DOI: https://doi.org/10.14232/wpcl.2022.7.3

Tools for data processing and visualization in the project VerbaAlpina

Beatrice Colcuc and Florian Zacherl[1]

1. Project overview

Since the beginning of the project in 2014, VerbaAlpina (https://www.verba-alpina.gwi.uni-muenchen.de/) has been a research project which represents a connection between linguistics and informatics from the perspective of the digital humanities. VerbaAlpina is a cooperation between the Institute of Romance Philology and the Center of Digital Humanities of Munich University (http://www.itg.lmu.de) and combines linguistics and information technology. An extensive, multilingual research environment with several functional areas has been created by using up-to-date media technology. Thus, a prototype for the transfer of traditional geolinguistics into digital humanities has been developed.

VerbaAlpina has always emphasised the importance of not considering the research activities solely from an internal perspective but inserting the project into a broader framework, by appropriating innovative methods offered by digital technology and, in particular, by developing these digital tools and methods further. In this way, VerbaAlpina has always envisioned the project from the perspective of FAIR thinking, even before the four cardinal principles of modern and digital research were formulated in 2016 by Wilkinson and Dumontier et al. (2016).

1.1 The area under investigation: the Alpine region

VerbaAlpina[2] is a research project based at Munich University which has been funded by the *DFG* (German Research Foundation) as a long-term project since 2014. VerbaAlpina seeks to investigate the linguistic and cultural area of the entire Alpine region from a transnational perspective. The area under investigation corresponds to the perimeter of the Alpine Convention (an international treaty between all bordering states of the Alpine region)[3] and covers a surface of 190,600 km^2. Thus, different countries are part of the project, namely Germany, Austria, Switzerland, Italy, France, Liechtenstein, Slovenia and Monaco.

From an ethnographic and topographic point of view, the Alpine area shows a constant homogeneity throughout the territory. From France to Slovenia livestock breeding, milk and cheese production, and also flora and fauna are common to the whole mountain region. However, when looking at the linguistic composition of the Alps, great heterogeneity emerges. In this area, varieties from three language families are spoken, namely Romance, Germanic and Slavonic. This plurality manifests itself not only

[1] We would like to thank our colleague Markus Kunzmann with whom we held our talk during the workshop. He contributed significantly to the conception and the contents of this paper.

[2] The complete name is “VerbaAlpina. Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit” (VerbaAlpina. The Alpine cultural region reflected through its multilingualism; Krefeld/Lücke 2014-).

[3] https://www.alpconv.org/en/home/.

in the number of standardised national languages (i.e. French, Italian, German, Romansh and Slovenian) spoken and written in the territory but above all in the various local varieties which are very well preserved in a large part of the region. The fragmentation of the Romanic area is more pronounced compared to the Germanic zone, and the latter, in turn, is more fragmented than the Slavonic zone. Alpine dialects are historically primary ones.[4] This means that they have developed before standardised languages could emerge and cover these as umbrella languages (cf. Krefeld 2020).

For VerbaAlpina it is fundamental to differentiate between the large (in terms of the number of speakers) standard languages spoken and written in the Alpine region (Italian, French, German, and Slovenian), the smaller standardised languages (Romansh, Ladin, and Friulan) and the language varieties of the dialect continuum. VerbaAlpina is mainly interested in the diatopic variation of the Alpine region. In other words, VerbaAlpina considers only dialectal words as its primary data. During the third stage of data processing, namely during the typification, dialectal words are grouped and matched with lemmas from standard dictionaries as will be shown below.

1.2 Research aims

As a research project, VerbaAlpina pursues several aims that concern both linguistic and IT (or infrastructural) aspects.

On the one hand, VerbaAlpina aims to investigate the vocabulary of the dialect varieties of the entire Alpine region in a selective (dealing with different semantic domains) and analytical (processing data in different steps) way. The project was planned in three stages: During the first stage (from October 2014 to October 2017) the focus was on vocabulary from the field of Alpine pasture farming, especially milk processing. The second stage (from November 2017 to October 2020) was dedicated to the vocabulary relating to flora, fauna, landscape formations, and weather. The subject of investigation of the third stage (from November 2020 to October 2023) is the vocabulary of modern life in the Alps, especially ecology and tourism. This approach helps to detect and to underline differences and similarities not only between linguistic varieties within the same language family but especially between linguistic varieties of different language families. One of the aims is therefore to recognize connections regarding the etymology of the individual dialectal words and to reconstruct their historical linguistic paths. Many words share a common etymology, even if this cannot be seen anymore at first sight: e.g. German *Butter*, French *beurre* and Italian *burro* are immediate developments from the Greco-Latin *butyru(m)*. By adopting this approach, VerbaAlpina overcomes the traditional boundaries of nation-states carrying out broader geolinguistic research.

On the other hand, VerbaAlpina is engaged in setting up a portal by using modern media technology: documentation, tools for data collection and processing, and collaborative development are aspects on which VerbaAlpina works.

In order to create a solid network of cooperation in the Alpine region, VerbaAlpina can count on numerous project partners. These include sister projects, cultural institutions in the Alpine region, and organisations involved in scientific research in the fields of language and information technology. The list of partners can be viewed at the following address: https://www.verba-alpina.gwi.uni-muenchen.de/en/?page_id=185&db=202.

[4] An exception is represented by the Walser community.

2. Data gathering and processing

VerbaAlpina works on a lexical level. Specifically, VerbaAlpina is about recording data from printed or digital sources in a structured way in a database. The data VerbaAlpina gathers and analyses derives, on the one hand, from printed linguistic atlases and geo-referenced dictionaries from the past one hundred years. For instance, VerbaAlpina deals with the *Sprach- und Sachatlas Italiens und der Südschweiz* (Jaberg and Jud 1928–1940), the *Atlas linguistique de la France* (Gilliéron and Edmont 1897–1900) and the *Dicziunari Rumantsch Grischun* (DRG; De Planta et al. 1938ff). VerbaAlpina also disposes of digital data from partner projects (e.g. the Bavarian Dialect Database which contains the surveys for the language atlas projects from the Bavarian Language Atlas).[5] Moreover, linguistic data acquired from atlases and dictionaries are supplemented through crowdsourcing. The crowdsourcing platform was designed within the project and directly addresses speakers of the Alpine dialects. By this new collection of current linguistic material, inconsistencies between the existing sources shall be evened out, gaps shall be eliminated and obsolete designations shall be marked as such. In this approach, the Alpine region can also be analysed from a diachronic perspective (cf. Krefeld and Lücke 2020a).

Concerning data processing, VerbaAlpina has to face the challenge that consists in the lack of uniformity of data from different sources as the data is not structured in the same way. In the VerbaAlpina portal, sources from different research traditions (Romance studies, German studies, and Slavonic studies) which were completed at different historical stages of dialectological research are brought together. To unify the linguistic material, data from printed sources first has to undergo a transcription process. The fact that data is so heterogeneous implied that there was a need for appropriate tools to satisfy the two opposing principles of reliability to the source and the easy comparability of the linguistic material.

