

Precise Annotation of Questionnaires for Dialect Research: The Bavarian Dictionary and its Digitization

Manuel Raaf

Bavarian Academy of Sciences and Humanities, Munich, Germany

E-mail: raaf@badw.de

Abstract

The goal of this paper is to present a new web-based application for the analysis of scanned paper-based questionnaires that serve as the basis of the Bavarian Dictionary, by offering project-specific semi-automatic functions which relate and assign image snippets to lemmas. The main requirement was to create software capable of handling several hundred thousand pages containing examples of dialect expressions in several million single parts of images. Additionally, in order to underline the expandability of the software, the requirements of the Franconian Dictionary are briefly described.

Since standard techniques for elaborating dictionary articles or, at least, the preparatory steps to that end, did not fit the particular needs of the project, after digitalization of the questionnaire the web-application LexHelfer was developed. The focus of the application is to aid the editors in the process of gathering relevant information for the lemma entries.

The short history of the application's development and use to date are fully described, so that readers have a complete overview and understanding both of the special type of data and the workflow for creating the entries of the dialect dictionary.

Keywords: annotation; dialect research; image-text-relation; lexicography; questionnaires

1. Introduction

In this paper, in order to provide a complete overview of the digitization of the Bavarian Dictionary and the development of the application, all of the stages are described, beginning with a brief history of the project. The main part of Section 2 then focuses on the application that has been developed for viewing and analyzing the collected material and assisting in compiling the lemma entries. There follow notes on sustainability and basic technical details.

1.1 The Bavarian Dictionary

In 1816, Johann Andreas Schmeller (Frommann: 1872–1877) began work on the first edition of the Bavarian Dictionary, completed in 1837. It was the first scientific work about the Bavarian Language and received extraordinarily high praise by Jacob Grimm (1854: XVII) and others. The second edition, edited by Georg Karl Frommann and published between 1872 and 1877, remains a standard reference work on Bavarian.

The collection of data for the new Bavarian¹ Dictionary has been ongoing since 1913 in order to provide a dictionary that extends beyond Schmeller. The new project also aimed, in cooperation with the Austrian Academy of Sciences, to include the Bavarian varieties spoken in Austria. At that time, today's technical possibilities were unthinkable; there was no thought of a digital strategy, the plan was to create a classical printed dictionary.

In 1961, the Austrian and Bavarian projects, both still ongoing, parted ways. From the mid-80's, the analysis of the material and the writing of lemma entries began, parallel to a survey using questionnaires to collect further information. Up until April 2016, the questionnaires remained almost entirely paper-based and filled out by hand. They were also analyzed by hand, i.e. by choosing the particular individual questionnaire archived in one of dozens of boxes (Figure 1). This meant that the linguists had to review several hundred questionnaires manually in order to determine the variants of the lemma in question. The lemma entries (i.e. the dictionary articles) were subsequently written in MS Word, summarizing all the information gathered on notepads, and thus also paper-based. There now are more than 450,000 paper pages from 240 questionnaires, containing in all over 6,000,000 individual documentations of dialectal expressions. These were digitized to 1.2 Terabytes of images.



Figure 1: paper-based questionnaires (© 2017 Bavarian Academy of Sciences and Humanities)

¹ Bavarian is a German dialect that consists of three main variants. It is spoken mainly in the State of Bavaria, Germany, and in Austria, by about 13 million speakers. The project “Bavarian Dictionary” at the Bavarian Academy of Sciences and Humanities focuses on the varieties spoken in Germany. Despite its current relatively large number of speakers, Bavarian is, according to UNESCO, a “vulnerable” language (Moseley: www.unesco.org/culture/languages-atlas/en/atlasmap/language-id-1019.html).

To directly access this large amount of image data for analysis and then use it to write the dictionary, the web-based application *LexHelfer* has been developed, which, among other features, is able to semi-automatically generate relations between questions in the questionnaires and the particular dialect expressions given as answers, enabling quick manual assignment of lemmas. Due to the specific procedure of drafting entries for the Bavarian Dictionary by referring to and analyzing handwritten questionnaires, its digitalization is very different from other methods of drafting digital and/or digitally-based dictionaries—both from the outset and in the concrete steps taken in working with the material.

