spazio comunicativo dell'Italia e delle varietà italiane*, Munich.
<https://www.kit.gwi.uni-muenchen.de/?p=109884&v=1>.
- Pons, Aline. 2024. Continuità (e distanza) tra la varietà occitana della val Germanasca e le varietà limitrofe. In Fratini, M. (eds.), *Vie di comunicazione e viabilità nelle valli Chisone, Germanasca e Pellice*, Perosa Argentina, LAR Editore: 227–236.
- Rabanus, Stefan. 2023. Nome di battesimo e articolo espletivo – crowdsourcing e cartografica linguistica nello studio della variazione linguistica in Trentino-Alto Adige e Veneto. In: Schöntag, Roger & Laura Linzmeier (eds.): *Neue Ansätze und Perspektiven zur sprachlichen Raumkonzeption und Geolinguistik: Fallstudien aus der Romania und der Germania*. Lausanne: Peter Lang, 93–134.
- Rabanus, Stefan. 2024. Two modes of contact-induced change in minority languages: Phonology and syntax vs. inflectional morphology. In: Hans-Bianchi, Barbara, Barbara Vogt & Chiara Truppi (eds.): *Speakers and Structures in Language Contact. Pluralistic Approaches to Change and Variation*. Berlin/Boston: De Gruyter: 65–92. <https://doi.org/10.1515/9783111188348-003>.
- Raimondi, Gianmario, Marco Angster, Marco Bellante, Paolo Benedetto Mas, Raffaele Cioffi, Livio Gaeta, Aline Pons & Matteo Rivoira. 2023. Going standard on a blank page: a corpus-based approach to the written varieties of the Western Alpine minorities (Occitan, Francoprovençal and Walser). In: Pustka, Elissa, Carmen Quijada Van den Berghe & Verena Weiland (eds.): *Corpus Dialectology*. Amsterdam/Philadelphia: Benjamins, 84–105.
- Rivoira, Matteo, Mattia Ravera, Nicola Duberti. 2023. Possessivi nell'occitano delle valli della provincia di Cuneo. In: Giraudo, Andrea, Walter Meliga, Giuseppe Noto, Aline Pons & Matteo Rivoira (eds.): *Centri e periferie. Atti del XIII congresso dell'AIEO*. Alessandria: Edizioni dell'Orso, 609–628.
- Rivoira, Matteo. 2024. L'espressione del plurale nella scripta dei manoscritti valdesi e nei dialetti occitani alpini. In: Baricci, Erica, Walter Meliga & Luca Sacchi (eds.): *Meminisse iuvabit Scritti in memoria di Federico Emidio Bo*. Milan: Ledizioni, 219–241.
- Rivoira, Matteo. 2024. L'occitano. *Linguistik Online* 130 (6), 281–307.
<https://doi.org/10.13092/lo.129.11159>.
- Rivoira, Matteo. 2024. Storia linguistica dei valdesi alpini. In: Peyronel Rambaldi, Susanna (ed.): *La storia dei valdesi. L'età moderna*. Turin: Claudiana, 715–735.
- Saracco, Caterina. 2024. Le minoranze walser dell'Italia settentrionale. Gressoney (AO) e la lingua titsch. In: Destro Bisol, Giovanni, Erica Autelli, Marco Capocasa & Marco Caria (eds.): *Gli Italiani che non conosciamo. Lingue, DNA e percorsi delle comunità storiche minoritarie*. Alghero: Edicions de l'Alguer, 1–9.
- Saracco, Caterina. 2024. Le varietà walser dell'Italia settentrionale: Gressoney e il titsch. *Linguistik Online* 130 (6), 235–255. <https://doi.org/10.13092/lo.129.11157>.
- Saracco, Caterina, Raffaele Cioffi, Dario Capelli & Livio Gaeta. 2024. Plurilinguismo e identità linguistica nelle isole walser. In: Verdiani, Silvia, Silvia Ulrich & Cristina Onesti (eds.): *L'intercomprensione tra tedesco e lingue altre*. Pisa: ETS.
- Tomaselli, Alessandra & Ermenegildo Bidese. 2023. Fortune and Decay of Lexical Expletives in Germanic and Romance along the Adige River. *Languages* 8 (1): 44.
<https://doi.org/10.3390/languages8010044>.
- Tomaselli, Alessandra, Ermenegildo Bidese & Andrea Padovan. 2022. Feature Borrowing in Language Contact. *Languages* 7 (4), 1–10. <https://doi.org/10.3390/languages7040288>.
- Tomaselli, Alessandra. 2023. *Dal particolare all'universale. Viaggio nella grammatica del cimbro*. Verona: QuiEdit.

WP7.2 Conferences

- Alber, Birgit & Joachim Kokkelmans. 6 July 2022. “Germanisch-romanischer Sprachkontakt im Sibilantensystem. Tiroler und Trentiner Dialekte.” 7th conference of the International Association of Dialectology of German (IGDD), University of Salzburg.
- Alber, Birgit & Joachim Kokkelmans. 6 October 2022. “South Bavarian rhotics in crowdsourced linguistic data from Northeastern Italy: a diachronic and qualitative comparison.” Conference “Beyond Borders: German-speaking Minorities in Italy and around the World”, University of Trento.
- Alber, Birgit. 26 May 2023. “Language Documentation via Crowdsourcing.” Invited talk at ZAS, Berlin.
- Alber, Birgit. 25 November 2023. “Sprachkontakt und sprachliche Vitalität im Südbairischen: Mòcheno, Zimbrisch und Hutterisch.” Presentation at “Mehrsprachigkeit in Südtirol vom Frühmittelalter bis zum Ersten Weltkrieg”, Meran.
- Alber, Birgit. 7 December 2023. “Crowd-sourcing and language documentation.” Presentation at UiT (Norges Arktiske Universitet), Tromsø.
- Alber, Birgit. 14 December 2023. “Crowdsourcing – Nuove possibilità per lo studio delle lingue locali.” Conference “La tecnologia incontra le lingue locali e minoritarie”, University of Parma.
- Alber, Birgit. 15 December 2023. “Sistemi fonologici: Cimbri, Mòcheno, e Tirolese.” Presentation at “Evoluzione e Diversità del Linguaggio Umano dalla Preistoria alla Modernità”, University of Trento.
- Alber, Birgit, Joachim Kokkelmans & Sabine Arndt-Lappe. 25 January 2024. “The predictability of name truncation – factoring in time and space.” Workshop “Nominal Inflection and Word-formation at the Phonology-Morphology Interface”, University of Barcelona.
- Alber, Birgit. 6 June 2024. “Tyrolean syllables under a typological perspective.” Workshop “Phonetics meets dialectology: Vowel dynamics in non-standard varieties”, University of Bozen-Bolzano.
- Alber, Birgit. 20 June 2024. “Crowdsourcing in language documentation – opportunities and challenges.” Presentation ‘Schultink lecture’ for the LOT Summer School in Leiden, Netherlands.
- Alber, Birgit. 26 October 2024. “Language contact at the phonological level.” Workshop “The Limits of Convergence in Language Contact Situations”, University of Verona.
- Alber, Birgit. 12 February 2025. “AlpiLinK and VinKiamo. Crowdsourcing and Citizen Science in Linguistics.” Invited talk, University of Potsdam.
- Benedetto Mas, Paolo & Gianmario Raimondi. 14 September 2023. “Variabilità diatopica, costruzioni culturali, distanza linguistica: per una ridefinizione del continuum dialettale della Valle d’Aosta.” 56th International Congress of the Italian Linguistic Society (SLI), University of Turin.
- Benedetto Mas, Paolo. 23 September 2023. “L’area francoprovenzale in Piemonte: note (socio)linguistiche.” “Giornata delle lingue minoritarie”, Pomaretto (Torino).
- Benedetto Mas, Paolo. 14 October 2023. “Gérer la complexité linguistique du francoprovençal: le corpus CLiMAlp.” Annual conference “Centre d’Etudes francoprovençales René Willien”, Saint-Nicolas (Aosta).
- Benedetto Mas, Paolo. 21 February 2024. “Minoranze linguistiche a scuola: un’inchiesta tra i futuri insegnanti della Valle d’Aosta.” 24th International Conference AItLa, University of Pavia.
- Benedetto Mas, Paolo. 13 June 2024. “Les matériaux inédits de la Carta dei Dialecti Italiani et des Tableaux phonétiques des parlers valdôtains: présentation et utilisations possible.” Conference “Le francoprovençal: des origines à aujourd’hui”, Lyon Catholic University.
- Bertollo, Sabrina & Stefan Rabanus. 24 November 2022. “VinKiamo: ein Citizen-Science-Projekt für Schulen zur Förderung von (sprach)übergreifenden Kompetenzen.” Conference “Digital Citizenship, Digital Wilds und Lehren & Lernen von Sprachen”, University of Salzburg.

- Bertollo, Sabrina & Stefan Rabanus. 2 September 2023. “Citizen science and the classroom: rewilding language learning.” Conference “Language Learning and Teaching in Digital Transformation”, Pädagogische Hochschule Luzern.
- Bertollo, Sabrina & Stefan Rabanus. 3 October 2024. “Citizen Science Initiatives for the Documentation and Promotion of Minority Languages in Northern Italy.” Annual Symposium of the Scientific Network RéAL2 “Acquisition, Teaching, and Promotion of ‘Non-Majoritarian’ Varieties”, University of Verona.
- Bertollo, Sabrina & Romano Madaro. 19 February 2025. “Resilience is Relative but How Resilient are Relatives? Some Insights from Germanic-Romance Contact.” 50th Generative Grammar Meeting, Pre-conference workshop “Resilient Syntax in Contact: Assessing Minority Languages”, University of Padua.
- Bertollo, Sabrina, Romano Madaro & Alessandra Tomaselli. 23 June 2025. “Relative clauses in the Germanic varieties in the Alps: some cues from the diachrony of Cimbrian and Timavese under the pressure of (Italo)Romance.” 26th Diachronic Generative Syntax conference, University of Oxford.
- Bidese, Ermenegildo & Michele Cosentino. 20 April 2023. “Il contatto linguistico germanico-romanzo lungo l’arco alpino italiano: parlarne attraverso il progetto AlpiLinK.” 7th conference at the “Centro Internazionale di Dialettologia”, Potenza.
- Bidese, Ermenegildo, Patrizia Cordin, Patrizia & Michele Cosentino. 5 October 2023. “Limits and advantages of modern technologies for documenting the Italian Alps linguistic varieties: from VinKo to Alpilink.” Conference “Documenting languages, Documenting cultures”, Naples.
- Bidese, Ermenegildo & Romano Madaro. 1 December 2023. “Routes of syntactic variation. The role of geographical distance in shaping variation among the German(ic) varieties in Northeastern Italy.” Workshop “Il ruolo dell’adaptation nelle lingue minoritarie in Italia”, University of Venice.
- Bidese, Ermenegildo & Michele Cosentino. 21 February 2024. “Un nuovo strumento per la documentazione, la valorizzazione e la promozione delle varietà dell’arco alpino italiano.” 24th International Conference AItLa, University of Pavia.
- Bidese, Ermenegildo, Romano Madaro & Alessandra Tomaselli. 25 October 2024. “The Historic Germanic Language Islands in the North-East of Italy and their ‘Adherence’ to the German Model.” Workshop “The Limits of Convergence in Language Contact Situations”, University of Verona.
- Bissolo, Serena. 28 November 2024. “How Do I Say (Not) in Dialect? – Image Description Task from AlpiLinK.” Workshop in occasion of the launch of the TreLinLab Linguistics Laboratory, University of Trento.
- Bissolo, Serena. 21 February 2025. “The AlpiLinK Project.” International Mother Language Day, University of Trento.
- Capelli, Dario. 10 September 2024. “Das Töitschu in Issime: Eine Untersuchung.” 21st “Alemann:innentagung”, University of Bern.
- Cioffi, Raffaele. 30 November 2023. “Fra traduzione e codifica di una lingua: la Parabola del Figliol Prodigo nelle varietà Walser italiane. Parte prima: fonti e documentazione.” Workshop “Il ruolo dell’adaptation nelle lingue minoritarie in Italia”, University of Venice.
- Cordin, Patrizia. 23 February 2024. “Educazione al plurilinguismo con lingue di minoranza: esperienze a confronto.” 24th International Conference AItLa, University of Pavia.
- Cordin, Patrizia. 21 February 2025. “What is the Mother Tongue”. International Mother Language Day, University of Trento.
- Cordin, Patrizia. 28 February 2025. “Minority Languages Beyond the Margins: A Challenge for Teaching.” Workshop “MinEdu – Supporting Minority Languages in Educational Contexts”, Free University of Bozen-Bolzano.

- Cosentino, Michele, Romano Madaro & Ermenegildo Bidese. 11 September 2023. “Contatto germanico-romanzo e variazione morfosintattica nel triveneto: due casi di studio.” 17th Cambridge Italian Dialect Syntax-Morphology Meeting, University of Zurich.
- Dal Negro, Silvia. 4 March 2024. “Lingue piccol(issim)e tra oralità e scrittura.” Invited talk at the University of Pavia.
- Dal Negro, Silvia. 3 July 2024. “Writing in a minority language/in a dialect: language policy and actual use.” Keynote talk at the conference “Minority Languages in European Societies” (MiLES), University of Turin.
- Dal Negro, Silvia. 28 September 2024. “Variazione linguistica e sociolinguistica sulle Alpi italiane: esplorando i dati di AlpiLinK.” Summer Seminar of the “Istituto di dialettologia e di etnografia valtellinese e valchiavennasca”, Ponte in Valtellina.
- Dal Negro, Silvia. 20 March 2025. “Die Wahrnehmung von Mehrsprachigkeit. Was Sprecher in deutschsprachigen Sprachinseln Norditaliens über ihre Mehrsprachigkeit sagen.” Opening lecture at “Diversität – Dynamik – Hybridität”, University of Heidelberg.
- Dal Negro, Silvia. 10 May 2025. “Il plurilinguismo endogeno nella prospettiva di una sociolinguistica educative.” Keynote talk at “A 50 anni dalle Dieci tesi per l’educazione linguistica democratica”, University of Bologna.
- Gaeta, Livio. 23 June 2022. “Between derivation and multifunctionality: in search of evidence for conversion.” Workshop on “Theoretical and empirical descriptions of conversion/zero-affixation” held during the Conference “Word-Formation Theories VI & Typology and Universals in Word-Formation V”, Pavol Jozef Šafárik University, Košice.
- Gaeta, Livio, Raffaele Cioffi & Caterina Saracco. 16 September 2022. “The passive complex in an isolated Walser German variety: the role of the corpus.” 16th Cambridge Italian Dialect Syntax-Morphology Meeting, Università Federico II, Naples.
- Gaeta, Livio. 3 August 2022. “Intense language contact and collapse of lexical strata: verbs ending with -*urun* in Issime.” 25th International Conference on Historical Linguistics, University of Oxford.
- Gaeta, Livio. 7 October 2022. “Germanische Sprachinseln in Italien: Ein (noch zu erforschender) Kulturschatz von hervorragender Bedeutung für Sprachvariation, -kontakt und -komplexität.” Conference “Beyond Borders. German-speaking Minorities in Italy and around the World”, University of Trento.
- Gaeta, Livio, Raffaele Cioffi, Caterina Saracco & Dario Capelli. 29 March 2023. “Plurilinguismo e identità linguistica nelle isole walser.” International academic seminar “Traduzione, multilinguismo e intercomprensione”, University of Turin.
- Gaeta, Livio. 30 August 2023. “Similitive constructions in the Walser German linguistic islands of Northern Italy.” Workshop “Similarity of quality and denominal similatives”, 56th International Annual Meeting of the Societas Linguistica Europaea, University of Athens.
- Gaeta, Livio & Robert Mailhammer. 10 October 2023. “A ditropic clitic in the Walser German variety Titsch (Gressoney, Northern Italy): the case of *ze* ‘to’.” Conference “German Abroad 5”, Catholic University of Eichstätt-Ingolstadt/Ludwig-Maximilians-Universität, Munich.
- Gaeta, Livio & Caterina Saracco. 27 October 2023. “Kontrastive Linguistik und Sprachdidaktik: Die Modalität und die Modalverben im Deutschen, Englischen und Italienischen.” Conference “DaF heute zwischen Anwendung und Empirie. Tendenzen und Perspektiven”, University of Trento.
- Gaeta, Livio. 30 November 2023. “Fra traduzione e codifica di una lingua: la Parabola del Figliol Prodigo nelle varietà Walser italiane. Parte seconda: Diacronia e evoluzione del repertorio linguistico.” Workshop “Il ruolo dell’adaptation nelle lingue minoritarie in Italia”, University of Venice.
- Gaeta, Livio & Raffaele Cioffi. 21 February 2024. “Banche dati per documentare la diversità linguistica delle Alpi.” Presentation at the International Mother Language Day, University of Turin.