2.1 Transcription

The linguistic material is entered into the MySQL relational database through a transcription system based only on ASCII (American Standard Code for Information Interchange) characters. The processing of linguistic attestations is therefore possible with any keyboard, by any user and at any time. The system used by VerbaAlpina, following the terminology used by the *Thesaurus Linguae Grecae* (TLG; Brunner 1972-), which first developed this system in the early 1970s for the electronic recording of ancient Greek texts, has been named “Beta Code”.

This system establishes that each linguistic character contained in an atlas or in a dictionary used by VerbaAlpina corresponds to a combination of ASCII characters.

In order to follow the two main principles mentioned above, which require the reliability of the source and the comparison between data, VerbaAlpina has decided to reproduce linguistic material graphically in two different ways:

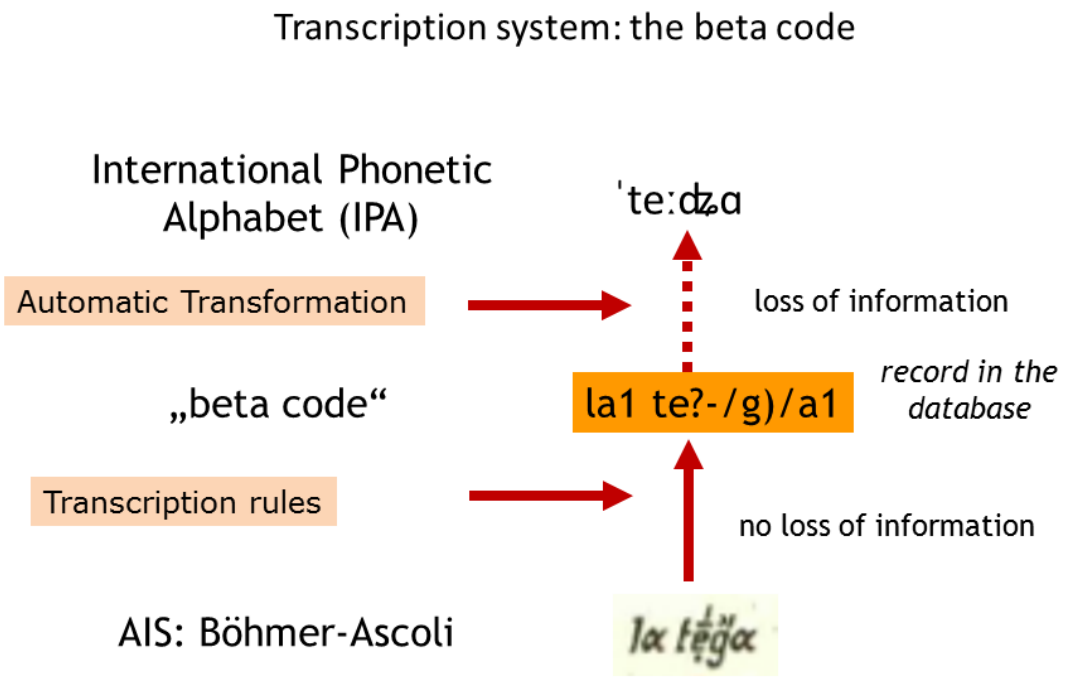
(1) Original transcription: As already mentioned, VerbaAlpina deals with sources that belong to different historical research periods. Therefore, it is necessary to respect the original transcription. Nevertheless, due to technical reasons, it is impossible to keep certain conventions unchanged, especially when several diacritical symbols overlap a letter. In VerbaAlpina, using the Beta Code, this vertical system of characterization of a letter is transferred to linear sequences of characters.

[5] For the complete list of atlases and dictionaries cf. Colcuc/Krefeld (2020).

For the beta encoding, ASCII characters that resemble (as far as possible) the original characters are used. This graphical rendering of the original transcriptions using Beta Code avoids possible loss of information.

(2) IPA: Using specific substitution routines, all Beta Codes are transferred to IPA characters. This is crucial for comparability and user-friendliness. The transformation of a sequence of characters into IPA can sometimes lead to a slight loss of information. In fact, the opening of vowels in IPA is much less precise than in transcription systems such as Böhmer-Ascoli or Theutonista.

Figure 1 *The transcription system of VerbaAlpina*



As briefly mentioned above, VerbaAlpina also works with data from digital databases and with data collected directly from speakers via crowdsourcing. Digital data from partner databases is automatically imported into VerbaAlpina and therefore does not need to be transcribed via the transcription tool. Some atlases are converted into beta code before being imported, others are imported in their beta code so that they can be further transformed into IPA. Other sources are directly imported in IPA transcription. We also deal with sources that are imported in an existing Unicode representation, but in these cases an IPA conversion can still be carried out. In principle, VerbaAlpina has no IPA representation for data from crowdsourcing because it is not possible to convert this type of attestations automatically.

2.1.1 Transcription rules and transcription tool

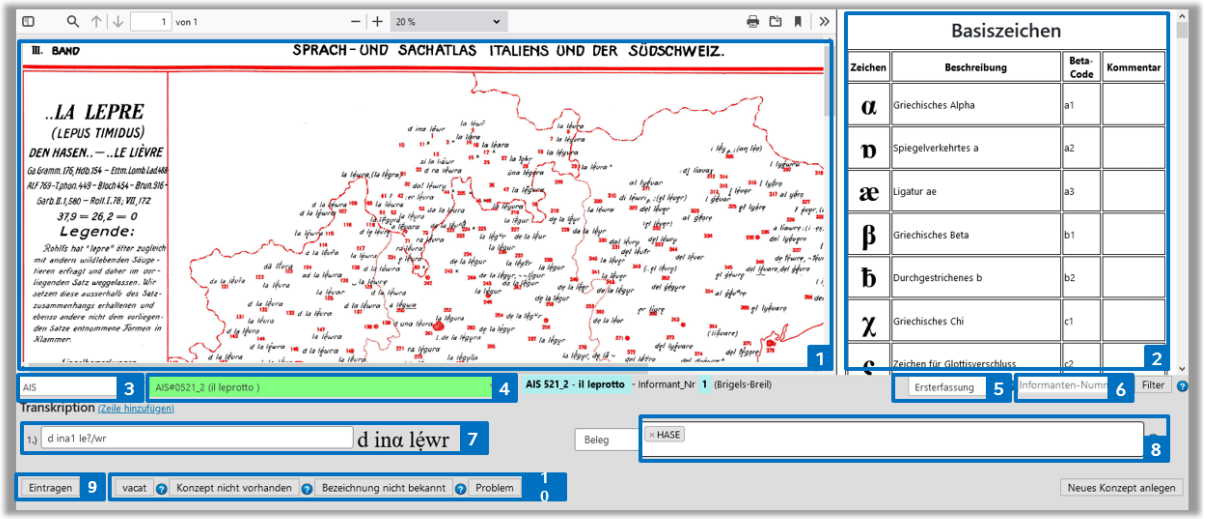
The differentiation between base signs (letters) located at the baseline and diacritics marked above and below base signs is crucial. Special sizes or positions of letters, for example when they are displayed smaller than other base signs in the source, are treated like diacritics.