Editors will soon be able to not only acribicly compile all occurrences of dialect words but also to automatically reduce this information to only the most important cases needed to create the print-version of the entry. In comparison to the manual practices previously used, both actions take only about one eighth of the time.

1.2 Digitization

After preparation of the questionnaires by student assistants, e.g. removal of attachments like nails, photographs or flowers contributed by the informant as clarifications for the dialect word, a service contractor² scanned the 450,000 A4 format pages to JPEG files at 300dpi. Because of the semantic structure of the file names (i.e. wordlist, place, region, informant, page number), which was set by hand by a service contractor in Vietnam with a very low error rate, each scan could be linked to the informant. More importantly, the information about place and administrative region could be directly extracted, entered into the database and used by the program to restrict the search to specific regions and locations. The creation of maps showing the distribution of a lemma with the aid of *SprachGIS* (Schmidt et al., 2008) is also based on this structure. A script parsed through the scanned images, inserting their information into the database.

Because the majority of the questionnaires were filled in by hand by hundreds of informants (i.e. too many different handwritings), OCR-techniques would not have been feasible for machine-aided text extraction of the scanned images.

As of April 2017, the informants have the option of filling the questionnaires in online, so that the material is digitally native and can be directly imported into the application. Though most informants are quite aged, this new method of gathering the desired information has so far been very well accepted; by almost 20%.

² The material was scanned by MFM Hofmaier (<http://www.mfm.de>); the file names were set by double-keying in Vietnam on behalf of MFM Hofmaier.

1.3 The Franconian Dictionary

The Franconian Dictionary was also founded, and is hosted by, the Bavarian Academy of Sciences and Humanities. It too is based upon questionnaires that contain examples of dialect expressions. However, the workflow is different: every example is gathered in an Excel file that contains only examples of one specific base lemma. Because of the precision in gathering all the information, each row refers to an image file showing the scanned original questionnaire. By importing the Excel contents into the database and adding table handling to the application, the researchers of the Franconian Dictionary are also able use *LexHelfer* and save time in preparing the material for dissemination to the scientific community as well as to the public.³ Thus, the Franconian Dictionary and its specific needs will also be part of the following description.

2. The Application: LexHelfer

LexHelfer is designed as a writing assistant for dictionary entries and as a search tool for researchers working on the dialect. Furthermore, the public is also able to search, for example, for expressions documented in their home region. The public view of the application is an automatically created restricted version of the researchers' full version, so that there is no need for the scientist to adjust contents or deactivate program features.

The application consists of different components for preparing digitized questionnaires for analysis and performing the analysis in order to gather information for creating dictionary entries and/or for acribically collecting and online-publishing language examples. From the start, it was developed in close cooperation with linguists, and with consultation about their requirements. Thus, the application fully respects the editors' scientific needs and demands on usability. Furthermore, it was assumed from the beginning that public users should also be able to search the database intuitively and therefore without the need to read (long) instructions.

2.1 From Image Files to Data Relations

As noted in Section 1.2, the names of the image files have a simple semantic structure: wordlist number, place, administrative region, informant ID, and page number. By creating a database table that fits this structure and reading all files into it, the application has all the information needed to point directly to the desired questionnaire and also to restrict searches to specific areas. The particular select-boxes of the HTML-formula for searching were created by a simple and grouped SQL-request

³ The Franconian Dictionary will not result in a classical dictionary, but rather in a collection of examples linked to standard lemmas enriched with grammatical information and comments.

on the rows containing the information about place and region. Thus, the process can be automated, avoiding the programmer having to type hundreds of place names. Furthermore, it permits the setting of a one-to-one relationship between a lemma and the specific source of examples after performing the actions described in the following sub-section.