- Gaeta, Livio. 8 August 2024. "Towards a Darwinian Model of Language Change." 21st International Congress of Linguists (ICL), University of Poznań.
- Gaeta, Livio. 23 August 2024. "Alz chént òn geit: COME and GO as Passive Auxiliaries in Titsch." 57th International Annual Meeting of the Societas Linguistica Europaea, University of Helsinki.
- Gaeta, Livio. 30 August 2024. "Swinging Between Simplification and Complexification: Morphologization as a Repair Strategy." 21st International Morphology Meeting, Vienna University of Economics and Business.
- Gaeta, Livio. 20 September 2024. "Word Formation in Minority Languages: Corpus Linguistics for Low-Density Varieties." Workshop on Data-based Research in Word Formation within "The Biennial of Czech Linguistics", University of Prague.
- Gaeta, Livio. 7 November 2024. "Between Productivity and Creativity: Snowclones We Live By." Department of Linguistics and Computational Linguistics, University of Saarland, Saarbrücken.
- Gaeta, Livio. 8 November 2024: "Innovation, Convergence, and Areality in Walser Linguistic Islands." Centre for European Regional and Minority Languages, University of Saarland, Saarbrücken.
- Gaeta, Livio. 2 December 2024. "What Happened to Töitschu? Language Contact and Change in Alpine Linguistic Islands." 4th AMC Symposium Contact and Language, University of Edinburgh.
- Glück, Alexander. 06/06/2024. "The vowel system(s) of German dialects in South Tyrol." Workshop "Phonetics meets dialectology: Vowel dynamics in non-standard varieties", Free University of Bozen-Bolzano.
- Kruijt, Anne. 14 July 2022. "VinKo: Varieties in Contact." 3rd VerbaAlpina conference "User Interfaces/User Experience vor dem Hintergrund traditioneller Arbeitsweisen", Ludwig-Maximilians-Universität, Munich.
- Kruijt, Anne. 11 September 2023. "Phrasal verbs in Venetan dialects." 17th Cambridge Italian Dialect Syntax-Morphology Meeting, University of Zurich.
- Pons, Aline & Gianmario Raimondi. 29 September 2023. "Paesaggio culturale, escursionismo e toponomastica in Valle d'Aosta. Il progetto DE-TOURS". Conference "Toponomastica Alpina", San Pietro di Cadore (Belluno).
- Pons, Aline. 4 October 2023. "Occitan alpin." Conference "Patois d'ici et patouà d'ailleurs", Centre Régional d'Études des Populations Alpines, Champsec (CH).
- Pons, Aline. 10 June 2024. "Il progetto 'Toponomastica delle Valli Valdesi': nomi di luogo nella rete del patrimonio culturale". Conference "Toponomastica e patrimonio culturale. Dalla raccolta alla restituzione", University of Turin.
- Pons, Aline. 12 June 2024. "Lingue di minoranza e cinema documentaristico: esperienze a confronto nelle valli alpine". Presentation at "Seminario di Linguistica", University of Turin.
- Pons, Aline & Gianmario Raimondi. 28 June 2024. "Le tracce della monticazione nella toponomastica. Il caso della Bassa Valle d'Aosta". Conference "Geografia e Patrimonio", Vercelli.
- Rabanus, Stefan & Anne Kruijt. 6 July 2022. "Zur Validität von mit Crowdsourcing-Methoden erhobenen morphologischen Daten." 7th conference of the International Association of Dialectology of German (IGDD), University of Salzburg.
- Rabanus, Stefan. 16 December 2022. "AlpiLinK: German-Romance Language Contact in the Italian Alps. Documentation, explanation, participation." "Conferenza Generale degli Studi Germanici in Italia", Rome, CNR.
- Rabanus, Stefan. 27 April 2023. "Citizen Science für Sprachwissenschaft und Sprachdidaktik". Conference "Neue Herausforderungen im Bildungssystem und neue sprachliche Realität", Yerevan State University.

- Rabanus, Stefan & Anne Kruijt. 24 October 2023. "From VinKo to AlpiLinK: web-based long-term storage and accessibility of information." Final VerbaAlpina conference "Publish and perish? Zur Persistenz webbasierter Forschung", Ludwig-Maximilians-Universität, Munich.
- Rabanus, Stefan. 30 November 2023. "Deutsch in den Südalpen: Isolation und Kontakt. Vorstellung des Projekts AlpiLinK." Presentation at "Gesellschaft für deutsche Sprache", Madison, WI.
- Rabanus, Stefan. 5 December 2023. "AlpiLinK. German-Romance Language Contact in the Italian Alps: Documentation, explanation, participation." Presentation at Max Kade Institute for German-American Studies, Madison, WI. The video of the talk can be found on the AlpiLinK homepage or at <https://youtu.be/RMZT67-dThE>.
- Rabanus, Stefan. 4 July 2024. "AlpiLinK. German-Romance Language Contact in the Italian Alps: documentation, explanation, participation." Keynote talk at the conference "Minority Languages in European Societies" (MiLES), University of Turin.
- Rabanus, Stefan. 26 October 2024. "Language Contact at the Morphological Level: Personal Pronouns in Cimbrian and Mòcheno." Workshop "The Limits of Convergence in Language Contact Situations", University of Verona.
- Rabanus, Stefan. 26 October 2024. "The AlpiLinK Database." Workshop "The Limits of Convergence in Language Contact Situations", University of Verona.
- Rabanus, Stefan. 17 December 2024: "Möglichkeiten und Grenzen der digitalen Sprachkartographie am Beispiel von VinKo und AlpiLinK." Presentation at the Berlin State Library, Cartography Department.
- Rabanus, Stefan & Barbara Vogt. 29 August 2024. "The role of contact in the history of research on the German language islands in the southeastern Alps." 16th International Conference on the History of Languages Sciences (ICHoLS), Ivane Javakhishvili University, Tbilisi.
- Rabanus, Stefan. 6 May 2025. "AlpiLinK – Deutsche Dialekte und Minderheitensprachen im Sprachkontakt in Norditalien". IDS-Workshop "Deutsch im europäischen Sprachraum", Leibniz-Institut für Deutsche Sprache Mannheim.
- Raimondi, Gianmario. 22 July 2022. "L'APV: eredità scientifica e prospettive di un atlante linguistico galloromanzo in Italia." Academic seminar "Dalla lingua alla cultura e ritorno: geolinguistica, etnolinguistica e varietà locali in Italia e in Europa. Prima parte", University of L'Aquila.
- Raimondi, Gianmario, Giovanni De Gasperi, Neri Binazzi & Matteo Rivoira. 1 December 2022. "Presentazione dell'Atlante Linguistico ed Etnografico Informatizzato della Conca Aquilano (ALEICA)." Academic seminar "Dalla lingua alla cultura e ritorno: geolinguistica, etnolinguistica e varietà locali in Italia e in Europa. Seconda parte", University of L'Aquila.
- Raimondi, Gianmario. 23 September 2023. "L'identità del francoprovenzale, fra gruppo linguistico e parlate locali." "Giornata delle lingue minoritarie", Pomaretto (Torino).
- Raimondi, Gianmario. 10 June 2024. "La media montagna valdostana e i suoi toponimi: il progetto DETOURS." Conference "Toponomastica e patrimonio culturale. Dalla raccolta alla restituzione", University of Turin.
- Raimondi Gianmario. 21 June 2024. "Il lessico della transumanza e della monticazione in Valle d'Aosta." Conference "Transumanza: lingua, memoria e innovazione", University of Turin.
- Saracco, Caterina. 11 July 2023. "Multilinguismo sulle Alpi: le comunità Walser dal Medioevo a oggi." Conference "Noi, loro, gli altri: il plurilinguismo medievale come campo d'azione", University of Bologna.
- Saracco, Caterina & Raffaele Cioffi. 9 February 2024. "Il nuovo dizionario Titsch – Italiano/Tedesco della parlata walser di Gressoney." Presentation at the "Circolo Linguistico Fiorentino", University of Florence.

- Saracco, Caterina. 19 September 2024. “Germanic-Romance Linguistic Contact in the Western Alps: Modal Verbs in the Walser Variety of Gressoney.” 57th International Congress of the Italian Linguistic Society (SLI), University of Catania.
- Tagliani, Marta, Chiara Melloni & Maria Vender. 12 July 2024. “Morpheme-based reading in typical and poor readers. A new tool for the assessment in Italian.” Workshop “Read & Lit’ Eval’ 2024”, INSPE Toulouse.
- Tagliani, Marta, Maria Vender & Chiara Melloni. 9 September 2024. “Enhancing Morphological Awareness in Typical and Atypical Readers: Effects of a Primary School Intervention on Morphological and Reading Skills.” Conference “Langues & Langage à la croisée des Disciplines LLcD 2024”, Sorbonne University Paris.
- Tagliani, Marta, Maria Vender, Chiara Melloni, Sabrina Piccinin & Serena Dal Maso. 28 November 2024. “The Role of Morphological Awareness in Reading and Text Comprehension in Italian.” LinE – Language in Education Conference “Communication Strategies in Teaching: Languages, Language Varieties, and Disciplines”, University of Trento.
- Tomaselli, Alessandra, Romano Madaro & Ermenegildo Bidese. 5 June 2024. “On the correlation between Germanic V2 and the low realization of the infinitival marker. Evidence from Germanic varieties spoken in NE-Italy.” Conference “Formal approaches to minority, minoritized or less studied languages in contact situations”, IKER (CNRS), Bayonne.
- Videsott, Ruth. 12 October 2024. “The Fear of Linguistic Contact in the Teaching of Ladin.” International Conference “De Bèlt um ins, the Mòcheno and Other Germanic Languages: Strategies and Perspectives”, Palù del Fersina, Mòcheno Cultural Institute.

WP7.3 Advanced academic teaching

- Alber, Birgit. 12 April 2023: Presentation (2 hours) of VinKo, AlpiLinK and VinKiamo at the MA seminar “Deutsche Dialektologie”, MA programme in Applied Linguistics, Free University of Bozen-Bolzano.
- Alber, Birgit. 25 November 2022. “Wortbildung im Bairischen – Word formation in Bavarian”. Workshop (4 hours), PhD programme in Linguistics, University of Bozen-Bolzano.
- Alber, Birgit & Angelica Bonelli. 24 May 2025. “Un territorio, tante lingue. Mehrsprachigkeit in der Euregio.” Presentation (1 hour) at “Euregio-Akademie”, Rovereto.
- Benedetto Mas, Paolo. 14 July 2023. “Regard (socio)linguistique sur le domaine francoprovençal en Italie”. Lecture (2 hours) at the Summer School “Langues et cultures alpines en devenir: l’aire francoprovençale”, Saint-Nicolas (Aosta).
- Bertollo, Sabrina. 24 October 2024: “VinKiamo as a Best Practice for Citizen Science: The Example of German Varieties in the Alps”. PhD seminar (4 hours), PhD programme in Linguistics, University of Verona.
- Gaeta, Livio. 7–8 February 2023. Cycle of classes (7 hours) on “The German-speaking minorities in Italy and in Friuli”, Scuola Superiore Universitaria di Toppo Wassermann, University of Udine.
- Gaeta, Livio. Second semester 2022/2023. Cycle of classes (18 hours) on “Digital German Linguistics”, MA programme in Language Technologies and Digital Humanities, University of Turin.
- Gaeta, Livio. Second semester 2023/2024. Cycle of classes (18 hours) on “Digital linguistics, Low-Density Language, and Walser varieties”, MA programme in Language Technologies and Digital Humanities, University of Turin.
- Gaeta, Livio. 23–24 May 2024. “The historical Germanic minorities in Italy between complexity and simplification”, PhD seminar (6 hours), PhD programme in Linguistics, University of Rome III/University “La Sapienza”.

- Gaeta, Livio. 11–22 November 2024. “Morphologie und Morphologisierung kontrastiv.” Cycle of MA classes (20 hours), University of Osnabrück.
- Gaeta, Livio. 11 December 2024: “The Walser Islands: Digital Resources and Language Preservation,” Lecture (3 hours), MA programme in Foreign Languages and Literatures, University of Ferrara.
- Kokkelmans, Joachim. 19 April 2023. “Halbautomatisierte Annotation von Dialektaufnahmen (VinKo & AlpiLink).” Guest presentation (2 hours) at the MA seminar “Deutsche Dialektologie”, MA programme in Applied Linguistics, Free University of Bozen-Bolzano.
- Rabanus, Stefan. 2 July 2024. “Investigating linguistic variables using AlpiLink data.” Workshop (4 hours) at the Summer School “Minority Languages in European Societies” (MiLES), University of Turin.
- Rabanus, Stefan. 26 February 2025. “Investigating Linguistic Variables Using Oral Data: The AlpiLink Corpus.” PhD seminar (4 hours), PhD programme in Linguistics, University of Verona.
- Rabanus, Stefan. Second semester 2024/2025. “Sprachvariation und Sprachkontakt am Beispiel der deutschsprachigen Inseln in Norditalien.” MA course (30 hours), Alpen-Adria-Universität, Klagenfurt.
- Raimondi, Gianmario. 6 December 2022, 20 December 2022. “Le francoprovençal et son «identité”, “Du paysage au paysage humain: l'anthroponymie francoprovençale.” Two lectures (2 hours) for the “Cours transfrontalier d'introduction au francoprovençal” (online), organized by Fondazione Natalino Sapegno.
- Saracco, Caterina. 19 May 2023–25 June 2023. Cycle of classes (8 hour) on Romance-Germanic plurilingualism in the Walser communities of Piemonte and Valle d'Aosta for various MA programmes at the University of Wrocław.
- Tomaselli, Alessandra. Second semester 2023/2024. Cycle of classes on Cimbrian syntax (18 hours) at MA seminar “Comparative Germanic Syntax”, MA programme in Linguistics, University of Verona.
- Tomaselli, Alessandra. Second semester 2024/2025. Cycle of classes on Cimbrian syntax (18 hours) at MA seminar “Comparative Germanic Syntax”, MA programme in Linguistics, University of Verona.

WP8: Public outreach: speech-community participation

WP8.1 VinKiamo

The term “VinKiamo” originates from a collaboration between the University of Verona and the Regional School Office for Veneto in 2021. Since then, it is used to cover all AlpiLinK activities carried out in collaboration with schools, regardless of region. VinKiamo constituted the most important part of the participatory aspect of AlpiLinK. Team members collaborated with numerous schools in Veneto, Friuli-Venezia Giulia, South Tyrol, Trentino and Aosta Valley and realized three types of interconnected activities:

- Raising awareness of multilingualism/plurilingualism: this was always part of the training in all VinKiamo activities but formed the central focus in the ‘linguistic autobiography’ activity.
- Data collection and active citizenship: secondary school students took active part in scientific data collection by acting as intermediaries between participants and AlpiLinK’s crowdsourcing data collection. They assisted with finding participants for the linguistic questionnaires as well as for the Linguistic Landscape part of the project.
- Data analysis and scientific method: these activities provided training that creates a link between linguistic research and education in secondary schools. The training aimed at promoting the use of innovative teaching methods that make use of databases and primary-source language materials and consisted of meetings with teachers on the one hand and interactive workshops for students on the other.

Including the VinKiamo activities that ran in Veneto during the VinKo project, the numbers concerning VinKiamo activities are listed below. While the activities dedicated to data collection ended with the closure of the project, the activities for raising awareness of multilingualism/plurilingualism and data analysis and scientific method will continue in the future.

VinKiamo	Veneto	Trentino	Südtirol	Valle d'Aosta	Friuli-Venezia Giulia	Total
Schools	18	6	9	2	1	36
Students	627	84	369	56	25	1,161
Teachers	80	7	16	5	5	113

The following schools were involved in the project activities.

- VinKiamo Veneto: Liceo Ginnasio Statale G.B. Brocchi, Bassano del Grappa (VI); IS Galilei Tiziano, Belluno; Liceo Renier, Belluno; IIS Nightingale, Castelfranco Veneto (TV); Istituto Marconi, Conegliano (TV); IIS G.B. Ferrari, Este (PD); IIS Sartori-Rosselli, Lonigo (VI); IIS Marie Curie, Garda (VR); Liceo De Fabris, Nove (VI); IC Bosco Chiesanuova, Roveré Veronese (VR); IIS Pasini, Schio (VI); Istituto Duca Degli Abruzzi, Treviso; IIS Trissino, Valdagno (VI); Liceo Fracastoro, Verona; Liceo Montanari, Verona; IIS Michele Sanmicheli, Verona; LS Quadri, Vicenza; Liceo Medi, Villafranca (VR).
- VinKiamo Trentino: ITET Pilati, Cles (TN); Liceo B. Russell, Cles (TN); IIS Martino Martini, Mezzolombardo (TN); Istituto San Michele all'Adige (TN); Liceo Classico G. Prati, Trento; Liceo Linguistico S. M. Scholl, Trento.
- VinKiamo Südtirol: Sozialwissenschaftlichen Gymnasium Meran; Sprachengymnasium Meran, Kunstgymnasium Cademia, St. Ulrich; Wirtschaftsfachoberschule Raetia, St. Ulrich, Oberschulzentrum (Linguistische Linie) La Ila; Technologische Fachoberschule Bruneck; Oberschule Sand in Taufers; Oberschulzentrum Mals; Landesberufsschule für das Gast- und Nahrungsmittelgewerbe Emma Hellenstainer, Brixen.
- VinKiamo Valle D'Aosta: Liceo Classico Bilingue, Aosta; Institut Agricole Régional, Aosta.
- VinKiamo Friuli: ISS Paschini-Linussio, Tolmezzo (UD).

Furthermore, the VinKiamo project was selected in 2024 for the so-called “VQR Terza Missione” (Research Quality Evaluation). As a result, VinKiamo was one of the top eight projects from the University of Verona which were submitted to the Ministry of University and Research (MUR) for the afore-mentioned purposes.

The following sections provide an overview of the VinKiamo activities per unit. Cf. also the VinKiamo table in *WP5.2 Data collection*.

Free University of Bozen-Bolzano: VinKiamo Südtirol

Team

Birgit Alber

Joachim Kokkelmans

Emily Siviero (as a member of the iNEST project)

Anna Pilsbacher (as a member of iNEST project)

Ruth Videsott

Alexander Glück

Angelica Bonelli

For up-to-date information, see the dedicated website here: <https://vinkiamo.projects.unibz.it/>.