Base signs that exist as ASCII characters are retained: That is to say all Latin characters. They are transcribed by a combination of a letter and a numeral. Diacritics are always placed after the base sign to which they are assigned. When several diacritics are assigned to one base character, they are transcribed from the bottom left to the top right. Each ASCII character used for the transcription of a diacritic may occur only one time per base character. For the repetition of the same diacritic, the Beta Code requires the use of special rules, e.g. \2 for a double grave accent. For the complete list of transcription rules cf. Lücke and Zacherl (2020).

Figure 2 *Differentiation between base sign and diacritics and transcription order*

The transcription tool was designed within the project. The main screen that appears using the tool contains first of all an original map from the source. On the side, there are the transcription rules. Below the map are two filters that allow to select the source and the map (stimulus, i.e. the question asked by the explorer during the data collection) that has to be transcribed. Once the map has been selected, the system asks to enter transcriptions for each survey point. The transition from one point to another is automatic: for each source, the system recognises which points are geographically located within the Alpine Convention and offers only those points for the transcription. The window for entering the transcription using the Beta Code rules is located below the filters and, to its right, the space for selecting and applying a concept to the transcribed utterance.

Figure 3 *The transcription tool of VerbaAlpina*



1 – Original map
2 – Transcription rules
3 – Data source
4 – Stimulus
5 – Entry/correction/problem
6 – Selection of the informant number
7 – Transcription in Beta Code and original representations
8 – Concept assignment
9 – Entry button
10 – alternative utterance format

2.2 Tokenization

The tokenization represents the second step towards the unification of the linguistic material. Through tokenization, the linguistic expressions which were transcribed, imported or entered into the database via crowdsourcing are fragmented into single tokens. After tokenization, the linguistic expressions are prepared to be used for the third step of the data processing, i.e. for the typification.

Table 1 *Overview of the tokenization process*

Attestation in Beta Code	**Attestation in IPA**	**Concept**
una1 mu:g/a1 da1 va/c)/	unɑ myʤɑ dɑ vˈatɕ	HERD OF COWS
TOKENIZATION		
una1	unɑ	ARTICLE
mu:g/a1	myʤɑ	HERD
da1	dɑ	PREPOSITION
va/c)/	v ˈatɕ	COW

Tokenization is carried out through a special tool that was developed within the project by entering the ID-number of the stimulus which corresponds to the map that one wants to tokenize. At this stage, if transcription errors are detected or if it is not possible to convert a string into IPA, the system displays the notification so that any errors can be corrected before the linguistic material is tokenized. Once tokenized, the linguistic attestations can be found via the interactive map. The search works through different filters which allow searching data from an onomasiological perspective (from the concept to the designations) as well as in a semasiological way (from the designation to the concepts) (cf. chapter 3).

2.3 Typification

After they have been tokenized, linguistic attestations can be further typified. The morphological typification of the linguistic data, which means the grouping of data according to their internal linguistic characteristics, is one core task of VerbaAlpina. Through the typification, VerbaAlpina aims at structuring the complex variety of the numerous linguistic attestations (tokens) so that comparisons among data are possible. A morpho-lexical type is defined by the following properties: language family, part of speech, single word vs. affixed words, gender, lexical base type. The morpho-lexical types represent the central category in the management of linguistic data and they are comparable to the lemmas which are listed in dictionaries.

Morpho-lexical types are specific to one language family. The form by which a morpho-lexical type should be represented (also in the search function of the interactive map) is given by the lemmas of selected reference dictionaries such as DWB (for the Germanic varieties), TLFi and Treccani (for the Romanic varieties) and SSKJ (for Slavonic) (cf. Krefeld and Lücke 2020b). In the case of Germanic and Slavonic tokens, it is rather easy to find the appropriate form because these two language families are represented by only one standardised language (German and Slovenian). For example, all the phonetic forms of Alemannic and Bavarian, which are variants of one morpho-lexical type, can be grouped under the same standard form. If standard variants do not exist, the lemmas of the most important regional reference dictionaries (Idiotikon, WBOE) are used (cf. Krefeld and Lücke 2020b). In the case of the Romance language family, the situation is more complex because of the numerous minor languages, some of which are still not sufficiently standardised. VerbaAlpina decided to represent all morpho-lexical types by the French and Italian standard forms (e.g. *beurre/burro* ‘butter’; *lait/latte* ‘milk’). If only one of these two standard languages has a suitable form, the type is constituted by this single term as in the case of *ricotta*, for which any corresponding form lacks in French. Similar to Germanic and Slavonic forms, if neither TLF nor Treccani contains any appropriate lemma, VerbaAlpina uses dialectal reference dictionaries (LSI, BLad) (cf. Krefeld and Lücke 2020b).

The typification of the linguistic material represents a further step towards the harmonisation of data, but above all, it allows to highlight linguistic convergences and divergences between the three language families of the Alpine region.