2.2 Semi-Automatic Creation of Snippets

Each of the 240 questionnaires contains 60 questions on four pages. This does not result in just 240×60 coordinates of questions/answers, because each questionnaire was distributed to some hundred places all over the State of Bavaria and completed in each place by between one and a dozen informants. Hence, the amount of potential relevant material totals more than 6,000,000 coordinates. Handling so many examples by hand is not feasible in an acceptable period. However, since on each of the 240 questionnaires there is a fixed position for every question and its answer, a simple tool could be created that would allow student assistants to capture these positions and store their coordinates together with the number of the particular question in the database. It was not deemed necessary to choose a specific best-suited copy of a questionnaire to represent all other copies, so the decision of simply taking the first one in the directory of the file system came very quickly. In the final analysis of the results, this decision was spot-on. A script subsequently used the gathered coordinates and created database entries for every single occurrence. Later, the coordinates could be adjusted in the application by researchers in cases where the informant entered information above, below or right next to the designated answer area.⁴

Coding and using the tool for the semi-automatic snippet-creation took about four days, which is a modicum of time compared to manually capturing over six million coordinates. Another day was needed to capture the personal information of the informants, i.e. the address dates, with due respect for data privacy. These areas were then overwritten for the public view.

2.3 Status of a Snippet: None, Irrelevant, Finished

2.3.1 None

The inherent status of each snippet is “none”, i.e. the snippet—or better: its content—is not yet reviewed. Snippets having the status “none” will always occur in the search results.

⁴ Please see the appendix for examples of a questionnaire and its semi-automatically created snippets.

2.3.2 Irrelevant

Fortunately, there are many valuable examples. Nonetheless, there are also many impractical ones, i.e. empty answers, which slow the researchers down when they are analysing the given examples because they interrupt the browsing and analysis of the material. To avoid this, each snippet can be set to “irrelevant” and will not then appear in the list of search results unless irrelevant snippets are explicitly demanded on loading. Filtering out impractical snippets can be undertaken quite quickly by student assistants, so that costs are low.

2.3.3 Finished

After finishing the work on a snippet by relating at least one lemma to it, the status “finished” excludes the snippet from (re-)occurring in the search results unless explicitly so defined. Hence, reviewing the contents of the snippets will result in fewer search results after each round, each of which is usually concentrated to a specific region.

2.4 Displaying Evidences and Relating them to Lemmas

The workflow of the Bavarian Dictionary generally starts by choosing a specific question from a wordlist questionnaire, with or without filtering of the search results by place name or administrative region. The snippets then appear ten to the page, surrounded by several action handlers for changing their status, assigning a lemma, or loading the full scan of the corresponding page. In addition, the clip of the snippet can be scrolled up or down by mouse in order to look at information that is potentially written beyond the designated position of the question.

Viewing the full page and choosing a lemma opens a new internal window in which the content in question appears. In this view, one is able to modify the snippet as well as the image file itself, e.g., in the few cases where the scan was made upside-down, by rotating the image. For quick assignment to an existing lemma within the same questionnaire, a drop down menu (<select>) is presented in order to relate the snippet to a lemma in only two clicks.

2.5 Table Handling for the Franconian Dictionary

Table handling is currently used only by the Franconian Dictionary. It allows researchers to modify the content of columns in order to perform error correction faster than in Excel, since the server is more powerful than a typical workstation computer. Sorting 5000 rows or changing their content in Excel might even lead to crashing the system, whereas the database server executes both actions with ease.

Moreover, the table contents can be worked on collaboratively and there is no need for colleagues to first finish and save their changes to the previously used single Excel file.

However, there still are limits relative to the workstation used in accessing the application: the lower the performance, the higher the risk of receiving a blank HTML page as a result of a request for several thousand rows. Fortunately, this only affects reception of the output and not the execution of a server-side command, e.g. changing the content of a field based on that of another field for thousands of entries.

2.6 Export of Search Results

For the Bavarian Dictionary, the search results (list of examples for a particular chosen lemma or list of search results of the concrete search requested by user input) can be downloaded as a PDF file. This contains the snippets and the information about the related lemma (if present) as well as the information extracted from the semantic structure of the particular file name. This function was requested in order to provide a printable and storable list of lemmas and their examples. The same is true of a complete list of examples for a particular chosen lemma.