The pilot project aimed at the data collected within the VinKo project. The pilot was conducted in the second semester of the scholastic year 2022/2023 at the Sozialwissenschaftliche Gymnasium Meran. It included four sessions of teaching and presentation of the project at the school itself. Results: 43 sets of Tyrolean data and diversifying the sample by raising the overall participants' age (from 24 to 31 year) and by attaining a more balanced gender distribution with respect to the previous data collection.

After that, the activities were organized in study days organized by the Free University of Bozen-Bolzano that invited students from local schools to the Faculty of Education in Brixen/Bressanone - "Tag der historischen Mehrsprachigkeit/De dl plurilinguism storich/Giornata del plurilinguismo storico" – which offered a variety of different workshops on a large range of linguistic topics by team members and guests. Subsequently, the students engaged in finding participants for the AlpiLinK data collection.

- 1st edition, 24 October 2023, 170 students.
- 2nd edition, 5 March 2024, 25 students.
- 3rd edition, 23 October 2024, 60 students.
- 4th edition, 11 March 2025, 33 students.

University of Trento: VinKiamo Trentino

Team

Patrizia Cordin

Ermenegildo Bidese

Michele Cosentino

Serena Bissolo

For up-to-date information, see the dedicated website here: <https://vinkiamo.unitn.it/>.

The VinKiamo Trentino activities were organized in training units on multilingualism and the AlpiLinK project held at the University of Trento or at the participating school, followed by autonomous project work of the students (finding informants and supporting them in participating at the AlpiLinK data collection).

- 1st edition, totaling the full participation of 48 students:
 - 10, 17, 24 October 2023: six hours of training provided at the University of Trento for the students from the participating schools in the city of Trento (see list above).
 - 19 February 2024: four hours of training provided for the students of the school in Mezzolombardo.
 - 21 May 2024: concluding event "Vivere il Multilinguismo. Le varietà linguistiche del nostro territorio", featuring a presentation on the benefits of multilingualism and the contribution of the students in the project (Patrizia Cordin, Bilinguismo Conta Trento), the awardings of distinctions to the most active students and the interactive lab "Le lingue dentro di noi: disegnare la propria sagoma linguistica".
- 2nd edition, featuring presentations AlpiLinK presentations directly at the school in San Michele all'Adige:
 - 9 January 2025: AlpiLinK project presentation (Ermenegildo Bidese, Serena Bissolo).
 - 18 February 2025: AlpiLinK training session for German language teachers (Serena Bissolo).
- 3rd edition, totalling the participation of 33 students from schools in Cles:
 - 12 March 2025: AlpiLinK training unit (Serena Bissolo).

University of Verona: VinKiamo Veneto/Friuli

Team

Stefan Rabanus
Sabrina Bertollo
Marta Tagliani
Francesca Bussola
Katharina Knapp
Ilaria Driussi

For up-to-date information, see the dedicated website here: <https://sites.hss.univr.it/vinkiamo/>.

The “classical” VinKiamo activities were organized in training units on pluri- and multilingualism and the AlpiLinK project, followed by autonomous project work in which the students found and supported informants who participated at the data collection. Additionally, there were workshops and training units in specific topics.

- VinKiamo.3. 10/2022–01/2023. The project involved 120 students from seven schools in the provinces of Verona, Vicenza, Padova, Treviso and Belluno. The training comprised three classes held at the University of Verona by Stefan Rabanus, Sabrina Bertollo, Marta Tagliani and Francesca Bussola. Results: 382 datasets for the VinKo data collection (368 for Venetan, eight for Trentino, two for Tyrolean, and one each for Saurano, Sappadino, Anpezan Ladin, Badiot Ladin).
- Training on linguistic autobiography. 03/2024. Teacher training and workshops in class in five schools in the provinces of Verona, Vicenza and Treviso with 120 students (Katharina Knapp).
- VinKiamo.4 Friuli. 01/2024–04/2024. The project involved 25 students at the secondary school “Paschini Linussio” in Tolmezzo (province of Udine). The training comprised three meetings in Tolmezzo (two of them in mixed mode, online and in presence) conducted by Stefan Rabanus, Sabrina Bertollo and Marta Tagliani on the AlpiLinK project and data collection. During their project work the students found and assisted informants who provided 63 sets of audio recordings elicited by AlpiLinK’s linguistics questionnaire (54 for Friulian, four for Sappadino, five for Timavese).
- Training on data analysis and scientific method, entitled “La tosa, la buteleta o la fiola? Esplorazioni nelle lingue del Veneto.” 11/2023–03/2024. The training consisted of lectures (3 hours) given at the University of Verona (08 November 2023) for involved teachers. Subsequently (12/2023–03/2024) classes (4 hours each) were taught in four schools in the provinces of Verona and Vicenza by Stefan Rabanus, Sabrina Bertollo and Francesca Bussola.
- Event “Il multilinguismo fa scuola!”. 30 April 2024. The event was organized for teachers and students involved in VinKiamo.3 and VinKiamo.4. In total, 25 students and nine teachers from five schools in Veneto and Friuli-Venezia Giulia participated.
 - The morning featured presentations on AlpiLinK (Alessandra Tomaselli, Sabrina Bertollo, Stefan Rabanus) and the benefits of multilingualism (Patrizia Cordin and Maria Vender on behalf of “Bilinguismo Conta Trento”).
 - The ceremony of the awards for students with outstanding contributions received a lot of attention by the local press, see *WP8.2 Public relations*.
 - The afternoon featured five interactive laboratories on languages and multilingualism offered by team members of the AlpiLinK units from UniVR e UniBZ:
 - “Cosa fai mentre leggi? Il multilinguismo attraverso gli occhi della tecnologia” (Marta Tagliani, Andrea Nardon).
 - “Alla ricerca delle radici comuni. Un gioco di carte sulle lingue indoeuropee” (Stefan Rabanus).
 - “PuzzLing, il Puzzle del Linguaggio” (Roberto Zamparelli).
 - “Le lingue dentro di noi: disegnare la propria sagoma linguistica” (Angelica Bonelli).
 - “Il gioco delle sette differenze acustiche. El zugo dee sete difarènses acustiche. Le jüch dles set desfarènzies acustisches” (Joachim Kokkelmans).
- VinKiamo.5 Friuli. 01/2025–05/2025. The project involved 15 students at the secondary school “Paschini Linussio” in Tolmezzo (province of Udine). The training comprised three meetings in Tolmezzo (two of them in mixed mode online and in presence) conducted by Stefan Rabanus, Sabrina Bertollo, Marta Tagliani and Ilaria Driussi on the AlpiLinK project and data collection. During their project work the

students found and assisted informants who provided audio recordings elicited by AlpiLinK’s linguistics questionnaire and photographs for the “Linguistic Landscape” section.

- Two AlpiLinK presentations in collaboration with two “eco musei” in Friuli, organized by Ilaria Driussi with support of team members of the UniVR and UniBZ units and researchers of the University of Udine (members of the iNEST project). The presentations were meant to increase the number of audio recordings for less-covered areas of Friuli.
 - 17 March 2025: “Eco Museo Lis Aganis” in Maniago (province of Pordenone).
 - 5 May 2025: “Eco-Museo delle Acque” in Gemona del Friuli.
- Training in linguistic landscaping
 - 1st edition, 02/2025–04/2025. The training involved 28 students at the secondary school “Florence Nightingale” in Castelfranco Veneto (province of Treviso) and consisted of six lessons with a total amount of 15 hours, including active linguistic landscaping, i.e. taking photographs that were then inserted in the map in the “Linguistics Landscape” section (Katharina Knapp).
 - 2nd edition, 03/2025–05/2025. The training involved 24 students of the middle school “IC Bosco Chiesanuova” in Roveré Veronese (province of Verona) and consisted of two lessons of two hours each and an excursion to the Cimbrian village of Giazza in order to take photographs of the multilingual territory which were inserted in the map in the “Linguistics Landscape” section (Ilaria Driussi, Stefan Rabanus, Sara Scalia).

Aosta Valley University: VinKiamo Valle D’Aosta

Team

Gianmario Raimondi

Paolo Benedetto Mas

The pilot project “Minority Languages and New Technologies” was designed in 2023 and carried out by Gianmario Raimondi and Paolo Benedetto Mas (AlpiLinK), in collaboration with the Regional Education Authority of the Autonomous Region of Aosta Valley. The project involved 56 upper secondary school students (Aosta: Institut Agricole Régional and Liceo Classico Bilingue), and included three training sessions, an AlpiLinK data collection campaign, and a final session for discussion and feedback. Classes taught included the following topics:

- “Le lingue romanze e il francoprovenzale” (Gianmario Raimondi)
- “I repertori linguistici della comunità valdostana” (Paolo Benedetto Mas)
- “L’Atlas des Patois Valdôtains - APV/1 Le lait et les Activités Laitières” (Gianmario Raimondi)
- “Il progetto AlpiLinK: indicazioni tecniche e metodologiche su come condurre le inchieste” (Paolo Benedetto Mas)
- “Méthodes de conduite d’une enquête linguistique et transcription des données phonétiques” (S. Belley)
- “Esercizi pratici a partire dall’interpretazione e lettura delle carte linguistiche dell’APV” (Gianmario Raimondi)

The units of UniVR and UniBZ bought various types of material to promote the VinKiamo activities. The following table shows various typologies.

Typology	Costs	Unit
Tote bags with AlpiLinK logo (500 items)	€ 1,348.10 (VAT included)	UniVR
AlpiLinK roll-up for conferences, workshops and VinKiamo activities	€ 122 (VAT included)	UniVR
Gadgets (powerbanks) for awardings students with outstanding contributions	€ 231.80 (VAT included)	UniVR
Gadgets (books) for awardings students with outstanding contributions	approx. € 150 (VAT included)	UniVR
Gadgets (various) for awardings students	approx. € 150 (VAT included)	UniBZ

WP8.2 Public relations

The public communication for the project has been expedited to the PR company Blum (contact: Roberta Voltan). They were in charge of the organization and communication of the project to news agencies and liaison with press offices of the university.

Unit	Costs
UniVR	€ 4,514 (VAT included)
UniBZ	€ 4,514 (VAT included)
UniTN	€ 4,514 (VAT included)
UniTO	€ 4,514 (VAT included)
UniVDA	€4,514 (VAT included)

Press conferences:

- 30 January 2024: Press launch of the AlpiLinK project; hosted at the University of Verona. It resulted in 49 articles written on the project.
- 30 April 2024: Event “Il multilinguismo fa scuola!”. The event was organized for teachers and students involved in VinKiamo.3 and VinKiamo.4. The ceremony of the awards for students with outstanding contributions received a lot of attention from the local press, cf. the interview in the table below (30 April 2024).
- A press conference dedicated to the conclusion of the project and future perspectives will take place on 5 December 2025.

Press releases issued by Blum in the period 10/2023–05/2025 generated 226 items of newspaper or website articles and radio or TV-station interviews to AlpiLinK team members. The overall newspaper and website space dedicated to AlpiLinK articles had an “Advertising Value Equivalency” (AVE) of € 523,000, the articles reached 4,800,000 potential readers.

The “Press” section of the AlpiLinK website features a sample of Italian and German articles concerning the various language areas and topics. The full list of published articles and released interviews is available here:

<https://public.reputation.onclusive.com/Public/ReviewKiosk?ticket=80EC2BFCDAB29C5AE7566934D192355734B566ACC02CCF5F3FFB43E63B39C6E37A3B9EE117A59A5532FCF012657547A23BF1776050290134AB6D3B03DF4C42AF9A129FF1FFDD0233ADA0D559911ED125EC342499A5DDAF71EDD4041859B31711D29766D783333D61421706247E9E8AB2DEFEC122F92E15AF5310B50CD14C14B4>

Additionally, the press office of the University of Verona informed about the proceedings of the project regularly in the “UniVrMagazine”, see here:

<https://www.univrmagazine.it/2025/01/15/giornata-nazionale-dei-dialetti-il-report-alpilink/>

The following table contains a sample of TV or radio-station interviews which at the time of writing (July 2025) are still available online.

Date	Channel	Interviewee	Unit	Link
15 October 2023	RAI Radio 3	Birgit Alber	UniBZ	https://www.raiplaysound.it/audio/2023/10/Futuradio---La-festa-di-Rai-Radio-3-a-Bolzano-del-15102023-f0bdf0b3-4132-426e-b38d-29a54b3c93ae.html
30 January 2024	RAI TGR Veneto (TV)	Stefan Rabanus	UniVR	https://www.rainews.it/tgr/bolzano/video/2024/02/alluniversita-di-verona-presentato-un-portale-dedicato-ai-dialetti-e-alla-lingue-minoritarie-7aa3abb9-7573-434d-9f56-32eefdcfd49d.html
9 February 2024	RAI Trentino (TV)	Ermenegildo Bidese	UniTN	https://www.rainews.it/tgr/trento/video/2024/02/i-dialetti-trentini-non-spariranno-progetto-alpilink-studio-dialetti-arco-

				alpino-5157564b-167b-4d8f-90d0-9e93888b4b3c.html
16 February 2024	Radio Number One	Stefan Rabanus	UniVR	https://radionumberone.it/podcast/alpilink-stefan-rabanus-dialetti-non-sono-italiano-difettoso
23 February 2024	Centro Studi Dialogo (TV)	Stefan Rabanus	UniVR	https://www.youtube.com/watch?v=LuPrNs_jU0jQ
5 March 2024	RAI Südtirol (radio)	Birgit Alber, Stefan Rabanus, Ruth Videsott, Emily Siviero, Joachim Kokkelmans	UniBZ, UniVR	https://www.raibz.rai.it/feed.php?id=111
30 April 2024	Tele Arena	Stefan Rabanus	UniVR	https://www.telearena.it/programmi-di-informazione/tg/tg-provincia-ed-economia/tg-provincia-ed-economia-1.10694219
27 May 2024	7 Gold Tele Padova	Sabrina Bertollo, Stefan Rabanus	UniVR	https://www.youtube.com/watch?v=Ek3dZjdVtqg
9 June 2024	Tele Arena	Stefan Rabanus	UniVR	https://www.telearena.it/programmi/cronache-dalla-provincia/il-dialetto-e-ancora-la-nostra-lingua-1.10732005
2 December 2024	Radio Veneto 24	Stefan Rabanus	UniVR	https://www.veneto24.it/podcasts/54715/episodi/1060179 (from minute 57)
17 January 2025	Radio Veneto 24	Sabrina Bertollo	UniVR	https://www.veneto24.it/podcasts/54715/episodi/1115759 (from minute 32)

WP8.3 Community support

An important objective of the project is to actively support the language communities in taking care of their languages. The most important activity in this area is the agreement signed with the cultural association “Tzimbar” for the creation of an up-to-date dictionary of the Cimbrian variety of Giazza. The dictionary will be added to the online platform for the dictionaries of 7-Municipalities Cimbrian (*Bóartpuuch*: <http://dizionario.cimbri7comuni.it/>) and Lusérn Cimbrian (Zimbarbort: <https://zimbarbort.istitutocimbri.it/>). With AlpiLinK and other research funding, an 8-months research scholarship has been financed (scholarship holder, Riccardo Ferracin) and IT services have been ordered from Chambrà d’Oc.

Team	Description	Unit
Alessandra Tomaselli	Scientific supervision	UniVR
Stefan Rabanus	Scientific supervision	UniVR
Riccardo Ferracin	Lexicographic work	“Tzimbar”
Nicolò Boniolo	Lexicographic work	“Tzimbar”
Carlo Zoli	Technical supervision	UniBZ

Funding	Objective	Costs
AlpiLinK	Research scholarship	€ 8,382.50
Alessandra Tomaselli	IT services	€ 10,400 (VAT included)

Appendix 2.1 Website pages

This is the complete overview of all pages in the AlpiLinK platform and their latest updated versions (only major content updates; references and internet resources in the language descriptions were continuously updated). The pages are indicated with their English names. Translations in the four platform languages (Italian, German, English, French) were provided and/or corrected by Anne Kruijt, Stefan Rabanus, Marta Tagliani, Ilaria Driussi and Joachim Kokkelmans. The updates concerned all four language versions.