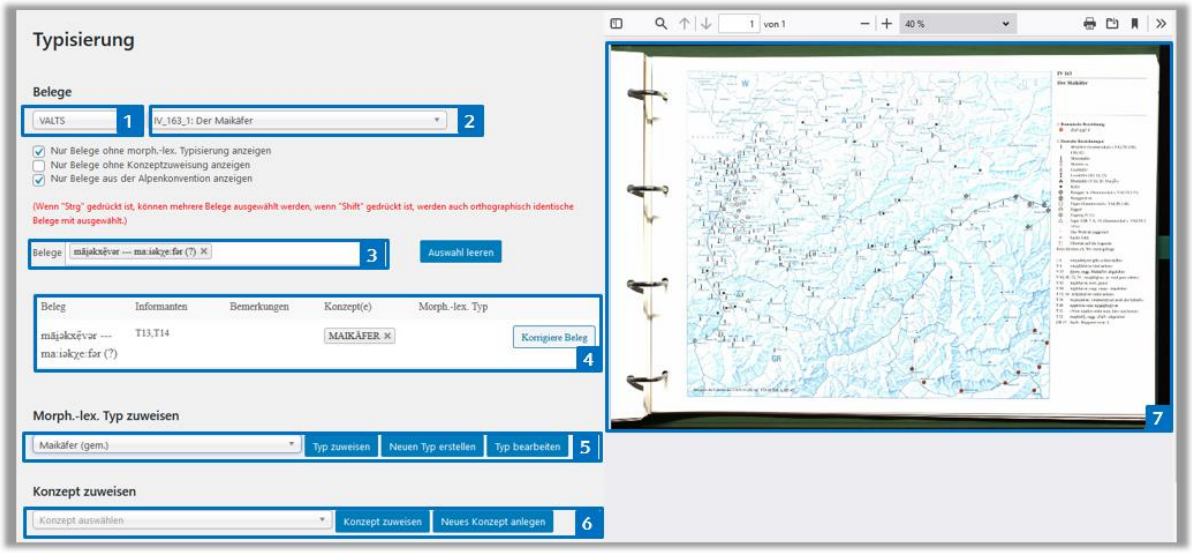
Each morpho-lexical type is described by entering linguistic information such as language family, part of speech, presence of affixes, and gender. At least one lemma from one of the reference dictionaries mentioned above is then applied to each morphological type. In addition, a so-called "base type" is applied to each morpho-lexical type. A base type is to be understood as the first historical attestation from which morpho-lexical types developed. In many cases, the base type corresponds to the etymon, with the only difference that an “etymon” refers to the immediate historical antecedent of a word, while a base type, as the name suggests, represents the older form (cf. Krefeld and Lücke 2018). Specific reference dictionaries (e.g. FEW, Georges, AWB) are also used for the selection of the base types.

Table 2 *Example of typification*

Token	kˈa:vra	kabrˈuŋ	kavrˈɛt	kawrˈɛt
Language family	roa	roa	roa	roa
Part of speech	noun	noun	noun	noun
Affix	-	+	+	+
Gender	f	m	m	m
Morpho-lexical type	capra	caprone	capretto	capretto
Base type	lat. capra	lat. capra	lat. capra	lat. capra

If a base type is marked as unknown or as not sufficiently clear by the reference dictionaries, VerbaAlpina applies a question mark as in the example *(?) battuere*. When it is not possible to determine the base type, we tend to use an unknown type represented by only one question mark.

Figure 4 *The typification tool of VerbaAlpina*



1 – Data source (linguistic atlas)
2 – Stimulus
3 – Selection of the attestations
4 – Attestations and properties
5 – Selection / creation / editing of morpho-lexical types
6 – Concept assignment
7 – Original map

3. Publication and Visualization

VerbaAlpina provides two main access points to the collected linguistic information, namely an interactive map and the so-called *Lexcion Alpinum*, a dictionary-like view on the data. This goes in accordance with the traditional ways of publishing in either linguistic atlases or dictionaries. A key difference to these established sources is the manner in which the information is provided. Whereas a linguistic atlas generally offers an onomasiological perspective and a dictionary a semasiological one with no (or at least very restricted) possibilities to approach the other way round, the digital counterparts that are developed for VerbaAlpina intentionally allow both perspectives each and therefore underline that the existing restrictions are mainly imposed by the paper-based publication form and not by the basic concepts of information representation.

3.1 Geolinguistic representation

As mentioned in chapter 2, all of the utilised linguistic data is geographically classified on principle which is a necessary condition to spatially visualize it in its entirety. To attain a presentation as consistent as possible VerbaAlpina utilises the respective administrative municipalities as a reference system, i.e. each linguistic attestation that is documented within the borders of a municipality is assigned to its geometrical centre.[6]

In addition to the linguistic core data, there are multiple other geo-referenced datasets added for the visualisation that are not strictly linguistic, but contain information relevant especially in the Alpine context. This includes areal data (e.g. the traditional language areas or administrative borders like nation-states, districts etc.) and point data (e.g. archaeological find spots or ancient and contemporary infrastructure placements). The principle idea in including this material is to allow the user to combine it with the linguistic data at will and thereby assist them to find patterns and draw conclusions about extra-linguistic factors that affect linguistic phenomena.