The Franconian Dictionary yet does not need this functionality; instead, it offers a download of the table of search results as a CSV file.

2.7 Subject Categories (Basic Type of Ontology)

Many dictionary users are interested not only in the meaning and distribution of a (dialect) expression but also in superordinate categories. For instance, before official holidays and celebrations of local traditions, people came to the researchers with questions like “are there any, maybe little-known, dialect words for Christmas in the region of ...?”. From time to time, journalists of local newspapers have similar interests. As answering such requests must not take too much of the lexicographer’s time, the solution could be hierarchical categorization of the examples in semantic groups for speedy reference so that a quick search and potential additional examples would satisfy all needs.

A positive side effect of a reliable link between a lemma and a semantic group would be the possibility of adding images automatically to the dictionary. Bearing in mind the widely held opinion that younger generations have less and less knowledge about their dialect, this could be exploited by using new media in preschools to spark off children’s interest and qualification. Moreover, such a link would be a good basis for enabling complete accessibility to the content of the dictionary by enriching the database with additionally linked sound files, which would help disabled people to have full access to the evidences by audio- and image-based indication.

Because there is little scientific work on German ontology and none regarding dialect expressions, this task was resolved using the “subject groups” from the “Badisches Wörterbuch” (Dictionary of Baden Dialects) as well as in the “Pfälzisches Wörterbuch” (Rhenish Palatinate Dictionary), which base their groups upon those of Hallig and von Wartburg (1963). This material is not yet printed or published online. We thankfully received a legal copy of the groups and corresponding lemmas from the Rhenish Palatinate Dictionary by courtesy of Rudolf Post and matched the material with the basic form and the meaning of the dialect examples of the Franconian Dictionary. Where no direct hit could be achieved, the Levenshtein algorithm was used to calculate the shortest distance between two strings, where one is the Franconian evidence (in two rounds: first the basic form, second the meaning) and the other the entry of the aforementioned list of subject groups from the Rhenish Palatinate Dictionary. With the aid of some basic normalizations on the strings (i.e. deletion of articles, pronouns, conjunctions, punctuation, spaces, converting all to lower-case) and accounting for the possibility of attributive metathesis (e.g. *young beautiful woman* vs. *beautiful young woman* as meaning of the example), the error-rate could be reduced a little. However, there are still a very great number of erroneous results, because the dataset available for this task is not yet sufficiently large.

Controlling and correcting the results of the automatic assignment by one student assistant is ongoing and will probably be completed by the end of 2017. The method in its two variants⁵ and the particular results are also part of a Bachelor thesis at the Chair of Computer Science and Modern German Literary History at the University of Würzburg, Germany, in cooperation with the Chair of German Linguistics at the University of Erlangen-Nuremberg, Germany. The issue is which method is better, or if both are equally good or bad depending upon the present dataset.

With time, the task will be completed with fewer incorrect results, because the amount of data will increase with each involved dictionary and thus the script will have more clear hits. The long term aim is to create a widespread list of German subject groups that is based on dialects spoken in the State of Bavaria, not on Standard High German. The list could also be used for other dictionaries.

2.8 Lists

LexHelfer offers the possibility of listing all entries by different options in order to produce an overview of what has already been gathered. The options are: listing in

⁵ Basically, using Levenshtein as described above. However, there are some possibilities for preferring the strings for comparison: the dialect expression always has the field “meaning” and “basic form” (i.e. basic lemma). For the first option, the meaning serves as first comparative value. The second option can then be either every meaning within the loop through all the meanings or the loop of only the search results for the meaning or (logical *or* in SQL) the basic form in the list of the Rhenish Palatinate Dictionary and its subject groups.

alphabetical order (lemma or basic form; soon: subject group), wordlist and question number, grammatical annotation (for the Franconian Dictionary only). For example, the list of grammatical annotations helps researchers to verify whether the annotation is already present, and how it is written. Every entry in a list links to the page containing the particular search results (e.g. grammatical annotation for “Art AkkSgN partitive”⁶ lists all entries for the article (*Art*) in case accusative (*Akk*) singular (*Sg*) neuter (*N*)). Rare constructions and/or rare uses of grammatical structures can thereby be found quite easily. Linguists can thus describe uncommon cases by browsing the list. This is also a good method to recognize typos, e.g. “AdjSgM” instead of the correct notation “Adj SgM” (i.e. adjective singular masculine).