Page	Content creator(s)	Unit(s)	Created	Updated	Update
Home	Anne Kruijt Stefan Rabanus	UniVR	12/2022	02/2024 04/2025 06/2025	Addition video Rewrite text “Listen & Explore” section, update screenshot Rewrite project description (data collection closed)
AlpiLinK	Anne Kruijt Stefan Rabanus	UniVR	12/2022	06/2025	Rewrite introduction
VinKo	Anne Kruijt Stefan Rabanus	UniVR	03/2023	01/2025	Rewrite text
VinKiamo	Stefan Rabanus	UniVR	03/2023	12/2023 06/2024	Rewrite text Rewrite text
Linguistic Landscape	Anne Kruijt	UniVR	03/2023	05/2025	Map update
Participate	Anne Kruijt	UniVR	03/2023	06/2025	Disabled
Listen & Explore	Anne Kruijt Stefan Rabanus	UniVR	12/2022	02/2025	Rewrite text Continuous updates of the numbers of the audio recordings
Team and Contact	Anne Kruijt	UniVR	03/2023		
Partner	Anne Kruijt Stefan Rabanus	UniVR	03/2023	06/2025	Eurac added
Events	Anne Kruijt	UniVR	06/2023	06/2025	Disabled
Press	Anne Kruijt	UniVR	02/2024	05/2024 07/2024	New article added New articles added
Results	Anne Kruijt Stefan Rabanus Angelica Bonelli	UniVR UniBZ	08/2024	12/2024 05/2024	New language version Rewrite data-collection article
Our varieties	Anne Kruijt	UniVR	12/2022	02/2024	Map update
Cimbrian	Alessandra Tomaselli	UniVR	03/2023		
Francoprovençal	Paolo Benedetto Mas Gianmario Raimondi Matteo Rivoira Aline Pons	UNIVDA UniTO	03/2023		
Friulian	Barbara Vogt Stefan Rabanus Ilaria Driussi	UniVR	03/2023	11/2024	Correction and rewriting after feedback from ARLeF

Ladin	Patrizia Cordin Jan Casalicchio Ruth Videsott	UniTN UniBZ	03/2023	02/2024	Correction and rewriting after feedback from Veneto Ladin communities
Lombard	Patrizia Cordin Anne Kruijt	UniTN UniVR	11/2022		
Mòcheno	Birgit Alber	UniBZ	03/2023		
Piedmontese	Matteo Rivoira	UniTO	03/2023		
Occitan	Paolo Benedetto Mas Gianmario Raimondi Matteo Rivoira Aline Pons	UniVDA UniTO	03/2023		
Resian	Stefan Rabanus	UniVR	01/2024		
Sappadino	Barbara Vogt	UniVR	03/2023		
Saurano	Barbara Vogt	UniVR	03/2023		
Timavese	Barbara Vogt	UniVR	12/2023		
Tyrolean	Birgit Alber	UniVR	03/2023		
Trentino	Patrizia Cordin	UniTN	03/2023		
Val Canale Slovenian	Stefan Rabanus	UniVR	01/2024		
Val Canale German	Barbara Vogt	UniVR	12/2023		
Venetan	Alessandra Tomaselli Anne Kruijt	UniVR	03/2023		
Walser German	Paolo Benedetto Mas Gianmario Raimondi Matteo Rivoira Aline Pons Silvia dal Negro?	UniVDA UniTO UniBZ	01/2023		
Citation	Anne Kruijt	UniVR	06/2023		

Appendix 2.2 How to use Phonic

Sign in on <https://www.phonic.ai/> using the following credentials: User: anne.kruijt@univr.it. Password.

Check incoming responses

Go the tab ‘Survey’. Here you can see the number of credits that have been used and that remain for the month. Click on ‘AlpiLinK’ to open the AlpiLinK survey. You will see 5 tabs: Respondents, Questions, Insights, crosstabs and Share.

Most important tab for keeping tab on participants, new and old, is ‘Respondents’. It provides a list of all respondents starting from the oldest to the newest ones. It provides the full list of all answers provided for a specific respondent. At the top it will also tell you the session id (important for identifying participants) and the status (finished or in progress), which indicates if the participant completed the questionnaire or not. If a participant does not click ‘finish’ at the end of the questionnaire it will remain ‘in progress’, so have a scroll down to make sure that new respondents are indeed still in progress or if they did complete the whole questionnaire. In the ‘Share’ tab, you find the link to share the survey. This is the link embedded on the AlpiLinK website.

Keep track of new respondents by checking which participants have come in and if they completed the questionnaire or not. Check the status and scroll through the responses of the questionnaire to check if audio has actually been stored and if the respondent is not a test (age=1). If the respondent is completed, check the check mark next to the respondent. See image below: black indicates the survey has been bookmarked. White that it hasn’t been bookmarked. By bookmarking all completed questionnaires, it is easy to export only the surveys that you want for the data export.

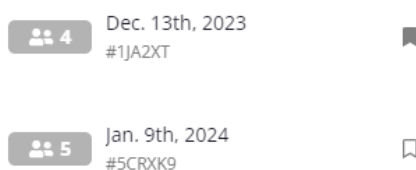


Figure 10: Bookmarks in Phonic.

If a questionnaire is entirely empty or is a test questionnaire (age=1), it can be deleted by clicking the triple dots in the respondent questionnaire and click ‘delete respondent’. See figure below. **Note: This cannot be undone and any data deleted is entirely deleted from the server.** Be sure all data has either been stored outside of phonic or it is really not needed!

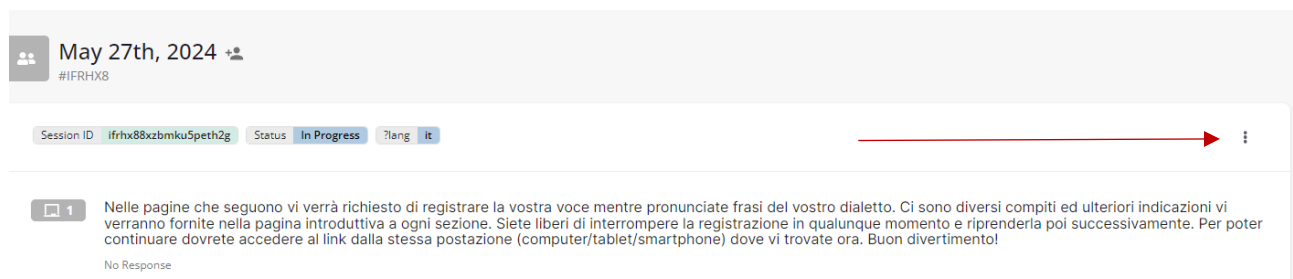


Figure 11: Delete respondent.

Surveys that come in and that include real participants but that have not been completed should remain on the server without a bookmark up to 6 months before being exported. See section ‘Export the data’ for what steps to take with the incomplete questionnaires.

Adapt the survey structure and/or questions

To edit the survey structure or to change the questionnaire in any way, open the questionnaire in the Survey tab and go to Options > Edit Survey.

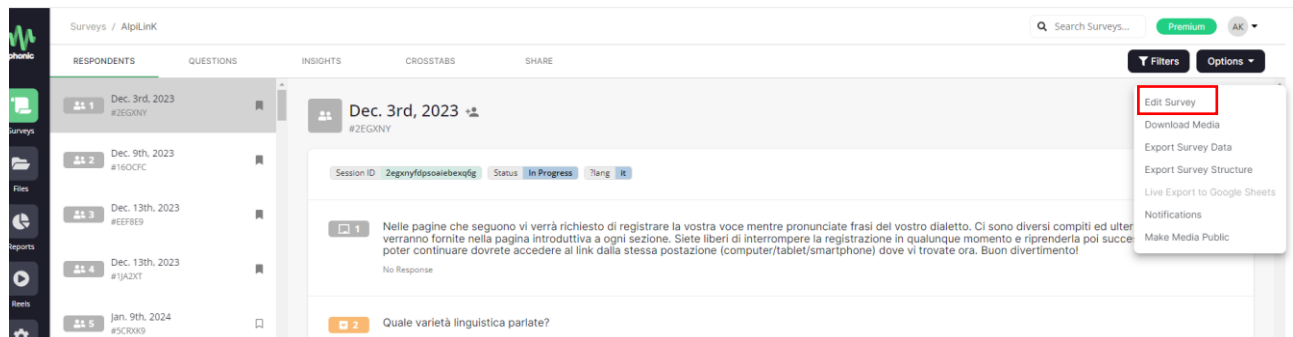


Figure 12: Phonic edit survey.

This will open the Survey Builder (Note: it might take a few seconds to open). On the left it gives the text and type (multiple choice, audio etc) of the questions, and on the right, it provides a preview, see image below.

Note on Multiple choice options: If you change the options or the order of the multiple-choice options (e.g. language variety), **this change will affect ALL COLLECTED QUESTIONNAIRE, so ALSO previously collected questionnaires.** It won't be possible to afterwards find the data that was changed, so make sure to always first save a copy of the questionnaire data BEFORE making any changes. For example, if a speaker in October chooses option 2 (Francoprovençal) and in December you change option 2 into Venetan, the system will henceforth label this speaker as Venetan (or say option Changed in case of deletion of an option). This goes both for relabeling as well as for shuffling of the options. Option 2 for the system indicates the second option in the list as it is saved in the current version of the questionnaire, regardless of what it was in the past.

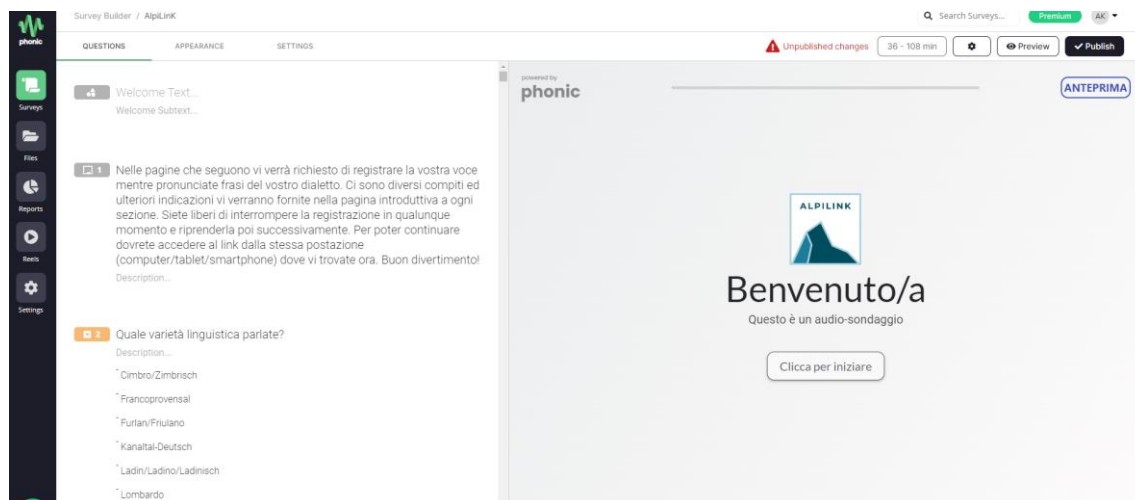


Figure 13: Phonic survey builder.

When hovering over a question you will see the symbols below appear (circled in red). The first one (alphabet) allows you to edit translations. Here you can modify the German translation of a given question. The second (image) allows you to add multimedia files to the questions. For example, this is often done in task instructions, in the form of audio to give participants a sample of a possible answer. The third (connections) is the place to edit question logic, more about this below. The fourth (cog) opens the general settings of a question, here you will find the Data Label (e.g. S01), and among other things, whether a question is optional or not, min/max

time, min/max selection options, and randomization. With the fifth symbol (three dots) you can duplicate or delete a question.

Note on randomization: There are currently two randomization groups in the questionnaire; the translation section (group 2) and the image description section (group 1). When using randomization, **the final question of the group must be present in BOTH ITALIAN AND GERMAN**. If not, the internal logic will skip the question if the other survey language is picked and this will phase the participant into the next task, regardless of how many of the questions in the section they did. The name truncation task cannot be randomized because it would not keep the “None of the above”, other part of the question together with the initial question.

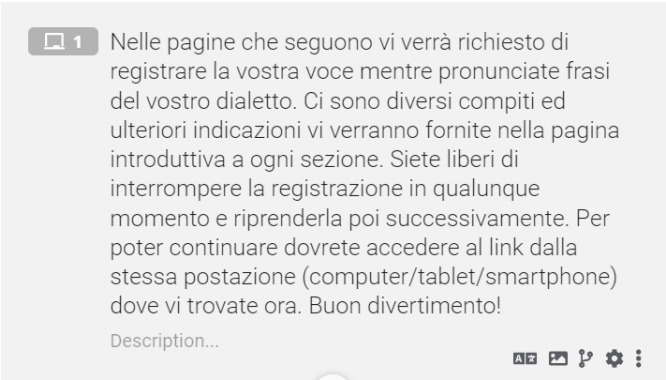


Figure 14: Phonic survey question.

The question logic tab is very important for determining which participants see what questions/which language the survey is in. Apart from language selection, there are two conditions set in the questionnaire. Condition 1 is called ‘German’, see below under (a). If participants get attributed ‘yes’ in this condition, they will receive the questions which are exclusively for German speakers and skip the questions which are exclusively for Romance speakers. If they have the attribute ‘no’, they get the Romance questions.

A) Condition ‘German’	B) Applying it in a question
<div><div><div>If</div><div>Answer</div><div>Is</div><div>Tiroler Dialekt</div><div>Then</div><div>Set language to</div><div>German</div></div><div><div>If</div><div>Answer</div><div>Is</div><div>Tiroler Dialekt</div><div>Then</div><div>Set condi...</div><div>German</div><div>=</div><div>yes</div></div><div><div>If</div><div>Answer</div><div>Is n...</div><div>Tiroler Dialekt</div><div>Then</div><div>Set condi...</div><div>German</div><div>=</div><div>no</div></div><div><div>If</div><div>Answer</div><div>Is n...</div><div>Tiroler Dialekt</div><div>Then</div><div>Set language to</div><div>Italian</div></div></div>	<div><div>Question Logic</div><div>Add conditional logic to your survey. See documentation for help.</div><div><div><div>If</div><div>Condition</div><div>:</div><div>German</div><div>Is</div><div>no</div></div><div><div>Then</div><div>Skip to</div><div>Next Question</div></div><div><div>Run</div><div>Before Question</div></div></div><div><div>+ Add Logic</div></div><div><div>Done</div></div></div>

Figure 15: Phonic question logic (a).

For example, in (b), see the use of the condition. In (b), the condition will be run before the question is presented. If the participant has ‘no’ in the German condition, they skip the question and are automatically moved to the next question. If they have ‘yes’, they will see the question. This condition is set in the language variety question (and if applicable, the ‘which language do you want the survey in’ question). NOTE: this is not the same as the ‘set language to’ logic, which set the survey language (automatic translates the buttons and it triggers the German translations as given under the ‘translation’ symbol).

The second condition is called DifName, and it determines whether participants get the full name truncation task or the single open question (N19). For all minority languages, the condition is set to ‘yes’. For the larger dialect areas, it is set to ‘no’. Below in (a) is the setting of the condition (in the language variety question) and in (b) the use of it.

a) Condition ‘DifName’

Add conditional logic to your survey. See [documentation](#) for help.

If

Condition

:

DifName

Is

yes

Then

Skip to

Question 142

Run

Before Question

+ Add Logic

Done

If

Answer

Cont...

Ladin

Then

Set condi...

DifName

=

yes

If

Answer

Cont...

Cimbri

Then

Set condi...

DifName

=

yes

If

Answer

Cont...

Möcheno

Then

Set condi...

DifName

=

yes

If

Answer

Cont...

Plodarisch

Then

Set condi...

DifName

=

yes

b) Applying it in a question

Figure 16: Phonic question logic (b).

Apart from the conditions, you can also set the ordering and skipping to questions in the tab skip. For example, below, participants that pick Cimbri/Zimbrisch are diverted straight to Q5 (Which province do you live in?) and Ladin speakers are first asked to specify the variety and which survey language they prefer (Q3).

AlpiLinK report

59

Question Logic

CONDITIONS	SKIP	DISPLAY OPTIONS
Cimbro/Zimbrisch		Q5. Provincia ▼
Francoprovensal		Q5. Provincia ▼
Furlan/Friulano		Q5. Provincia ▼
Kanaltal-Deutsch		Q5. Provincia ▼
Ladin/Ladino/Ladinisch		Q3. La varietà li... ▼
Lombardo		Q5. Provincia ▼

Figure 17: Phonic question logic (c).