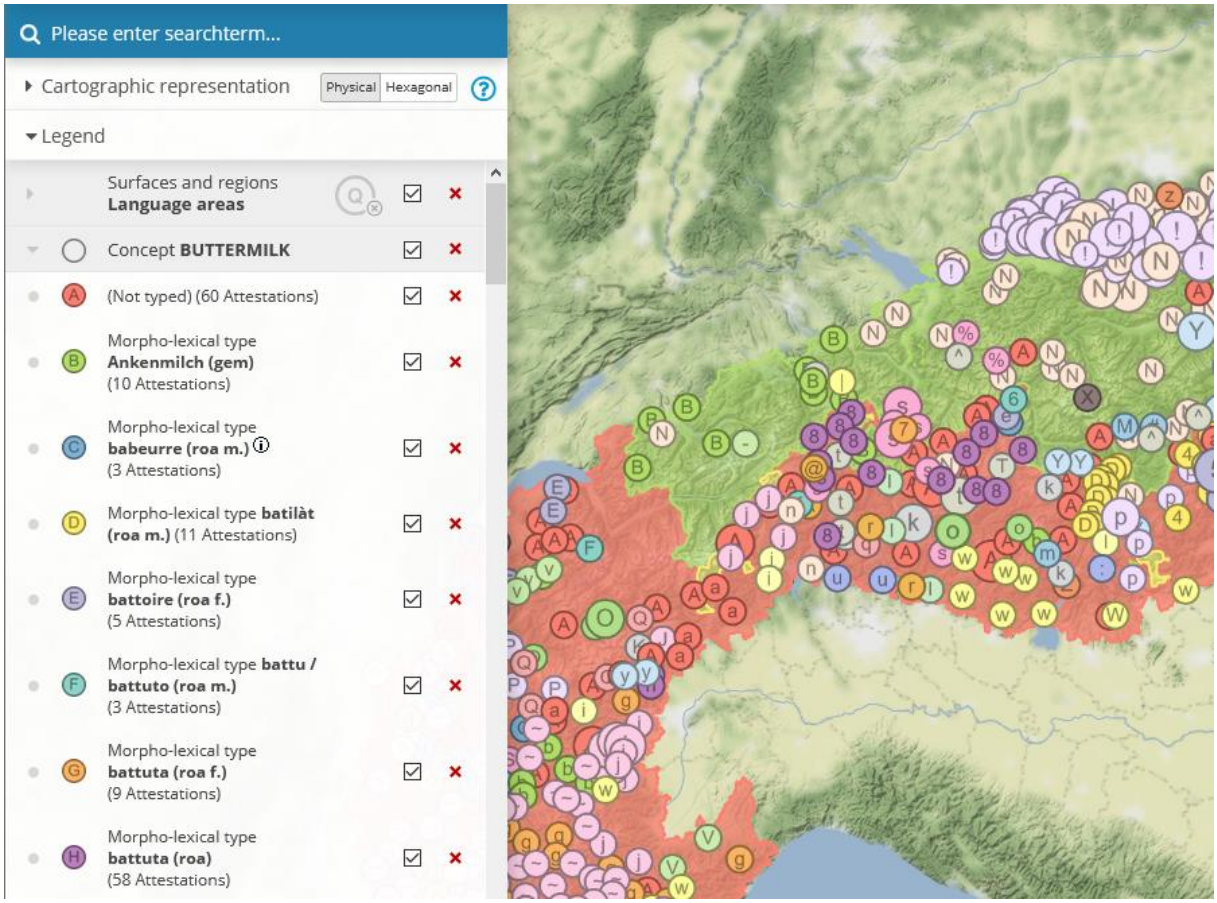
3.1.1 The Interactive Map

The main design choice concerning VerbaAlpina's interactive map is to allow the user to freely choose from the available linguistic and extra-linguistic data and to combine, filter and visualise in a way that fits their needs and interests best. Consequently, it starts off with an empty map that only contains the language areas in the Alpine region (which also can optionally be removed). The section *Cartographic representation* then allows to add new data to the map whereas the section *Legend* explains the symbols and colours that are shown on the map for the current selection.

The linguistic attestations themselves can be selected and grouped by three principal categories which occurred in the previous chapters: Morpho-lexical types, base types and concepts. Each one allows to show all attestations that are connected to the respective category and to group them appropriately. For example, the selection of a morpho-lexical type or base type allows (among others) grouping by concept which represents the classical semasiological perspective, whereas the selection of a concept vice versa allows grouping by one of the type variants to create an onomasiological representation. Figure 5 illustrates that for the concept BUTTERMILK and the respective morpho-lexical types.

[6] The interactive map allows this default behaviour to be changed via the options menu and each attestation to be shown exactly at the point where it is localised. This can be useful for source material with a particularly tight informant net in which many informants are located within one municipality.

Figure 5 *Legend and map representation for the selection of the concept* BUTTERMILK

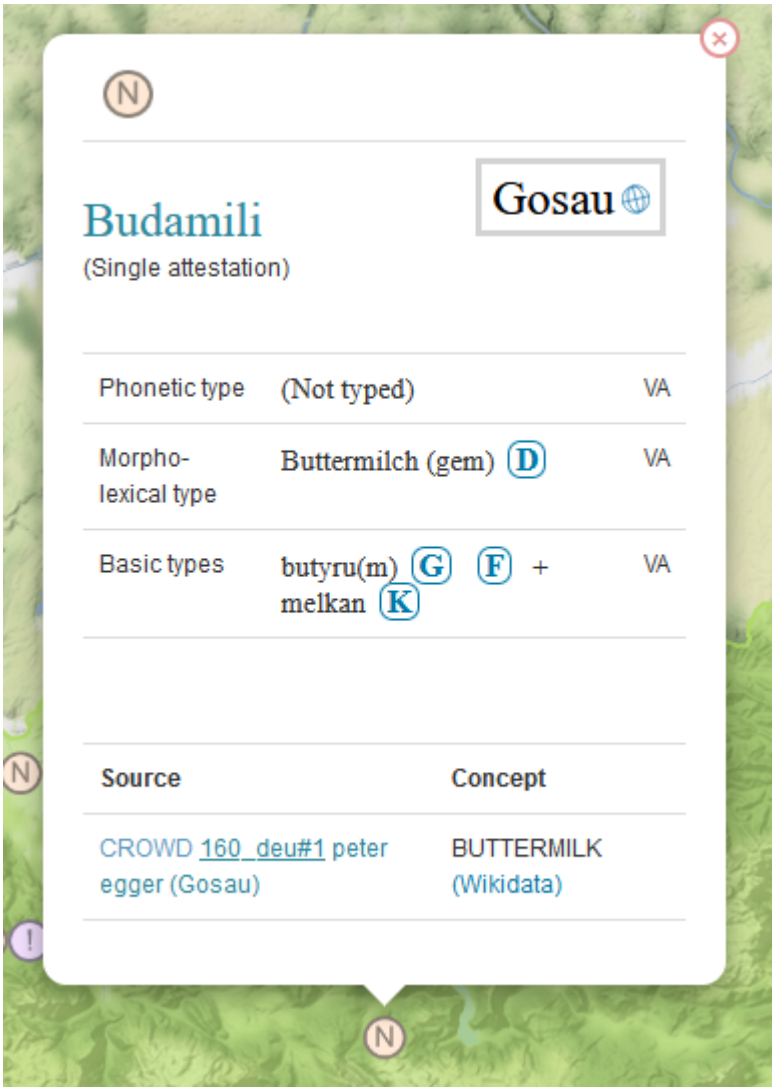


Depending on the particular category, there are also various filter and sorting possibilities to further customize the display. Additionally, the elements in the legend can be manually filtered so that sub-elements which are not relevant can be hidden or removed.

All point data (either linguistic attestations or extra-linguistic point data) is visualised on the map using coloured symbols that contain letters or numbers. The main category is specified by the symbol shape (circle, square, hexagon, etc.) whereas the colour and letter/number indicate the sub-category (cf. Figure 5). This allows multiple groups of attestations to be shown at once while they still can be distinguished from each other. Multiple symbols at the same location are joined to a larger symbol which slowly grows on a logarithmic scale until it reaches a maximum size. All areal data is represented by coloured part-transparent polygon overlays.

While the map representation alone gives an overview over the distribution of the different types or concepts in space, it is also possible to access the full information about one single attestation. Following the so-called *visual information seeking mantra* "overview first, zoom and filter, then details on demand" (Shneiderman 1996) this can be achieved by opening an extra popup window for each overlay on the map. Among others this contains details about the utterance itself, its typification and the source it came from. Many elements in this window are interactive to give further information or link to external resources like the dictionary references added during the typification process.