2.9 Upcoming: Compiling Fascicles

To date, researchers write the lemma entries in MS Word. Changing the current workflow to an XML-base is time-consuming because the structure needs to be created in a precise and user-friendly fashion, so that linguists only need a short time for the migration. The plan is to integrate this process into *LexHelfer*, thereby allowing the researchers to prepare the material and directly import it into the function for compiling lemma entries. After reducing the information, the entry can be finished and directly published online or printed (alone or with other entries) to PDF for further processes by the publisher. As of May 2017, the XML-structure is being created in close consultation with the researchers.

3. Sustainability

The questionnaires are present in high-resolution JPEG files, one scan per page, and saved on a highly professional long-time archive system at the governmental Leibniz-Rechenzentrum⁷, Garching, Germany, which is part of the Bavarian Academy of Sciences and Humanities. The decision about what format to use was easily taken in favour of compressed JPEG instead of (lossless) TIFF or PNG because of the comparatively low quality of the sources: the questionnaires consist of yellowish paper containing black typewriter font and answers written in either blue, grey (lead pencil), or black. Hence, important visible image information loss by choosing a lossy format

⁶ “Art AkkSgN partitive” stands for: article (*Art*) in the case accusative (*Akk*) partitive (partitive), number singular (*Sg*), gender neuter (*N*). The case *partitive* is not present in German nor in Germanic languages as morphological case. However, it is in the semantics, since atelic actions might force the object of the action to be in a partitive state (depending on the context). For example, the phrase “I’m drinking tea” in the sense of “I’m drinking tea sip by sip” is *partitive* from semantic point of view, whereas “I’m drinking tea” in the sense of “I will completely gulp this cup of tea right now” is not.

⁷ www.lrz.de

was not to be expected. The possibility of losing not only visible information but the whole file due to bit-errors⁸ was considered, but neglected for two reasons: 1) The total amount of disk capacity needed for storing around 450,000 files in high-quality JPEG is 1.2 Terabytes, whereas in an uncompressed format it would be at least ten times more. The service contractor also recommended JPEG, with regard to the large amount of data. It is true that high quality scans stored to an uncompressed format would be best for long-term archiving and for reducing the risk of information loss caused by bit errors. However, the compromise with JPEG was regarded as the best choice for the project. 2) The original questionnaires could be re-scanned without much effort, if necessary.

The database containing all the data of the application and thereby the digital base of the dictionary is stored in a MySQL-database. Since MySQL is an open and widely used database management system that can be converted to another (standard) format (e.g. XML), and since JPEG can be read on any device and be converted to any other image format, the data will be available and convertible for a long time. Although technological changes usually happen quickly, the technological base behind the changes does not. Therefore, the current formats will guarantee technical sustainability better than proprietary or less common ones.

4. Technique

LexHelfer runs under Apache 2.4 on a virtual machine running a modern Linux distribution. On the server, PHP 5.6 is the primary language in the scripts handling requests and creating the HTML-output for the client. The latter then uses pure JavaScript (i.e. no libraries) for modifying the output and preparing the requests. The scripts for importing the information about the scanned image files and for the creation of the snippet coordinates for each single occurrence are written in Perl 5. As database, the latest free open-source version of MySQL 5 is used.

On the server, PHP and Perl were used because of the higher power of Perl in processing text information, especially with the aid of regular expressions, which are built-in in Perl. Nevertheless, all actions performed by Perl-scripts could also be performed by PHP-scripts.

Following the open-access-policy of the Bavarian Academy, the source code is licensed under ALv2.⁹ After expanding the application in order to have a single one that can be configured by authorized users via the integrated administration panel, e.g. having the

⁸ In compressed formats, even a single error-bit can lead to a total and unrecoverable loss of all the information of the file. In uncompressed formats, on the other hand, a destroyed bit would not lead to other bits also being defective. As yet, there is no such case known in the project.