Appendix 2.3 Map, included stimuli

Frase	Sätze	
IT	DE	stimulus
AlpiLink	AlpiLink	
Mario ieri ha ballato tanto con Sara	Martin hat gestern lange mit Sara getanzt	S01
Zappa questa aiuola subito, l'altra falla dopo	Grab dieses Beet jetzt gleich um, und dann das andere	S02
Il pane l'hai già mangiato o lo devi ancora mangiare?	Hast du das Brot schon gegessen oder wirst du es noch essen	S03
La frutta l'hai già portata o posso portarla io?	Hast du schon Obst mitgebracht, oder soll ich es mitbringen?	S04
Questa casa qua è più fredda di quella là	Das Haus hier ist kälter als das andere da	S05
Quello bianco è il tuo cane	Der weiße da ist dein Hund	S06
Questo è il mio cane, non quello	Das hier ist mein Hund, nicht der da drüben	S07
La casa di Sara è questa	Saras Haus ist das da	S08
Avete mica visto Mario?	Habt ihr Martin gesehen?	S10
Mi chiedo se Mario sia più furbo di me	Ich frage mich, ob Martin schlauer ist als ich	S11
Hai visto come Giovanni ha costruito la sua casa?	Hast du gesehen, wie Johannes sein Haus gebaut hat?	S12
Tu dormi più di lui	Du schläfst mehr als er	S13
Sara dice che mangiamo troppo	Sara sagt, dass wir zu viel essen	S14
Ognuno va per conto suo: noi con la nostra auto, voi con la vostra	Jeder fährt alleine: wir mit unserem Auto, ihr mit eurem	S17
D'estate i miei vicini tornano nella loro vecchia casa	Meine Nachbarn fahren im Sommer immer zu ihrem alten Haus	S18
In piazza non ci sono alberi alti	Auf dem Platz stehen keine großen Bäume	S19
Alcuni bambini piccoli parlano ancora in dialetto	Einige kleine Kinder sprechen noch Dialekt	S20
Ho visto pochi vecchi amici che stavano festeggiando	Ich habe viele alte Freunde gesehen, die am Feiern waren	S21
Ricordo ancora la tua vecchia gattina	Ich erinnere mich noch an dein Kätzchen von früher	S22
C'era Martina in piazza, l'hai vista? Cosa faceva?	Martina war draußen unterwegs, hast du sie gesehen? Was hat sie gemacht?	S23
Giovanni si sta lavando i capelli	Johannes wäscht sich gerade die Haare	S24
Mi siedo qui?	Ich setze mich hierher?	S27
So che ieri i miei amici hanno dovuto lavorare tutto il giorno	Ich weiß, dass meine Freunde den ganzen Tag haben arbeiten müssen	T01
So che Mario potrà comprarsi una macchina nuova	Ich weiß, dass sich Martin ein neues Auto kaufen können wird	T02
In alto in montagna è nevicato e in basso è piovuto	Oben auf dem Berg hat es geschneit, unten hat es geregnet	T03
Ieri non avete incontrato nessuno qui	Ihr habt hier niemanden getroffen	T04
Abbiamo parlato con Mario	Wir haben mit Martin gesprochen	T05
Ho abbracciato Sara	Ich habe Sara umarmt	T06
VinKo	VinKo	
In questa stanza Maria è la più bella	In diesem Saal ist Maria die schönste	S0015
Oggi tu hai studiato più di tutto	Du hast heute am meisten gelernt	S0016
Tuo fratello, lo voglio vedere domani	Deinen Bruder, den will ich morgen sehen	S0028
I compiti, li comincio a fare adesso	Die Aufgaben, die fange ich jetzt an	S0037





A Gianni, gli vado a portare una bottiglia di vino	Dem Hans, dem bringe ich eine Flasche Wein	S0038
Dimmi perché parte domani	Sag mir mal, warum er morgen abfährt	S0078, S0079
Mi ha chiesto se tua mamma ha ancora la tosse	Er hat mich gefragt, ob deine Mutter noch hustet	S0080
Non so quale autobus parte per primo	Ich weiss nicht, welcher Bus zuerst abfährt	S0082
Non gli ho mai chiesto quanti anni ha	Ich habe ihn nicht gefragt, wie alt er ist	S0083, S0084
Vorrei sapere chi è stato a rompere la finestra!	Ich wollte mal wissen, wer das Fenster zerbrochen hat!	S0086
Sai cosa ha comprato quello signora?	Weisst du, was jene Frau gekauft hat?	S0087
Chiedi ai tuoi zii chi hanno invitato alla festa	Frag deinen Onkel und deine Tante, wen sie zur Party eingeladen haben	S0088
Non so dove sia andata Marco	Ich weiss nicht, wohin Markus gegangen ist	S0115



Storie	Geschichten	
IT	DE	stimulus
In un paese di montagna, ogni sera i bambini devono raccogliere la legna nel bosco	In einem abgelegenen Dorf müssen die Kinder jeden Abend im Wald Feuerholz sammeln	T0101
Una sera, due di loro vedono una strega che ha visto altri bambini, ma non loro	Eines Abends sehen zwei von den Kindern eine Hexe, die die anderen Kinder gesehen hat, aber sie nicht	T0102
Tornano di corsa in paese e chiedono aiuto al cacciatore	Sie rennen ins Dorf zurück und bitten den Jäger um Hilfe	T0103
Noi siamo scappati, ma loro sono stati catturati	Wir sind entkommen, aber sie sind gefangen	T0104
Siete fortunati che lei non abbia visto voi!	Ihr habt Glück gehabt, dass sie euch nicht gesehen hat!	T0105
Ti do il mio fucile per difenderti	Ich gebe dir mein Gewehr zum Schutz	T0106
Se dai il fucile a lui, a me cosa dai?	Wenn du ihm das Gewehr gibst, was gibst du mir?	T0107
Trovano la casa della strega nel bosco buio	Sie finden das Haus der Hexe im dunklen Wald	T0108
La casa è vecchia e i bambini la trovano tremenda	Das Haus ist alt und die Kinder finden es schrecklich	T0109
La strega vuole cucinare i bambini presi	Die Hexe will die gefangenen Kinder kochen	T0110
Non mangiare lui, mangia me! Sono molto più grassa di lui!	Iss nicht ihn, iss mich! Ich bin viel dicker als er!	T0111
Non deve mangiare neanche te! Non deve mangiare nessuno di noi!	Nein, sie darf auch dich nicht essen! Iss kleinen von uns!	T0112
Il cacciatore dice alla strega di fermarsi, e lei tenta di fargli una magia	Der Jäger befiehlt der Hexe aufzuhören, aber sie versucht ihn zu verzaubern	T0113
Allora lui prende il suo fucile e uccide la strega	Er aber nimmt das Gewehr und tötet die Hexe	T0114
Libera i bambini e li porta a casa	Er befreit die Kinder und bringt sie nach Hause	T0115
Il cacciatore ha dato il fucile a noi, ma non a loro. Loro sono ancora troppo piccoli!	Der Jäger hat uns ein Gewehr gegeben, aber ihnen nicht. Sie sind noch zu klein!	T0116
Gli abitanti del paese ringraziano il cacciatore	Die Dorfbewohner danken dem Jäger	T0117
C'è una madre che ha due figlie, di cui una è molto pigra mentre l'altra è molto brava	Eine Mutter hat zwei Töchter, von denen eine sehr faul ist, während die andere sehr fleißig ist	T0301
La ragazza molto brava è seduta sul bordo del pozzo e ci cade dentro	Während das fleißige Mädchen am Brunnen arbeitet, fällt sie/es aus Versehen hinein	T0302

La ragazza pigra decide di seguirla e insieme arrivano in un altro mondo	Das faule Mädchen will ihr folgen und beide landen in einer anderen Welt	T0303
C'è un albero pieno di mele che le chiama: "Scuotetemi, ho frutti maturi!"	Da ist ein Baum voller Äpfel, der zu ihnen ruft: "Schüttelt mich, ich habe reife Äpfel!"	T0304
"Fai tu" dice la ragazza pigra a quella brava e lei aiuta l'albero e lo scuote	"Mach du es!", sagt das faule Mädchen zu dem fleißigen, und sie hilft dem Baum und schüttelt ihn	T0305
Poi arrivano a un forno che urla loro: "Tirate fuori il pane, è pronto!"	Dann kommen sie zu einem Ofen, der sie anspricht: "Nehmt das Brot raus, es ist fertig!"	T0306
"Adesso tocca a te", dice la ragazza brava a quella pigra, ma lei non aiuta il pane	"Jetzt bist du aber dran", sagt das fleißige Mädchen zu dem faulen, aber sie hilft dem Brot nicht	T0307
"Va bene, l'aiuto io", dice la ragazza brava e lei prende il pane dal forno	"Gut, dann helfe ich ihm", sagt das fleißige Mädchen, und sie nimmt das Brot aus dem Ofen	T0308
Avete raccolto voi le mele?	Habt ihr die Äpfel gepflückt?	T0309
Sì, le abbiamo raccolte noi!	Ja, wir haben es gemacht!	T0310
Non ti ho visto lavorare, ho visto lavorare solo lei	Ich habe dich nicht arbeiten sehen, nur sie	T0311
Ma l'albero ha chiesto a voi due, non solo a lei	Aber der Baum hat euch beide gefragt, nicht nur sie	T0312
La vecchia dà alla brava ragazza oro e diamanti e ricopre di cenere la ragazza pigra	Die alte Frau gibt dem fleißigen Mädchen Gold und Diamanten und dem faulen nur Asche	T0313
Poi le ragazze tornano a casa e chiamano la mamma	Dann kehren die Mädchen nach Hause zurück und rufen ihre Mutter	T0314
La vecchia vi ha dato qualcosa per il vostro lavoro?	Hat euch die alte Frau für die Arbeit belohnt?	T0315
Non ha dato qualcosa a noi, ma solo a me!	Sie hat nicht uns belohnt, nur mich!	T0316
La vecchia vi ha dato dell'oro come ricompensa?!	Hat die Alte euch Gold als Belohnung gegeben?!	T0317

Suoni	Laute	
IT	DE	stimulus
Come pronunciamo il suono [r]	Wie wir den Laut [r] aussprechen	
rot (rosso) – ros (rosso o marrone) / presa (fretta) / será (rame) / ran (rame)	rot – ros (rot oder braun) / presa (Hast) / será (geschlossen) / ran (Kupfer)	W0005, W0418, W0474, W0093, W0059, W0563, W0526, W0160, W0335, W0277, W0219, W0390
Jahr (anno) – morto/nord/smortsar (spegnere)	Jahr – morto/nord/smortsar (löschen)	W0021, W0497, W0100, W0062, W0421, W0570, W0533, W0164, W0281, W0339, W0349, W0223
Come pronunciamo il suono [s] prima di consonante	Wie wir den Laut [s] vor Konsonant aussprechen	
Stadl (fienile) / Stall (stalla) / Stark (forte) – stufa / stufo (stanco) / stua (soggiorno)	Stadl / Stall / Stark – stufa / stufo (müde) / stua (Wohnzimmer)	W0015, W0483, W0068, W0038, W0395, W0541, W0504, W0124, W0242, W0356, W0183, W0300
Spitz (cima) - specchio	Spitz - specchio (Spiegel)	W0013, W0539, W0502, W0036, W0394, W0066, W0122, W0181, W0240, W0354
Mist (letame) – Agosto / impastare / estate	Mist – Agosto (August) / impastare (mischen) / estate (Sommer)	W0026, W0498, W0074, W0043, W0401, W0547, W0510, W0368, W0196, W0313, W0255, W0137

Wespen (vespe) – responder (rispondere)	Wespen – responder (antworten)	W0024, W0400, W0366, W0194, W0073, W0042, W0135, W0311, W0253, W0546, W0509
Aspekt (aspetto) / ospedale (ospedale) / Vesper – ispettore	Aspekt / Vesper/ ospedale (Krankenhaus) - ispettore (Inspektor)	W0087, W0055, W0413, W0156, W0331, W0273, W0215, W0386, W0521, W0558
Suoni sibilanti interessanti	Interessante Sibilantenlaute	
cervello	cervello (Gehirn)	W0001, W0489
giocare / giallo / giovane	giocare (spielen) / giallo (gelb) / giovane (jung)	W0002, W0476, W0283, W0341, W0109, W0225, W0168
Säge (sega) / lessen (leggere) - sale	Säge / lesan (lesen) - sale (Salz)	W0080, W0003, W0473, W0450, W0106, W0224, W0282, W0340, W0165, W0553, W0516

Immagini	Bilder	
IT (displayed text)	DE (displayed text)	stimulus
I01.jpg (Salame)	I01.jpg (Salami)	I01 
I02.jpg (Tetto)	I02.jpg (Dach)	I02 
I03.jpg (Piatti)	I03.jpg (Geschirr)	I03 
I04.jpg (Lettera)	I04.jpg (Brief)	I04 
I05.jpg (Acqua)	I05.jpg (Wasser)	I05 

I06.jpg (Capelli)	I06.jpg (Haare)	I06	
I07.jpg (Mosche)	I07.jpg (Fliegen)	I07	

Appendix 2.4 Linguistic Landscape workflow

Data collection

Statement of purpose: the collection of any written sign in the public space (in any of the northern regions of Italy). Both monolingual and multilingual signs are fine, and all languages are welcome.

What type of data will be collected?

- Images (format: currently set to “image/*, pdf” which should include all image extensions, “image/*, pdf” which should include all image extensions and, additionally, pdf files that may contain comments or tags provided by the informants, especially by the students in the context of school projects)

What metadata needs to be collected for tagging the images?

- Location as precise as possible (geographic coordinates or street, community name, province) (coordinates will be added by researcher if not provided by the informant)
- Comment (optional); for example, which languages are included on the sign; who placed the sign; what is the sign communicating? Etc.

Where can data be collected?

- Public spaces only, only text, no people (checked by researchers).

Data collection is done via a contact form on the AlpiLinK page. The Wordpress contact form can be linked to any email address (currently to vinko@ateneo.univr.it) to which the entire form plus image is send (max. size is 10MB).

Data storage and representation

Anne Kruijt’s proposal is to leave the organized data storage with the Lingscape project and their servers. They have a workflow set-up for external projects, in which you can easily add your own data. Kruijt created an AlpiLinK project on the Lingscape project list area: <https://lingscape-app.uni.lu/pin/list/> (For a full tutorial on how to manage projects with them, see: https://lingscape-app.uni.lu/mobile/tutorial_eng.pdf)

Username: Anne_Kruijt. Project name: AlpiLinK. Lingscape app: password for current project.

Once data comes in through the contact form, the researcher has to check the data for the following things:

- Suitability: adherence to the Lingscape terms and conditions (see terms and conditions on the app, <https://lingscape-app.uni.lu/mobile/en/terms.html>), location is in the target area
- Coordinates of location (can be extracted from photo by looking at properties, or using Exiftool (<https://exiftool.org/>))
- Languages present in the photo
 - Only languages with a ISO-693 code are part of this list; this means that *Tirolerisch*, *zahrar* and *plodar* are grouped under bav (Bavarian).
- Tagging (everything that can be deduced from the image and comment; cf. <https://lingscape.uni.lu/taxonomies/>). Suggested tagging; directedness, discourse, and form.
 - Directedness:
Description: relates to the authorship for a sign, distinguishing between “official” (institutional or administrative) and “private” (basically everything else) authors; the category private subsumes such a disparate groups like companies or private persons. Plus, the scope of the distinction in linguistic landscapes literature varies between different researchers: **bottom-up** (private authors, including commercial communication), **top-down** (institutional and/or administrative authors).
 - Discourse:
Description: indicates different socio-pragmatic domains of language use also referred to as “discourses”, e.g., artistic expression, signs of protest, public regulation, or information boards: **artistic** (e.g., graffiti and street art), **commemorative** (e.g., memorial plaques or inscriptions), **commercial** (e.g., advertising and shop signage), **expressive** (e.g., statements in relation to societal, cultural, or private facts), **informatory** (e.g., information about public events, such as

holidays or elections), **infrastructural** (e.g., information about public infrastructure, such as street signs or instructions on trashcans), **political** (e.g., statements by political parties or private protest), **regulatory** (e.g., information about the regulation of behavior in the public, such as prohibition signs), **subcultural** (e.g., signs addressing a specific subculture, such as Hip Hop or skating), **technical** (e.g., signs indicating machinery or other technical infrastructure), **other**

- Form

Description: describes the appearance or “type” of sign: **display panel, graffiti, mural, neon sign, note, object (non-sign), plaque, poster, stand, sticker, street sign, window, writing, other**

If the numbers of images are relatively low, Kruijt suggests adding the photo’s one by one, as tagging needs to be done by hand anyway. However, they do have a mass upload workflow available, should it be necessary at any point (contact lingscape for details).

Once the data is present in the Lingscape project and made public, it is displayed on their map and freely available. Data hosted with Lingscape should be cited in the following way: <https://lingscape.uni.lu/license-citation/>

<Your name> (<year of first publication>): <Your project name>. In: Purschke, Christoph / Gilles, Peter (2016 ff.): Lingscape – Citizen science meets linguistic landscaping. Esch-sur-Alzette: University of Luxembourg. Published online at: <https://lingscape-app.uni.lu>

Data curation and preservation

Anne Kruijt suggests a **long-term** storage after completion of the project, as a single upload to a general repository (e.g. Zenodo). At the time of writing (July 2025) no repository has been created, i.e., photographs and their tagging are stored exclusively by Lingscape. However, Lingscape has an export project button (see below) – this results in a .zip folder with images and a structured .csv file containing the metadata in a structured format. Once the project is complete, the entire dataset could easily be exported and stored with a repository. The data would need to be structured only minimally, and a ReadMe file should be added.



Figure 18: Lingscape export function.

References

European Organization For Nuclear Research and OpenAIRE (2013). Zenodo. <https://www.zenodo.org/>

Purschke, Christoph & Peter Gilles (2016 ff.): Lingscape – Citizen science meets linguistic landscaping. Esch-sur-Alzette: University of Luxembourg. <https://lingscape.uni.lu>.