Figure 6 *Detail view for a specific attestation*

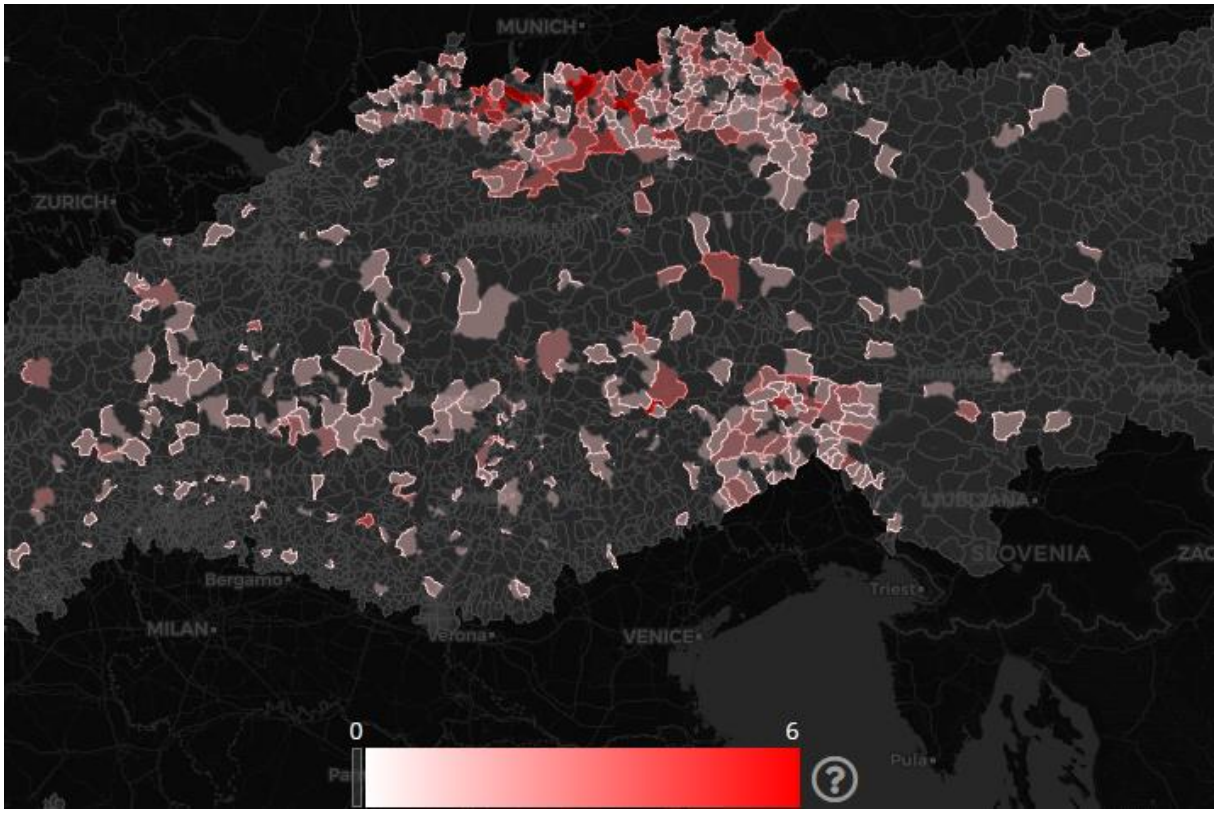


An important feature of the interactive map is the possibility to store its current state. This includes data selection, position and zoom of the map and other options like the different visualisation modes presented in the next chapters. In principle there are two possibilities to achieve this: The creation of a so-called *synoptic map* or the creation of a share-link. Technologically both variants store the current state in the VerbaAlpina database, but while a synoptic map contains a name and a description and can be made available for other users, the share-link solely produces an URL which can be used to re-create the current state.

3.1.2 Qualitative vs. Quantitative Presentation

In addition to the symbol-based visualisation that was presented in the previous chapter and which is activated by default, the interactive map offers a second, more aggregated way to show the data. The *quantitative view* produces a heat map to depict the distribution of attestations. It is possible to choose any of the polygon layers provided and project the current data selection on it using the *Q* button in the respective legend entry. Figure 7 shows the map from figure 5 in quantitative view.

Figure 7 *Quantitative view for the concept* BUTTERMILK

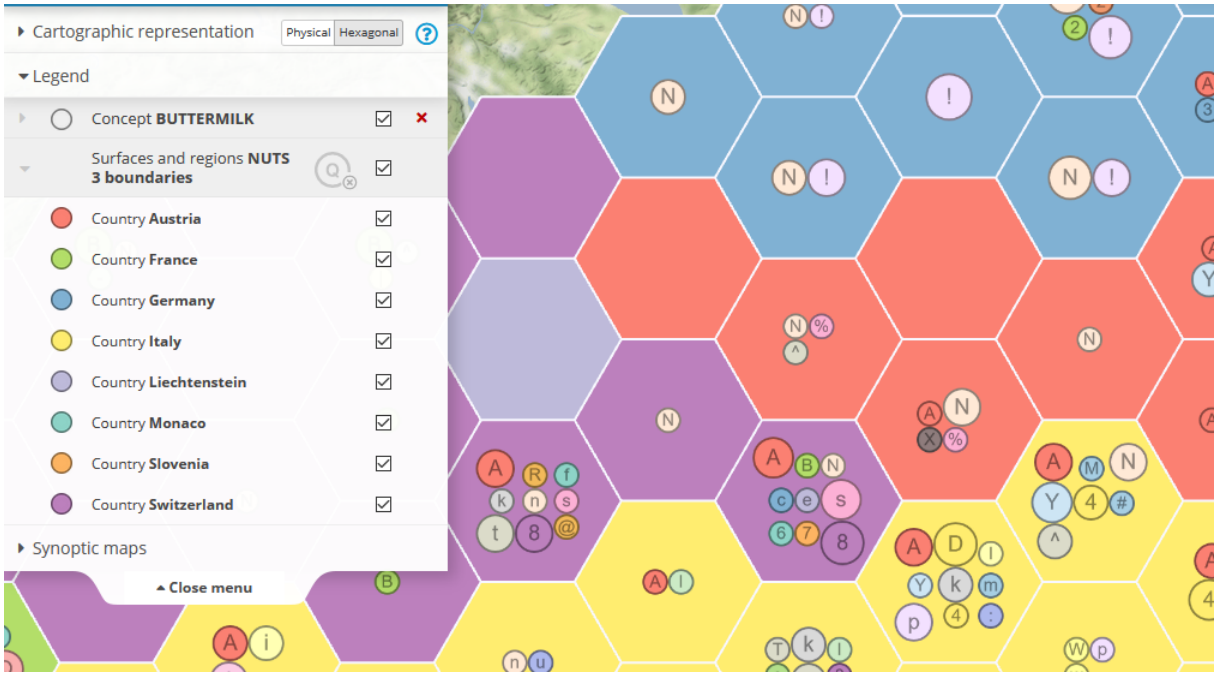


While the normal qualitative representation allows to find patterns in the visualised data and to access all details, the quantitative view can be used for more specific tasks. It allows the evaluation of the denseness for a certain portion of the data, which on the one hand helps internally to find gaps in the data, which for example can be specifically addressed by promoting crowdsourcing for certain areas. On the other hand, it is also useful for interpretation of the visualised data as it can be compared with population density maps or the like and may also provide information about the geographic focus of the data material from specific sources.

3.1.3 Geographic vs. Abstract Presentation

A second possibility to alter the nature in which the interactive map is displayed is to change to the abstract mode. In opposition to the *geographic mode* the different polygon layers are simplified to hexagons and arranged in a grid that tries to retain the neighbourhoods of the original polygons as close as possible. At the moment this is only possible for the so-called NUTS-3 borders, that are administrative borders one level above the municipalities, and the languages areas. In the future another grid consisting of the municipality borders will be added. In qualitative mode the point symbols are arranged at the centres of the respective polygons, in quantitative mode the hexagons are coloured just like the original polygons.

Figure 8 *Abstract view for the concept* BUTTERMILK

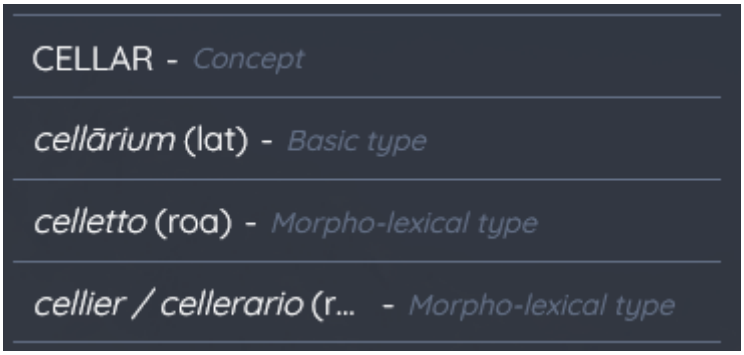


The main reason for introducing this extra mode of visualisation is to counter the effect that larger regions might be considered more important than smaller regions. Especially in the Alpine region, large municipalities or other administrative units are often very thinly populated whereas bigger towns show up very small on a geographic map. If each one of them is presented with the same size, this effect is removed, although precision and the direct recognition of specific geographic features get lost.

3.2 Lexicographical representation

The second entry point to the collected VerbaAlpina data is the dictionary-like *Lexicon Alpinum*. Its surface consists of a title bar on the left and the main area that contains a list of open articles. The title bar lists all elements from the main categories (morpho-lexical types, base types and concepts) in alphabetical order. Figure 9 shows a small section that illustrates this.

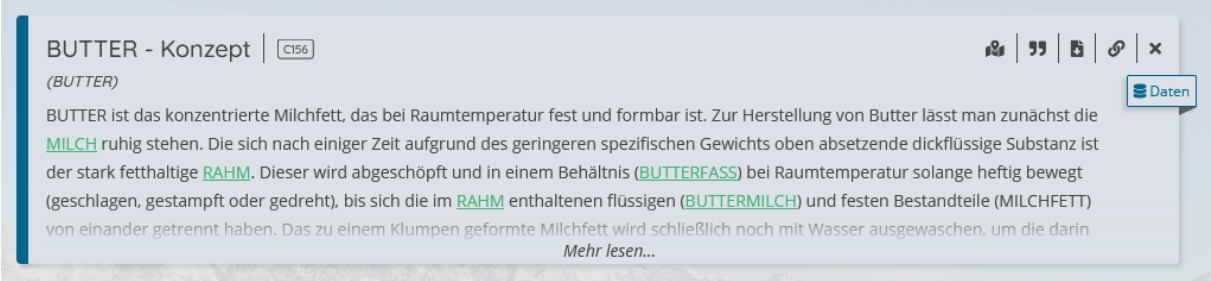
Figure 9 *Excerpt from the article list of the Lexicon Alpinum*



Each entry is visualised on a card which has a front and a back side. The front side contains general information about the specific entry and links to connected sites like the representation on the map and an API call to receive the underlying data. If existing, it also contains a comment explaining specific

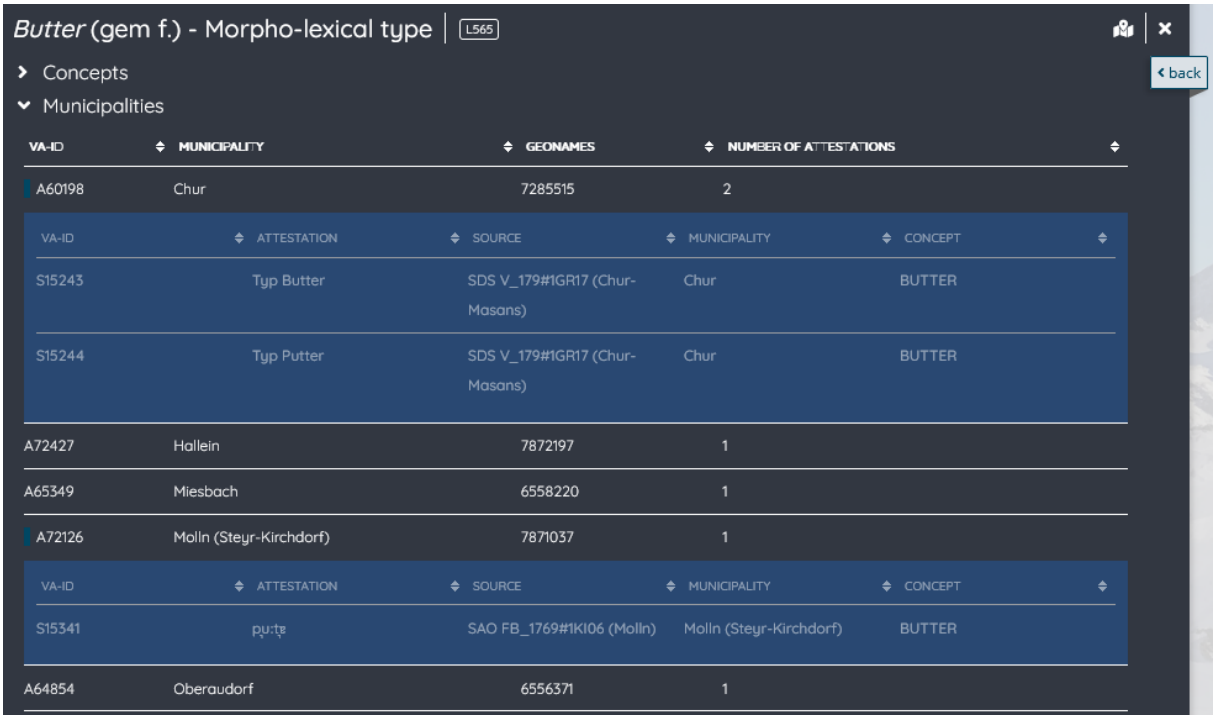
information about this entry[7] and a template to cite this specific entry. The majority of entries does not have an explanatory text, though, since it is only added, if there are no other online resources available which contain this information. Especially for concepts the connection to Wikidata (and indirectly to Wikipedia) generally holds more value than an own mostly redundant description.

Figure 10 *Front*



The back side allows to access the actual linguistic data in detail. Like for the map, there are different sub-categories depending on the main category, e.g. concepts for both kinds of types and types for concepts vice-versa. Like mentioned before, this also mirrors the two traditional lexicographic perspectives on the data. Also, all entries contain a sub-category *Municipalities* which allows to group all attestations geographically (cf. figure 11). Hence it is possible to get a representation very similar to a traditional dialectal atlas that lists variants for a number of locations. For every sub-category the elements can be further extended to show information about the specific attestations, similar to the details info window on the interactive map.