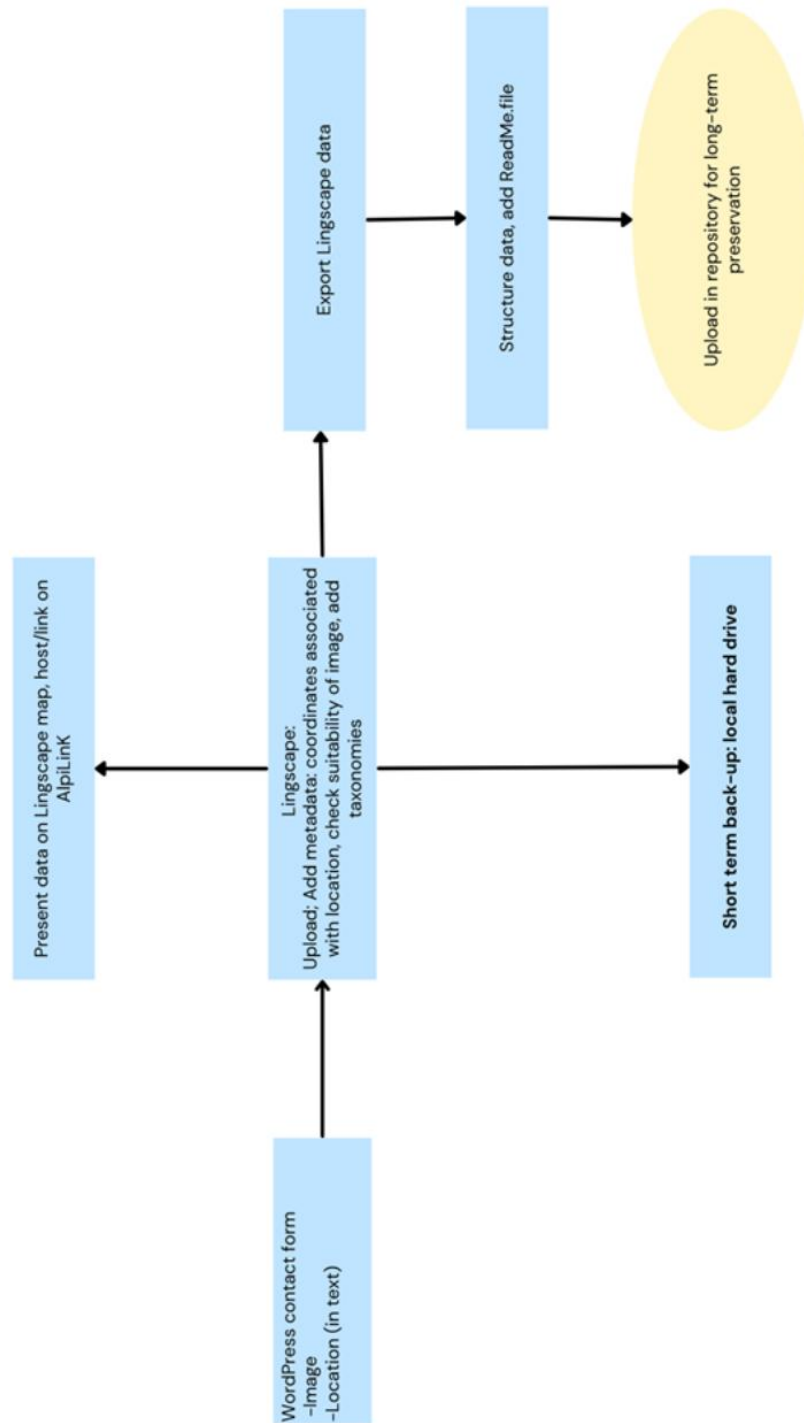


Figure 19: Linguistic Landscape workflow (proposal).

Appendix 3.1 Data Management Plan

The Data Management Plan has been authored by Anne Kruijt. It does **not** include the data of the Linguistic Landscape subproject.

1. Data Summary

The purpose of the data collection is to supply the data needed to obtain the objectives of the project, namely the investigation of linguistic phenomena in the non-standard language varieties spoken in the Alpine regions of Italy. The collected data will come in two types: audio data and written data. The audio data will be stored in the .flac format. The written data will be stored in a .csv file. The origin of the data is the participants of the project; participation in the project is on voluntary basis and is done offline without active assistance from the research team. The expected size of the data is currently estimated to 1.400 participants, resulting in 60.000 audio files (roughly 20 GB) and 13.250 written responses. The resulting data might be useful to other researchers working in the fields of syntax, dialectology, morphology, phonology, and sociolinguistics. The resulting data might also be of interest to local speech communities, dialect/minority language organizations and for educational purposes.

2. FAIR data

2.1 Making data findable, including provisions for metadata

The produced data are discoverable by means of their DOI. All versions of the dataset can be identified using the following DOI: <https://doi.org/10.5281/zenodo.8360169>.

The naming conventions of the audio files are as follows: The audio file name always mentions the stimulus ID (e.g. S01) followed by the abbreviation of the language variety (e.g., cim) and ending in the user ID (e.g., U0003). This means that audio file S01_cim_U0003 is a Cimbrian translation of stimulus S01 by speaker U0003. The first letter of the stimulus ID indicates the task design of the stimulus. Full details are available in the ReadMe file of the dataset.

The csv files are named with the version number attached for clarity.

Keywords provided are: Italian dialects, German minority languages, Cimbrian, Francoprovençal, Friulian, Mòcheno, Ladin, Lombard, Occitan, Piemontese, Sappadino, Saurano, Trentino, Tyrolean, Venetan, Walser German, Cross-linguistic comparison, German-Romance language contact, Audio, Timavese, Resian, Slovenian dialects.

Zenodo provides clear versioning numbers, which will be reflected in the naming of the .csv files. Each new upload has a distinct DOI allowing for clear referencing to specific versions of the dataset.

Provided metadata includes basic information about the participants, a readme file with the methodology and tasks clearly described, and the stimuli used to elicit the data.

2.2 Making data openly accessible

All produced data can and will be made openly available. The data is made available by deposition in a repository (Zenodo). The data can be accessed with any text processor and any audio processor.

Data and associated metadata are stored in the Zenodo (<https://about.zenodo.org/>) repository. Zenodo is a recognised and certified repository funded by CERN, OpenAIRE, and the European Union.

There are no restrictions on access to the data.

2.3 Making data interoperable

The data produced are in standard and commonly-used formats which have a large degree of interoperability, allowing for data exchange and re-use by any interested researcher.

The vocabulary used is detailed in the documentation associated with the dataset (readme file).

2.4 Increase data re-use (through clarifying licences)

The data will be licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0): <https://creativecommons.org/licenses/by-nc-sa/4.0/>

The data is immediately available for re-use for any party, including third parties. Technically the data is remains re-usable to infinity. Data quality is assured by spot checks in the export and upload process.

3. Allocation of resources

The costs for the data management and archiving are included in the project funding and included in the work package of the RTD-A position at the University of Verona.

4. Data security

Data security is ensured by the data security protocols at the University of Verona. During the project, data is preserved on the server of the external survey provider (Phonic) for a maximum of 2 months. Within this period, the data will be transferred to the University server where they will reside for the remainder of the project. After transferring the data, it will be deleted from the external survey provider's server.

Long term storage on Zenodo is ensured by Zenodo for the period of 20 years at least. They further state that "in the case of closure of the repository, best efforts will be made to integrate all content into suitable alternative institutional and/or subject based repositories".

5. Ethical aspects

Informed consent is acquired from all participants via the website and the planned data sharing and long term preservation are included. Without providing consent of the use of their data, participants cannot enter the online survey. The privacy agreement is available in Italian and German here:

https://alpilink.it/wp-content/uploads/2023/05/TrattamentoDati_Datenverarbeitung_it-de.pdf

The collected data include the following personal data: age, gender, dialect variety, location of birth, and audio recordings of the participant's voice. Participants might be identified by people that know them from their voice in combination with their sociolinguistic information. The potential identification represents no or minimal risk to participants, as no sensitive information is included in the questionnaire and participants can control which information they provide.

6. History of changes

Version	Publication data	Change	Responsible person
1.0	10 October 2023	Initial version	Anne Kruijt
1.1	28 May 2024	Audio format changed from .wav to .flac. This format is lossless but compressed resulting in a lower processing load and easier storage of the data. The predicted number of participants and audio files has been adjusted slightly upwards in line with current predictions. Keywords added: Timavese, Slovenian dialects, Resian.	Anne Kruijt
1.2	26 May 2025	Variables have been completely revised, see Readme file and questionnaire_v1.2.0.csv.	Stefan Rabanus

Appendix 3.3 Data export: from Phonic to AlpiLinK repository

Programs needed: Excel, Spyder (to run the python script)

Step 1: Download .csv data from Phonic.

Create a new Folder in your Export Phonic folder and name it export_yyyymmdd

Download the written data in a .csv-format from Phonic, by logging into the Phonic account.

User: anne.kruijt@univr.it. Password.

Go to Survey > AlpiLinK Survey > Options > Export Survey Data, select .csv, check boxes for “Use custom data labels as column headers” and “Export only bookmarked responses” and click export.

This downloads a file with the name phonic-63ed07834ec225fa096a4b21.csv (the number is the ID of the entire survey) – save file in the folder you created.

Step 2: Download Media data from Phonic.

Open Spyder and open api_export_by_session_phonic.py file located in the shared AlpiLinK folder in “Internal documents...”, “Documents_WP3...” (in case ask for the file via mail to vinko@ateneo.univr.it).

Run the script: this should create a new folder named “sessions” in the same folder as the script. It contains all audio data separated by participant id. Depending on how much data there is, this might take a while.

The script downloads data from all participants, not only those that are bookmarked (there is no way to download only bookmarked participants, I asked Phonic, but it is not available).

Once downloaded, move the folder “session” to the export folder you created in step 1.

Step 3: Clean up the .csv file so that it is ready for processing.

Open the csv. file in a new worksheet, by going to Data> Import from txt/csv, click “comma-delimited”, check the previous and if correct load into the excel sheet. Rename: cleaned_export_yyyymmdd.csv. Follow the following steps:

- Delete unnecessary columns: Location, User-Agent, URL
- Check if survey language is correct. For example, a Tyrolean speaker might use the it url to get into the survey, but once Tiroler Dialekt is selected, the survey language changes to de. Ladin varieties select which survey they do. The column registers only the survey language through which a participant first entered the survey.
- Merge Q6 to Q40 by adding a new column 'Comune' before Q6 and then using formula CONCAT: (in AutoSum / Formulas /Concat / Text 1 take whole range from Q6 to Q40) !! Then take the names in new 'Comune' Copy and Paste Special > Value. Now you can delete Q6 - Q40.
- Q41/Q42 - Merge Q41a/b and Q42a/b in one column (delete the remaining columns)- rename 'location_presence' - find and replace si/ja with Y and no/nein with N.
- Q43 - rename 'Age'
- Q44a/b/c, Q45a/b/c Merge in new column (delete remaining) - rename 'gender', find and replace Femmina/weiblich with F, and Maschio/männlich with M.
- Merge Q47 with Q46: then rename (find replace) categories; Ja/Sì = 5, Abbastanza bene/Spesso/Ganz gut/recht häufig = 4, Così così/Ogni tanto/Mittelgut/manchmal = 3, Non molto bene/Raramente/Nicht sehr gut/selten = 2, No/Nein = 1.
- Q 52 - Q90 - Delete unnecessary data columns: only the one called Qxx [S01] Response ID needs to be kept - delete columns called Back-up text, Sentiment Overall, Sentiment Positive, Sentiment Negative, Sentiment Neutral, Sentiment Mixed, and Audio.

- Delete Q 93 and Q 96
- Q 98 - Q103 - Delete unnecessary data columns: only the one called Qxx [S01] Response ID needs to be kept - delete columns called Back-up text, Sentiment Overall, Sentiment Positive, Sentiment Negative, Sentiment Neutral, Sentiment Mixed, and Audio.
- Keep Q104 [N19], will only be filled in for German minority languages, Slovenian, and Ladin.
- Q 105 - Q 140 (with [N01], etc) – Merge those with the same code between the brackets e.g. Q105 and Q106 both have [N01] – Use formula: =TEXTJOIN(" "; TRUE;xxxdatarangexxx)
Depending on the settings of your Excel, instead of semicolon “;” you might need to use comma’s “,”.
The same goes for all other formulas in this document.
- Reorder the N section to go in numerical order from N01 to N19: they are shuffled in the questionnaire to mix the age groups per category. Fix now to prevent problems later.
- Q141 - Q153 - Delete unnecessary data columns: only the one called Qxx [G01] Response ID needs to be kept - delete columns called Back-up text, Sentiment Overall, Sentiment Positive, Sentiment Negative, Sentiment Neutral, Sentiment Mixed, and Audio.
- Keep final question (open text, how did you learn about the questionnaire).

Step 4: Update metadata table

Open excel sheet User_metadata.xlsx (in “Internal documents...”, “Documents_WP3_...”) and add the new data to the existing file.

Fill in the columns: date, survey_language, language_variety, location_province, location_municipality, location_presence, age, gender, language_compentence, language_frequency, language_family, language_friends, other_languages, written_use, found_questionnaire.

Drag down the VLOOKUP formula in the ISTAT_name_IT, the GID, coordinates, and map_label columns.

language_variety_iso works on a formula, so either it updates automatically or draw down the existing formula (formula is:

```
=IF(E684="Tiroler  
Dialekt";"tir";IF(E684="Cimbro/Zimbrisch";"cim";IF(E684="Trentino";"tre";IF(E684="Lombardo";"lmo";I  
F(E684="Francoprovensal";"frp";IF(E684="Furlan/Friulano";"fur";IF(ISNUMBER(SEARCH("Ladin";  
E684)));"lld";IF(E684="Mòcheno/Bernstolerisch";"mhn";IF(E684="Occitan/Occitano";"oci";IF(E684="Piem  
ontese";"pms";IF(E684="Plodarisch/Sappadino";"plo";IF(E684="Zahrar  
sproche/Saurano";"zah";IF(E684="Tischlbong  
sproche/Timavese";"tis";IF(E684="Veneto";"vec";IF(E684="Walser";"wae";""))))))))))))))
```

(this will break if any of the language labels are changed)

Audio columns are left empty for now.

Step 5: create audio_conversion file

Copy file audio_conversion_empty to your new export folder and rename audio_conversion_yyyymmdd

Copy into this file User_ID and linguistic variety (tir, vec, etc.) from the User_metadata file. (linguistic variety copy as value, otherwise it doesn't work).

Now copy the audio labels from the cleaned_export_yyyymmdd.csv file into the columns S01-M07. DO NOT EDIT ANYTHING IN THE FOLLOWING COLUMNS.

This would remove the present formulas:

Formula_oldname: =IF(ISTEXT(C2); C2 & "-0.wav"; "") (C2 is variable depending on the row it is in)

Formula_newname: =IF(ISTEXT(C2); "S01_"&B2&"_"&A2; "") (S01 must be changed to reflect the actual stimuli. B2 and A2 are constant and should not change).

Select from the row below S01_oldname to M07_newname and drag down for all your data rows– watch the magic happen.

Check if empty input is properly read as empty or if it produces things like: -0.wav etc. If under the XXX_oldname columns you get -0.wav, this means that the original cells are not being read as empty. DO NOT CHANGE ANYTHING IN THE OLDNAME COLUMNS. The problem lies in the part where you input the data, so this is the place to fix it, even if visually you will not find a problem here. Follow these steps:

1. Select the data input area (S01, S02 etc). Only the area where you just pasted the audio labels. Do not select the oldname and newname columns.
2. Open Find & Replace with Ctrl+F.
3. Leave the Find section empty and click the 'Match entire cell contents'.
4. Click "Find all". This should find empty cells in the selected area. Use Ctrl+A to select all the found cells.
5. Click out of the Find dialogue box and in the Formula bar (where you type any text you want to put into a cell) click backspace or delete.
6. Hold Ctrl and hit Enter to empty all selected cells.

Copy and paste all the new audio names into the audio columns in the User_metadata.xlsx sheet. (remember we left this empty in step 4.) Please note: in the questionnaire S15-16 and T07-T06 are not in numerical order, however, the renaming sheet should also reorder them correctly. So all audio files should be ready to copy straight into the User_metadata sheet.

Step 6: Rename your audio files.

Create folder named 'audio' to your new export folder. In this folder copy all files of the bookmarked participants from the sessions folder. Check in the Phonic interface which participants are not bookmarked and delete these folders from sessions or check the column in the clean_copy_yyyymmdd file labelled Session ID.

Then copy all remaining audio files (without folder; <https://www.computroon.co.uk/2023/01/16/how-to-move-files-from-multiple-folders-to-one-folder/#:~:text=Select%20All%20Files%20From%20Search%20Result&text=Depending%20on%20the%20amount%20of,into%20the%20new%20destination%20folder.>) to the audio folder. Any files that are not present in your excel sheet will create an error message in the script.

Make a copy of the audio folder as a back-up.

Copy rename_phonicaudio.py to your created export folder and rename rename_phonicaudio_yyyymmdd.py.

Open Spyder and open rename_phonicaudio_yyyymmdd.py. (Note: This python script runs on a specific spyder-env (modified in Spyder > Preferences > Python interpreter), I needed to download miniconda to download the required packages to make the API work. So, if there is any package missing in future scripts, open Anaconda Prompt, go to the environment (conda activate spyder-env), and download new packages there (conda install "whateverpackage").

Edit the following lines:

- Line 5: Correct the path to excel sheet
names_df = pd.read_excel(r"audio_conversion_yyyymmdd.xlsx")
- Line 120: Correct path to audio folder
audio_folder = r'LINK\audio'

Run the program.

All the audio files should now be renamed in the format indicated in the _newname columns of the data_conversion file.

Convert all to .flac format. (I personally use 'speechnotes_wav_to_flac.bat'. Copy into the folder with the audio and run the program. It will automatically convert all files and put them in a new folder.)

Step 7: Update the corpus on Zenodo

Go to AlpiLinK>Corpus, duplicate the latest folder, and rename: AlpiLinK_corpus_vx.x.x (for guidelines on versioning see the ReadMe file of the corpus, Section 6).

Add the new audio data per language variety to the specific language variety folders. If no data on a variety is available, create new folder with the 3-letter language abbreviation.

Go to the metadata folder and update and rename with the correct version: users_results_vx.x.x.

If needed update the questionnaire_vx.x.x file. If not, simply rename with the appropriate version number.

(Questionnaire should only be changed if there are changes to the questionnaire, but if this is the case, a lot of things need to be fixed in previous steps too!!).

Update the ReadMe file with at a minimum the following things:

- Description: time period (final sentence).
- General: Date of creation version; last updated; number of audio files and tokens.
- How to cite; update reference.
- Abbreviations: add new language varieties if applicable.
- Data structure: add new folders if applicable, and sentence: “There are audio recordings in 11 language varieties.”
- Metadata folder: Update version numbers. Add changes if any have been made.
- Dissemination: Update.
- Updates: Provide list of changes made in new version; including the speaker IDs of new speakers added. Also list any deletions that might have been made in the quality control of the corpus.

Create a new map with the data locations using the REDE platform (<https://www.regionalsprache.de/SprachGIS/Map.aspx>). This can be easily done by using the functions Kartenebenen verwalten> Importiere Daten > Texteingabe and copying and pasting columns F (language_variety_iso) to J (GID) of users_results_vx.x.x.csv. Click ‘Aktualisieren’ and indicate language_variety_iso as ‘Wert’ and GID as ‘Rede GID’. The rest can be set to ‘Ignorieren’. Go to Style bearbeiten and get rid of the colored municipality areas. Then go to Visualisieren and plot. I usually plot with colors per variety first, so that any unexpected locations for a specific variety immediately stand out. If there are no peculiarities, set all colors to white and plot dots on the AlpiLinK area map. Export as a high-quality jpg image. Add a legend to the map with the text: The white dots indicate the locations for which there is data available in the corpus. Attribution: Erstellt mit www.regionalsprache.de.

Zip all audio folders and metadata folder and place in Zenodo upload folder. Add a copy of the ReadMe file and the map in the upload folder.

Log-in on the vinko@ateneo.univr.it Zenodo profile with the password. Create a new version of the existing corpus. Get a new DOI for the upload.

Upload all files in the Zenodo upload folder to Zenodo and correct the following metadata:

- Change the title to the correct version.
- Publication date.
- Update description.
- Update citation.
- (if applicable: update keywords).
- Update version.

If you have updated parts but are not ready to publish, click ‘save draft’. Once ready and checked, click ‘Publish’. Once published, the metadata can still be corrected but no files can be changed.

Publish new AlpiLinK corpus and notify all team members.

Appendix 5.1 Questionnaire

ID	Target sentence IT	Target sentence DE	Task	Variables	Unit(s)
G01	-	Gefauch(e)/ Vor dem Haus FAUCHEN zwei Katzen. Das erschrockene Kind fragt: Was ist denn das für ein ...?	Sentence completion	(deverbal iterative/ collective noun) (simpleonset: obstruent, fricative, labial, voiceless)	UniBZ
G02	-	Getropf(e)/ Das warme Eis TROPFT auf den Boden, aber das Kind merkt es nicht. Die Mutter schimpft: Was ist denn das für ein ...?	Sentence completion	(deverbal iterative/ collective noun) (complexonset: obstruent, stop, alveolar, voiceless)	UniBZ
G03	-	Gedrahn(e)/ Die Frau DREHT am Knopf des Radios, um ihren Lieblingssender zu finden. Ihr Mann fragt genervt: Was ist denn das für ein ...?	Sentence completion	(deverbal iterative/ collective noun) (complexonset: obstruent, stop, alveolar, voiced)	UniBZ
G04	-	Geroschl(e)/Die Kinder RASCHELN mit dem Geschenkpapier. Genervt fragt die Mutter: Was ist denn das für ein ...?	Sentence completion	(deverbal iterative/ collective noun) (simpleonset: sonorant, liquid, rhotic, voiced)	UniBZ
G05	-	Gesumme/Das Kind SUMMT gelangweilt vor sich hin. Die Mutter sagt: Was ist denn das für ein ... ?	Sentence completion	(deverbal iterative/ collective noun) (simpleonset: obstruent, fricative, alveolar, voiceless)	UniBZ
G06	-	Gewechsl(e)/ Maria WECHSELT dauernd die Schule. Hans findet das nicht gut, er sagt: Was ist denn das für ein ...?	Sentence completion	(deverbal iterative/ collective noun) (simpleonset: obstruent, fricative, labial, voiced)	UniBZ
G07	-	Kschittl(e)/ Der Junge SCHÜTTELT die Bierflasche, bis das Bier herausläuft. Sein Freund sagt: Was ist denn das für ein ...?	Sentence completion	(deverbal iterative/ collective noun) (simpleonset: obstruent, fricative, postalveolar, voiceless)	UniBZ
G08	-	Khock(e)/ Der Mann HACKT den ganzen Tag Holz und macht dabei viel Lärm. Seinen Nachbarn ärgert das und er sagt: Was ist denn das für ein ...?	Sentence completion	(deverbal iterative/ collective noun) (simpleonset: obstruent, fricative, glottal, voiceless)	UniBZ
G09	-	Genag(e)/ Der Hund NAGT schon seit einer Stunde an einem Knochen und macht dabei knirschende Geräusche. Sein Herrchen sagt genervt: was denn das für ein ...?	Sentence completion	(deverbal iterative/ collective noun) (simpleonset: sonorant, nasal, alveolar, voiced)	UniBZ
G10	-	Gejodl(e)/ Das Mädchen lernt gerade zu JODELN. Ihrem Freund gefällt das nicht und er sagt: Was ist denn das für ein ...?	Sentence completion	(deverbal iterative/ collective noun) (simpleonset: sonorant, approximant, palatal, voiced)	UniBZ
I01	L'uomo taglia (su) il salame.	Der Mann schneidet die Salami auf.	Image description	(3sg.m.preverbal) (clitic doubling) (phrasal verb)	UniVR, UniTN
I02	Gli operai salgono (su) sul tetto.	Die Arbeiter steigen auf das Dach (auf).	Image description	(3pl.m.preverbal) (clitic doubling) (phrasal verb)	UniVR, UniTN
I03	Le ragazze lavano (su) (i piatti).	Die Mädchen spülen (das Geschirr) ab.	Image description	(3pl.f.preverbal) (clitic doubling) (phrasal verb)	UniVR, UniTN

I04	L'anziana spedisce (/manda via) le lettere.	Die ältere Frau verschickt Briefe.	Image description	(3sg.f.preverbal) (clitic doubling) (phrasal verb)	UniVR, UniTN
I05	Dalla finestra viene dentro l'acqua.	Wasser kommt durch das Fenster herein.	Image description	(phrasal verb)	UniVR, UniTN
I06	Ho visto Giovanni lavarsi la testa.	Ich habe Johannes (beim Haarewaschen) gesehen (wie er sich gerade die Haare wäscht).	Image description	(1sg.preverbal) (3sg.m.preverbal) (clitic doubling) (3sg.reflexive) (progressive) (m.jamb.obj. postverbal) (subordinate clause)	UniTO
I07	Il ragazzo manda via le mosche.	Der Junge schickt die Fliegen weg.	Image description	(3sg.m.preverbal) (clitic doubling) (phrasal verb)	UniVR, UniTN
N01	Emily (26) mentre sta tornando a casa dalla biblioteca, incontra alla fermata dell'autobus la sua amica Francesca (28), sudata e con la borsa del tennis. Emily le dice: "Hey _____ (Francesca), hai appena finito l'allenamento? Come è andato?" Quale nome abbreviato sceglierebbe Emily, fra i seguenti? [è possibile selezionare più di una risposta] Fra, France, Franci, Cesca, Non saprei, Uno diverso (specifica).	-	Forced choice (no audio)	(name truncation)	UniBZ
N02	Luca (20) e Matteo (19) erano compagni di classe al liceo. Un anno dopo la maturità hanno organizzato con la loro classe una pizza per rivedersi. Luca, che lavora nell'azienda di famiglia, domanda a Matteo: Luca: Ehi, _____ (Matteo), come va con l'università? Matteo: Finora sta andando bene. Il carico di studio è tanto, ma mi piace. Quale nome abbreviato sceglierebbe Luca, fra i seguenti? [è possibile selezionare più di una risposta] Matti, Matte, Teo, Non saprei, Uno diverso (specifica).	-	Forced choice (no audio)	(name truncation)	UniBZ
N03	Giovanni (18) e Stefania (21) sono fratelli e stanno cenando con la loro famiglia. Stefania chiede a Giovanni se vuole che gli passi l'acqua: "_____ (Giovanni), vuoi un altro po' d'acqua?" Quale nome abbreviato sceglierebbe Stefania, fra i seguenti? [è possibile selezionare più di una risposta] Giò, Giova, Giovi, Vanni, Non Saprei, Uno diverso (specifica).	-	Forced choice (no audio)	(name truncation)	UniBZ
N04	Francesca (45) e Elisa (39) sono cugine e stanno festeggiando il compleanno del loro zio. Dopo aver mangiato una fetta di torta, Elisa chiede a Francesca se ne vuole un'altra fetta: "_____ (Francesca), vuoi un'altra fetta di torta? Ne è rimasta una piccola." Quale nome abbreviato	-	Forced choice (no audio)	(name truncation)	UniBZ

	sceglierebbe Elisa, fra i seguenti? [è possibile selezionare più di una risposta] Fra, France, Franci, Cesca, Non saprei, Uno diverso (specifica).				
N05	Matteo (34) e Giacomo (34) sono amici e si aggiornano di cosa è successo nell'ultimo periodo. Giacomo dice: "Sai _____ (Matteo), ho iniziato quel corso di teatro che ci aveva consigliato Sophia. È davvero fantastico!" Quale nome abbreviato sceglierebbe Giacomo, fra i seguenti? [è possibile selezionare più di una risposta] Matti, Matte, Teo, Non saprei, Uno diverso (specifica).	-	Forced choice (no audio)	(name truncation)	UniBZ
N06	Paolo (44) e Giovanni (50) sono amici e si sono dati appuntamento per mangiare la pizza in una pizzeria locale. Paolo chiede a Giovanni se gli piace la sua pizza, dicendo: "_____ (Giovanni), ti piace la tua pizza con le verdure?" Quale nome abbreviato sceglierebbe Paolo, fra i seguenti? [è possibile selezionare più di una risposta] Giò, Giova, Giovi, Vanni, Non Saprei, Uno diverso (specifica).	-	Forced choice (no audio)	(name truncation)	UniBZ
N07	Anna (63 anni) sta facendo una passeggiata in centro città un sabato pomeriggio e incontra la sua amica d'infanzia Francesca (63 anni). Anna saluta l'amica dicendo: "Ciao _____ (Francesca), è tantissimo che non ci vediamo! Come stai?" Quale nome abbreviato sceglierebbe Anna, fra i seguenti? [è possibile selezionare più di una risposta] Fra, France, Franci, Cesca, Non saprei, Uno diverso (specifica).	-	Forced choice (no audio)	(name truncation)	UniBZ
N08	Alessandro (68) e Matteo (67) fanno parte dello stesso gruppo di amici. Un giorno Alessandro chiama Matteo e gli dice: "Alla fine, _____ (Matteo), ieri sono andato a provare quel nuovo ristorante in via Cavour e devo ammettere che è davvero buono, grazie del consiglio!" Quale nome abbreviato sceglierebbe Alessandro, fra i seguenti? [è possibile selezionare più di una risposta] Matti, Matte, Teo, Non saprei, Uno diverso (specifica).	-	Forced choice (no audio)	(name truncation)	UniBZ
N09	Claudio (63) e Giovanni (65) sono appassionati di calcio e tifosi della Juventus. Oggi sono al bar del quartiere a guardare la partita Juventus-FC Bayern München. Dopo aver bevuto una birra, Claudio chiede a Giovanni	-	Forced choice (no audio)	(name truncation)	UniBZ

	se vuole prenderne un'altra: "____ (Giovanni), vogliamo fare un altro giro?" Quale nome abbreviato sceglierebbe Claudio, fra i seguenti? [è possibile selezionare più di una risposta] Giò, Giova, Giovi, Vanni, Non saprei, Uno diverso (specifica).				
N10	-	Emily (26) trifft auf dem Heimweg von der Bibliothek an der Bushaltestelle ihre Freundin Theresa (28), verschwitz und mit einer Tennistasche. Emily sagt zu ihr: "Hey (Theresa), warst du gerade beim Sport? Wie ist es gelaufen?" Welchen Kurznamen würde Emily am ehesten verwenden? [es können auch mehrere ausgewählt werden] Tess, There, Theri, Resi, Ich weiß nicht, Einen anderen (bitte angeben).	Forced choice (no audio)	(name truncation)	UniBZ
N11	-	Lukas (20) und Isabella (19) waren Klassenkameraden im Gymnasium. Ein Jahr nach dem Abitur organisieren sie mit ihrer Klasse ein Pizzeessen, um sich wiederzusehen. Lukas, der im Familienunternehmen arbeitet, fragt Isabella: "Hey, ____ (Isabella), wie läuft es mit der Uni?" Isabella: "Bisher läuft es gut. Die Studienbelastung ist hoch, aber mir gefällt es." Welchen Kurznamen würde Lukas am ehesten verwenden? [es können auch mehrere ausgewählt werden] Isa, Isi, Bella, Ich weiß nicht Einen anderen (bitte angeben).	Forced choice (no audio)	(name truncation)	UniBZ
N12	-	Matthias (18) und Stefanie (21) sind Geschwister und essen mit ihrer Familie zu Abend. Stefanie fragt Matthias, ob sie ihm das Wasser reichen soll: "____ (Matthias), möchtest du noch etwas Wasser?" Welchen Kurznamen würde Stefanie am ehesten verwenden? [es können auch mehrere ausgewählt werden] Mats, Matthi, Hias, Ich weiß nicht, Einen anderen (bitte angeben).	Forced choice (no audio)	(name truncation)	UniBZ
N13	-	Theresa (45) und Lisa (39) sind Cousins und feiern den Geburtstag ihres Onkels. Nachdem sie ein Stück Kuchen gegessen haben, fragt Lisa Theresa, ob sie noch ein Stück möchte: "____ (Theresa), möchtest du noch ein Stück Kuchen? Es ist noch ein kleines Stück übrig." Welchen Kurznamen würde Lisa am ehesten verwenden? [es können auch mehrere ausgewählt werden] Tess, There, Theri, Resi, Ich weiß nicht, Einen anderen (bitte angeben).	Forced choice (no audio)	(name truncation)	UniBZ

N14	-	Isabella (34) und Martin (34) sind befreundet und erzählen sich, was in letzter Zeit passiert ist. Martin sagt: "Weißt du _____ (Isabella), ich habe mit dem Theaterkurs begonnen, den Sophia uns empfohlen hatte. Der ist echt gut!" Welchen Kurznamen würde Martin am ehesten verwenden? [es können auch mehrere ausgewählt werden] Isa, Isi, Bella, Ich weiß nicht, Einen anderen (bitte angeben).	Forced choice (no audio)	(name truncation)	UniBZ
N15	-	Florian (44) und Matthias (50) sind Freunde und haben sich zum Abendessen in einer Pizzeria verabredet. Florian fragt Matthias, ob ihm seine Pizza schmeckt. Florian fragt: "_____ (Matthias), schmeckt dir deine Gemüsepizza?" Welchen Kurznamen würde Florian am ehesten verwenden? [es können auch mehrere ausgewählt werden] Mats, Matthi, Hias, Ich weiß nicht, Einen anderen (bitte angeben).	Forced choice (no audio)	(name truncation)	UniBZ
N16	-	Anna (63) geht an einem Samstagnachmittag in der Innenstadt spazieren und trifft ihre Jugendfreundin, Theresa (63), die einkaufen geht. Miriam begrüßt ihre Freundin mit den Worten: "Hallo _____ (Theresa), lange nicht mehr gesehen! Wie geht es dir?" Welchen Kurznamen würde Anna am ehesten verwenden? [es können auch mehrere ausgewählt werden] Tess, There, Theri, Resi, Ich weiß nicht, Einen anderen (bitte angeben).	Forced choice (no audio)	(name truncation)	UniBZ
N17	-	Alexander (68) und Isabella (67) gehören zum selben Freundeskreis. Eines Tages ruft Alexander Isabella an und sagt: "_____ (Isabella), gestern habe ich endlich das neue Restaurant in der Cavourstraße ausprobiert und ich muss zugeben, es ist wirklich gut, danke für den Tipp!" Welchen Kurznamen würde Alexander am ehesten verwenden? [es können auch mehrere ausgewählt werden] Isa, Isi, Bella, Ich weiß nicht, Einen anderen (bitte angeben).	Forced choice (no audio)	(name truncation)	UniBZ
N18		Klaus (63) und Matthias (65) sind Fußballfans und sehen sich in der Dorfbar das Spiel Juventus-FC Bayern München an. Nachdem sie ein Bier getrunken haben, fragt Klaus Matthias, ob er noch ein Bier trinken möchte: "_____ (Matthias) wollen wir noch eine Runde trinken?" Welchen Kurznamen würde Klaus am ehesten	Forced choice (no audio)	(name truncation)	UniBZ

		verwenden? [es können auch mehrere ausgewählt werden] Mats, Matthi, Hias, Ich weiß nicht, Einen anderen (bitte angeben).			
N19	Una domanda sulla abbreviazione dei nomi propri. Fra amici e in famiglia i nomi di persone vengono spesso abbreviati. Così una ragazza che si chiama MADDALENA potrebbe essere chiamata MADDI. Per favore indicate nomi che sono tipici per la vostra varietà. Per favore indicate per questi nomi anche le loro abbreviazioni, se esistono.	Eine Frage zu Kurznamen. Unter Freunden oder in der Familie werden die Namen von Personen oft abgekürzt. So wird dann vielleicht jemand, der ANTON heißt, TONI genannt. Bitte nennen Sie Namen, die typisch sind für die Varietät, die Sie sprechen. Bitte nennen Sie auch Abkürzungen für diese Namen, falls es welche gibt.	Open written response (no audio)	(name truncation)	UniBZ
S01	Mario ieri ha ballato tanto con Sara.	Martin hat gestern lange mit Sara getanzt.	Translation	(3sg.m.preverbal) (clitic doubling) (m.trochee.sbj.preverbal) (f.trochee.pp.postverbal)	UniVR
S02	Zappa questa aiuola subito, l'altra falla dopo.	Grab dieses Beet jetzt gleich um, und dann das andere.	Translation	(imperative.sg) (proximal and distal demonstratives)	UniTO, UniVDA
S03	Il pane l'hai già mangiato o lo devi ancora mangiare?	Hast du das Brot schon gegessen oder wirst du es noch essen	Translation	(2sg.postverbal) (3sg.acc.m/n) (infinitive/participle opposition) (left-dislocation) (interrogative clause)	UniTO, UniVDA
S04	La frutta l'hai già portata o posso portarla io?	Hast du schon Obst mitgebracht, oder soll ich es mitbringen?	Translation	(2sg.postverbal) (3sg.acc.f/m) (infinitive/participle opposition) (left-dislocation) (interrogative clause)	UniTO, UniVDA
S05	Questa casa qua è più fredda di quella là.	Das Haus hier ist kälter als das andere da.	Translation	(3sg.f/n.preverbal) (clitic doubling) (proximal and distal demonstratives) (comparative)	UniTO, UniVDA
S06	Quello bianco è il tuo cane.	Der weiße da ist dein Hund.	Translation	(3sg.m.preverbal) (clitic doubling) (possessive.2sg) (proximal and distal demonstratives)	UniTN
S07	Questo è il mio cane, non quello.	Das hier ist mein Hund, nicht der da drüben.	Translation	(3sg.m.preverbal) (clitic doubling) (possessive.1sg) (proximal and distal demonstratives)	UniTN
S08	La casa di Sara è questa.	Saras Haus ist das da.	Translation	(3sg.f/n.preverbal) (clitic doubling) (f.trochee.pp/gen.preverbal/postverbal) (possessor.f) (proximal and distal demonstratives)	UniTN
S09	Bisogna comprare del pane, dell'insalata, dei fagioli e delle ciliegie.	-	Translation	(impersonal verb "bisogna") (partitivity) (subordinate clause)	UniTO, UniVDA
S10	Avete mica visto Mario?	Habt ihr Martin gesehen?	Translation	(2pl.postverbal) (m.trochee.obj.postverbal) (presupposition)	UniTO, UniVDA