Figure 11 *Back*



[7] This text can also be accessed via the legend of the interactive map.

Analogously to the feature of the interactive maps, it is similarly possible to store the current selection in the *Lexicon Alpinum* by creating a share link. This stores the opened articles as well as more specific details (the position which the user scrolled to, if the front or back side of an article is shown, which elements are extended, etc.)

4. Conclusion

This paper describes the major tools that are used in the VerbaAlpina project to process and publish linguistic data. All program code for these tools is made public in half-year intervals via multiple GitHub repositories and can be found here: https://github.com/VerbaAlpina/. In addition to the main tools, which mostly are realized as separate WordPress[8] plugins, there is a multitude of smaller tools that are exclusively used internally for specific tasks. Examples are a small tool which allows the editing of the hierarchical representation of concepts and sub-concepts or one that checks the text published on the VerbaAlpina homepage for structural errors. These are (like the main tools) developed in close collaboration between the linguistic and informatics staff and are probably too specific to be used in different contexts. However, they are included in the main VerbaAlpina plugin repository at the previously given URL.

VerbaAlpina also provides an API through which the data can be accessed in defined formats programmatically. The selection of the data and the output format are controlled by URL parameters. The API of VerbaAlpina can be found at https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=8844&db=xxx.

References

AWB = Karg-Gasterstädt, Elisabeth and Frings, Theodor (eds.) 1952-2015: *Althochdeutsches Wörterbuch, digitalisierte Fassung im Wörterbuchnetz des Trier Center for Digital Humanities*, http://awb.saw-leipzig.de/ (accessed 27.05.2021).

BLad = Istitut Cultural Ladin Majon Di Fascegn 2007: *Banca lessicala ladina*. Vigo di Fassa, http://blad.ladintal.it/applications/dictionary/siteHistoric/index.jsp?_VP_V_ID=834120465 (accessed 27.05.2020).

Colcuc, Beatrice and Krefeld, Thomas 2020: *Sprachatlanten und Wörterbücher im Alpenraum*, Methodologie, VerbaAlpina-de, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D202%26letter%3DS%2352.

DRG = De Planta, Robert, Melcher, Florian, Pult, Chasper, and Giger, Felix 1938–1940: *Dicziunari Rumantsch Grischun*. Chur: Institut dal Dicziunari Rumantsch Grischun, http://online.drg.ch/ (accessed 27.05.2020).

DWB = Grimm, Jacob and Grimm, Wilhelm 1854-1961: *Deutsches Wörterbuch von Jacob und Wilhelm Grimm, digitalisierte Fassung im Wörterbuchnetz des Trier Center for Digital Humanities*, Version 01/21, https://woerterbuchnetz.de/?DWB, (accessed 27.05.2020).

FEW = Wartburg, Walter 1922-1967: Französisches etymologisches Wörterbuch. Basel: Zbinden, https://apps.atilf.fr/lecteurFEW/ (accessed 27.05.2020).

[8] https://wordpress.org/.

Georges = Georges, Heinrich 1913–1918: *Ausführliches lateinisch-deutsches Handwörterbuch*. Aus den Quellen zusammengetragen und mit besonderer Bezugnahme auf Synonymik und Antiquitäten unter Berücksichtigung der besten Hilfsmittel ausgearbeitet. Unveränderter Nachdruck der achten verbesserten und vermehrten Auflage. 2 Bände. Darmstadt: Wissenschaftliche Buchgesellschaft, 1998 (Reprint). Hannover: Hahnsche Buchhandlung, http://www.zeno.org/georges-1913 (accessed 17.05.2020).

Gilliéron, Jules and Edmont, Edmond 1902–1910: *Atlas linguistique de la France*. Paris: Champion.

Idiotikon = Schweizerisches Idiotikon. *Schweizerdeutsches Wörterbuch*. Basel. http://www.idiotikon.ch/index.php (accessed 26.05.2020).

Jaberg, Karl and Jud, Jakob 1928–1940: *Sprach- und Sachatlas Italiens und der Südschweiz*. Ringier: Zofingen.

Krefeld, Thomas and Lücke, Stephan (eds.) 2014: *VerbaAlpina*. Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit. München: online, https://dx.doi.org/10.5282/verba-alpina.

Krefeld, Thomas and Lücke, Stephan 2018: *Typisierung, Methodologie*, VerbaAlpina-de 20/2, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D202%26letter%3DT%2358.

Krefeld, Thomas and Lücke, Stephan 2020a: *Crowdsourcing, Methodologie*, VerbaAlpina-de 20/, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D202%26letter%3DC%2312.

Krefeld, Thomas and Lücke, Stephan 2020b: *Referenzwörterbücher, Methodologie*, VerbaAlpina-de 20/2, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D202%26letter%3DR%2351.

Krefeld, Thomas 2020: *Sprachen und Sprachfamilien im Alpenraum, Methodologie*, VerbaAlpina-de 20/2, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D202%26letter%3DS%2353.

LSI = Lurà, Franco (ed.) 2004: *Lessico dialettale della Svizzera italana*. Bellinzona: Centro di dialettologia e di etnografia.

Lücke, Stephan and Zacherl, Florian 2020: *Transkriptionsregeln, Methodologie*, VerbaAlpina-de 20/2, https://doi.org/10.5282/verba-alpina?urlappend=%3Fpage_id%3D493%26db%3D202%26letter%3DT%23150.

Shneiderman, Ben 1996: *The eyes have it: a task by data type taxonomy for information visualizations*. Proceedings 1996 IEEE Symposium on Visual Languages, 336–343.

SSKJ = *Inštitut za slovenski jezik Frana Ramovša: Slovar slovenskega knjižnega jezika*, http://www.fran.si/130/sskj-slovar-slovenskega-knjiznega-jezika (accessed 28.05.2021).

TLFi = *Trésor de la langue française informatisé*, http://atilf.atilf.fr/tlf.htm (accessed 19/05/2021).

TLG = Brunner, Theodore (ed.) 1972-, *Thesaurus Linguae Graecae*. Irvine: University of California, http://stephanus.tlg.uci.edu/ (accessed 24/05/2021).

Treccani = *Vocabolario Treccani online*, https://www.treccani.it/vocabolario/ (accessed 19/05/2021).

WBOE = Bauer, Werner, Kranzmayer, Eberhard and Institut für österreichische Dialekt- und Namenlexika (eds.) 1970: *Wörterbuch der bairischen Mundarten in Österreich*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Wilkinson, Mark and Dumontier, Michel et al. 2016: *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific Data 3, https://www.nature.com/articles/sdata201618 (accessed 25.05.2021).