				(interrogative clause)	
S11	Mi chiedo se Mario sia più furbo di me.	Ich frage mich, ob Martin schlauer ist als ich.	Translation	(1sg.preverbal) (3sg.m.preverbal) (clitic doubling) (1sg.reflexive) (m.trochee.sbj.preverbal) (comparative) (subordinate clause)	UniTO, UniVDA
S12	Hai visto come Giovanni ha costruito la sua casa?	Hast du gesehen, wie Johannes sein Haus gebaut hat?	Translation	(2sg.postverbal) (3sg.m.preverbal) (clitic doubling) (m.jamb.sbj.preverbal) (possessive.3sg.m) (subordinate clause) (interrogative clause)	UniTO
S13	Tu dormi più di lui.	Du schläfst mehr als er.	Translation	(2sg.preverbal) (clitic coupling) (comparative)	UniTN
S14	Sara dice che mangiamo troppo.	Sara sagt, dass wir zu viel essen	Translation	(3sg.f.preverbal) (1pl.preverbal) (clitic doubling) (f.trochee.sbj.preverbal) (subordinate clause)	UniTN
S15	Da bambino giocavo sempre con le mie sorelline.	-	Translation	(1sg.preverbal) (evaluative morphology/diminutives) (possessive.1sg)	UniTN
S16	-	Als Kind habe ich immer mit meinen Hündchen gespielt.	Translation	(1sg.postverbal) (evaluative morphology/diminutives) (possessive.1sg)	UniTN
S17	Ognuno va per conto suo: noi con la nostra auto, voi con la vostra.	Jeder fährt alleine: wir mit unserem Auto, ihr mit eurem.	Translation	(3sg.preverbal) (1pl.preverbal) (2pl.preverbal) (clitic doubling) (possessive.3sg) (possessive.1pl) (possessive.2pl)	UniTN
S18	D'estate i miei vicini tornano nella loro vecchia casa.	Meine Nachbarn fahren im Sommer immer zu ihrem alten Haus.	Translation	(3pl.m.preverbal) (clitic doubling) (possessive.3pl) (possessive.1sg)	UniTN
S19	In piazza non ci sono alberi alti.	Auf dem Platz stehen keine großen Bäume	Translation	(negative quantification)	UniTN
S20	Alcuni bambini piccoli parlano ancora in dialetto.	Einige kleine Kinder sprechen noch Dialekt.	Translation	(3pl.m.preverbal) (clitic doubling) (indefinite quantification)	UniTN
S21	Ho visto pochi vecchi amici che stavano festeggiando.	Ich habe viele alte Freunde gesehen, die am Feiern waren.	Translation	(1sg.preverbal) (progressive) (indefinite quantification)	UniTN
S22	Ricordo ancora la tua vecchia gattina.	Ich erinnere mich noch an dein Kätzchen von früher.	Translation	(1sg.preverbal) (1sg.reflexive) (evaluative morphology/diminutives) (possessive.2sg)	UniTN
S23	C'era Martina in piazza, l'hai vista? Cosa faceva?	Martina war draußen unterwegs, hast du sie gesehen? Was hat sie gemacht?	Translation	(3sg.f.preverbal) (2sg.postverbal) (3sg.f.postverbal) (clitic doubling) (3sg.acc.f) (f.jamb.sbj.preverbal/post-	UniTO, UniVDA

				verbal) (presentative) (interrogative clause)	
S24	Giovanni si sta lavando i capelli.	Johannes wäscht sich gerade die Haare.	Translation	(3sg.m.preverbal) (clitic doubling) (3sg.reflexive) (progressive) (m. jamb.sbj.preverbal)	UniTO
S25	Ho dato una bella lavata alla macchina.	-	Translation	(1sg.preverbal) (deverbal action noun [transitive])	UniTO
S26	La passeggiata che ho fatto fare ai bambini li ha stancati	-	Translation	(1sg.preverbal) (3pl.acc.m) (causative) (deverbal action noun [intransitive])	UniTO
S27	Mi siedo qui?	Ich setze mich hierher?	Translation	(1sg.postverbal) (1sg.reflexive) (interrogative clause)	UniTO, UniVDA
S28	Stamattina non abbiamo fatto colazione, non abbiamo mangiato niente	-	Translation	(1pl.preverbal) (negative quantification) (negative concord)	UniTO, UniVDA
S29	-	Unsere Verwandten sind keine Vegetarier, wir aber schon. Wir essen nie etwas, bei denen.	Translation	(1pl.preverbal) (negative quantification) (negative concord)	UniTO, UniVDA
S30	È nevicato in montagna ieri?	Hat es gestern in den Bergen geschneit?	Translation	(weather expletive) (aux, weather verbs) (interrogative clause)	UniVR
T01	So che ieri i miei amici hanno dovuto lavorare tutto il giorno./ So che i miei amici devono lavorare tutto il giorno. Metti la frase nel PASSATO: So che ieri i miei amici ...	Ich weiß, dass meine Freunde den ganzen Tag haben arbeiten müssen/arbeiten haben müssen/ Ich weiß, dass meine Freunde den ganzen Tag arbeiten müssen. Setzen Sie den Satz in die VERGANGENHEITSform: Ich weiss, dass ...	Tense transformation	(1sg.preverbal) (3pl.m.preverbal) (clitic doubling) (verb complex) (possessive.1sg) (subordinate clause)	UniVR
T02	So che Mario potrà comprarsi una macchina nuova./ So che Mario può comprarsi una macchina nuova. Metti la frase nel FUTURO: So che Mario ...	Ich weiß, dass sich Martin ein neues Auto "kaufen können wird"/"wird kaufen können/ Ich weiß, dass sich Martin ein neues Auto kaufen kann. Setzen Sie den Satz in die ZUKUNFT: Ich weiss, dass ...	Tense transformation	(1sg.preverbal)(3sg.m.preverbal) (clitic doubling) (verb complex) (m.trochee.sbj.preverbal) (subordinate clause)	UniVR
T03	In alto in montagna è nevicato e in basso è piovuto/ In alto in montagna nevica e in basso in valle piove. Metti la frase nel PASSATO: In alto in montagna ...	Oben auf dem Berg hat es geschneit, unten hat es geregnet./ Oben auf dem Berg schneit es, unten im Tal regnet es. Setzen Sie den Satz in die VERGANGENHEITSform: Oben auf dem Berg ...	Tense transformation	(weather expletive) (aux, weather verbs) (locative adverbials)	UniTO, UniVDA, UniVR
T04	Ieri non avete incontrato nessuno qui./ Non incontrate nessuno qui. Metti la frase nel PASSATO: Ieri ...	Ihr habt hier niemanden getroffen./ Ihr trefft hier niemanden. Setzen Sie den Satz in die VERGANGENHEITSform: Ihr ...	Tense transformation	(2pl.preverbal) (negative quantification) (negative concord)	UniVR, UniTO, UniVDA
T05	Abbiamo parlato con Mario. / Parliamo con Mario. Metti la frase nel PASSATO: ...	Wir haben mit Martin gesprochen./ Wir sprechen mit Martin. Setzen Sie den Satz in die VERGANGENHEITSform: ...	Tense transformation	(1pl.preverbal) (m.trochee.pp)	UniVR
T06	Ho abbracciato Sara./ Abbraccio Sara. Metti la frase nel PASSATO: ...	Ich habe Sara umarmt./ Ich umarme Sara. Setzen Sie den Satz in die VERGANGENHEITSform: ...	Tense transformation	(1sg.preverbal) (f.trochee.obj.postverbal)	UniVR

Appendix 5.3 Audio quality control workflow

The main goal of the quality control process is to ensure that the data published on the map is free from empty, incomplete, unintelligible, or inappropriate entries, and that the audio files are concise, containing only the relevant content without any extraneous comments, prolonged silences, or background noise. The following sections detail the process step by step.

1. Get the needed files

Download the latest version of the repository from <https://zenodo.org/doi/10.5281/zenodo.8360169>.

Then, download the ZIP archive corresponding to the language variety you are checking (e.g., *cim.zip* for Cimbrian, *tir.zip*, *tre.zip*, etc.), along with the *metadata.zip* folder.

The language variety archive contains all the available audio files for that variety, while the *metadata.zip* folder includes essential information about the questionnaire and the participants.

Inside *metadata.zip*, there is a file named *questionnaire_vx.x.x.csv*, which contains all the stimuli along with their IDs. For instance, *S01* corresponds to the translation of the sentence *Mario ieri ha ballato tanto con Sara/Martin hat gestern lange mit Sara getanzt*, while *I01* refers to the description of image *I01* (images can be found in the folder *metadata > I_images*), which targets the verb *tagliare su/aufschneiden*.

Additionally, the *metadata.zip* folder contains a file named *users_results_vx.x.x.csv*, which includes participant information and the IDs of all audio files produced by each participant.

2. Check the audio files

The audio that still needs to be checked is found on the following shared spreadsheet in Google Drive.

https://docs.google.com/spreadsheets/d/11piE3hy07Ln5Lljl68oz_3TcSvJHVswj0TIEWLvSI8/edit?usp=sharing

The spreadsheet contains two tabs: *Overview by speaker* and *errors_versionx.x.x*.

In the *Overview by speaker* tab, you will find a list of all speakers whose recordings still need to be checked. The list is organized alphabetically by language variety: for example, Cimbrian (*cim*) speakers appear first, while Walser (*wae*) speakers appear at the end.

Each row includes the participant's ID (e.g., U0391), their language variety, location, age, and gender. This is followed by the IDs of all audio recordings available for that speaker. Each ID follows this format: *M01_vec_U0831*, where *M01* is the stimulus, *vec* is the language variety (in this case, Veneto), and *U0831* is the speaker ID.

Columns A (*Quality Check*) and B (*Errors*) are used to track progress. In Column A, mark a cell in green once all audio files from a speaker have been checked, or yellow if only partially checked. In Column B, you may add a note to indicate which parts have been reviewed and which still need attention.

Each audio file should be checked for issues that might compromise data quality or make it unsuitable for publication. Depending on the type of issue, one of two actions should be taken: the file can either be **trimmed** or **deleted**.

- **Deletion** is appropriate when the file is empty, the spoken content does not match the stimulus, the audio quality is extremely poor, inappropriate language is used, or the sentence is cut off mid-way.
- **Trimming** may be used if there is a long pause at the beginning or end of the file, background noise before or after the utterance, or if another person prompts the speaker (e.g., saying “*vai*” before the start).

The table gives an overview of the variety of the issues found in the checks and the number of audio files it pertained to (until version 1.1.1)

Issue	Version 1.0.0	Version 1.0.2.	Version 1.0.5	Version 1.1.1
Empty audio	3	40	48	4
Light noise	72	10	1	0
Audio cut off	4	29	60	54
Audio starts late or conversation before or after	25	14	499	249
Noise	3	7	201	131
Content: not in line with stimulus	0	28	93	51
Total	144	122	902	489

Listen carefully to each audio file to assess its quality. You may choose the work method that suits you best. For example, you might select a subset of participants (e.g., all Lombard speakers with IDs between U0600–U0620), mark their audio files in grey in the spreadsheet, and begin reviewing them one by one.

Listening to several instances of the same stimulus (e.g., *S01*) across different speakers can help identify outliers, especially when dealing with a variety you're less familiar with (e.g., Francoprovençal). Once a file is reviewed, remove the grey highlight to track progress. To mark which participants you are currently reviewing, highlight Column A in yellow; once all their audio is verified, change it to green.

If you encounter an issue in a specific file, copy its ID (e.g., *S02_frp_U0871*) and switch to the *errors_version x.x.x* tab. In Column A (*audio*), paste the audio ID. In Column B, describe the issue, following the existing format (e.g., *extra: long pause at beginning*).

For student assistants or other collaborators responsible only for audio checking, this marks the end of their task. Any further steps, such as uploading and updating the repository, should be handled by the person responsible for corpus management, as it is crucial to work within a unified folder structure.

3. Trimming files

For the audio files that can be trimmed, you can use any audio editing software (e.g. Audacity). In the software you open the file in question (e.g. *S09_lmo_U0424.flac*), trim off the pause, noise or comment and then save the file with the exact same name in the same folder. This results in the original file being replaced and no duplicates can show up.

Nota bene: make sure that the output is in the original format (.flac).

The ID of the trimmed audio file must then be added to the *ReadMe.html* file, which is being prepared for the next version of the repository. Open the file using a text editor such as Notepad, and list the audio ID under **Section 6: Updates**. It should also be included in the list of corrections for the upcoming version (e.g., from *v1.1.1* to *v1.1.2*).

In the *errors_versionx.x.x* tab of the Google Drive spreadsheet, indicate that the trimming has been completed by writing “**trimmed**” in **Column C** (signifying that the file has been edited), and enter “**y**” in **Column D** (indicating that the update has been recorded in the ReadMe file).

4. Deleting files

For audio files that must be deleted, remove the audio file from the corresponding folder and delete its ID from the *users_results_vx.x.x.csv* file.

The ID of the deleted file must then be added to the *ReadMe.html* file, which is being prepared for the next repository update. Open the file in a text editor (e.g., Notepad) and record the deletion under **Section 6: Updates**. It should also be listed among the corrections included in the upcoming version (e.g., *v1.1.1* → *v1.1.2*).

To verify that the deletion has been carried out correctly, search for the deleted audio ID in the corpus: only the *ReadMe* file should appear. The audio file itself and the reference to it in *users_results_vx.x.x.csv* should no longer be present.

In the *errors_versionx.x.x* tab of the Google Drive spreadsheet, mark the deletion by entering “**deleted**” in **Column C** (indicating the file has been removed) and “**y**” in **Column D** (confirming that the change has been recorded in the *ReadMe* file).

5. Using the AAAA Tool (Automated Audio Acceptability Assessment) for Faster and Efficient Quality Control

To improve and speed up the process of reviewing, trimming, and deleting audio files in the AlpiLinK corpus (while reducing the risk of human error) a semi-automated quality control software based on PRAAT is available. This tool, called the Automated Audio Acceptability Assessment (AAAA) and developed by Joachim Kokkelmanns (see WP4.3), carries out automatic acoustic analyses to help prioritise which audio files need manual verification.

Once launched, AAAA presents the audio files of a specific speaker, which the user selects by modifying the *usercode\$* variable in the script (e.g., U1547). The script is run through Praat, and for each file, the researcher can either validate it (*SAVE-unmodified*), trim it using Ctrl+X and save it (*SAVE-trimmed*), or delete it if unsuitable (*REMOVE*). These actions are performed through a user-friendly interface with corresponding buttons, and all actions are automatically logged in the file *AAAA.log*.

Important note: This log file is overwritten with each new session, so users must copy the relevant lines (e.g., S06_vec_U1477 trimmed) into the shared Google Sheet immediately after completing their review, following the procedure mentioned in previous sections. Additionally, users must verify the content of files *S01–S30* to ensure that the utterance corresponds meaningfully to the intended stimulus (as listed in *questionnaire_xxx.csv* in the *metadata* folder). Utterances that are unintelligible should be marked as *incomprehensible*.

To use the tool:

- Download Praat: <https://praat.org/>.
- Open the script *AAAA.praat* in the AAAA folder.
- Edit line 6 to define the folder path, e.g.:
`foldername$ = “../AlpiLinK Corpus 1.2.0/vec”.`
- Edit line 8 to specify the speaker code, e.g.:
`usercode$ = “U1547”.`
- Run the script and follow the interface instructions
- After reviewing, send the zipped audio folder via WeTransfer to vinko@ateneo.univr.it.

This optional tool significantly improves efficiency and consistency in quality control, particularly for larger speaker subsets or unfamiliar varieties.