⁹ Apache License Version 2, see <http://www.apache.org/licenses/LICENSE-2.0>.

focus set to the tables (Franconian) or lemma-snippet-assignment (Bavarian), the code of *LexHelfer* will be available via Git. Hence, the direct usage as well as forking the program will be possible and appreciated, since this would transform the application from its status as an in-house solution to a more generic one. The community of users would create synergies and hopefully benefit greatly from it.

5. Public Effectiveness

Government funded projects must always serve not only the scientific community, but also the public. Since a public version of *LexHelfer* is available on-the-fly without researchers having to unlock the contents, anybody is able to access the material on the website.¹⁰ From February 1st until May 16st 2017, around 13,000 public requests were recorded¹¹—for a project that is not yet commonly known and that targets a relatively small number of prospective users, this is a relatively high rate of requests by the public.

In addition, the Franconian Dictionary is used by students at the Universities of Erlangen-Nuremberg and Würzburg and has registered about 45,000 cases of public access in the period from January 1st to May 16st 2017.

6. Summary

The schedule of digitizing the Bavarian and Franconian Dictionaries was maintained throughout the whole process. Despite some minor problems, which are to be expected in any project, all went well, thanks to the constant close contact with researchers. Without this, development of the application to meet the scientists' needs would have taken much longer. Scanning the paper-based examples and developing the application have both had a timesaving effect on the linguists' work. The additional financial expense compared to zero expense options, i.e. not going digital, was completely worth the effort, and the additional expenditure was mostly for the scanning of several hundred thousand questionnaires. The speed of gathering all available information for one specific question/lemma has increased about eight fold compared to the analogue manual work.

The options regarding further development¹² are promising and will gradually be instigated in coming years.

¹⁰ Database of the Bavarian Dictionary: <http://www.bwb.badw.de/bwb-digital/datenbank>.
Database of the Franconian Dictionary: <http://www.wbf.badw.de/wbf-digital/wörterbuch>.

¹¹ In respect of data privacy, only the action of requesting data is recorded/counted, not the request itself or any user-specific data.

¹² Among others: compiling fascicles in XML, improving the automation on relating examples to subject groups, adding images as well as sound files spoken by native speakers to the subject groups as described in Section 2.7

7. References

- Christmann, Ernst & Krämer, Julius (ed.) (1965 – 1997): *Pfälzisches Wörterbuch*. Steiner. Wiesbaden.
- Fromman, Georg Karl (ed.) (1872 – 1877): *Bayerisches Wörterbuch von J. Andreas Schmeller*. Oldenbourg. Munich. Accessed at <http://daten.digital-sammlungen.de/~db/bsb00005027/images/> (19.05.2017)
- Grimm, Jacob (1854): *Deutsches Wörterbuch. Band 1. A – Biermolke*. Hirzel. Leipzig
- Hallig, Rudolf & von Wartburg, Walther (1963): *Begriffssystem als Grundlage für die Lexikographie*. Akademie-Verlag. Berlin
- Moseley, Christopher (ed.) (2010): *Atlas of the World's Languages in Danger*, 3rd edition. UNESCO Publishing. Paris. Accessed at <http://www.unesco.org/culture/en/endangeredlanguages/atlas> (18.01.2017)
- Ochs, Ernst & Müller, Karl Friedrich & Post, Rudolf et al. (eds.) (1925–today): *Badisches Wörterbuch*. Oldenbourg. Lahr/Munich
- Raaf, Manuel (2016–today): *LexHelfer BWB*. Bavarian Academy of Sciences and Humanities. Munich. Accessed at <http://www.bwb.badw.de/en/digital-platform/database> (16.05.2017)
- Rowley, Anthony et al. (eds.) (1995–today): *Bayerisches Wörterbuch*. Oldenbourg. Munich.
- Schmidt, Jürgen Erich & Herrgen, Joachim & Kehrein, Roland (eds.) (2008 ff): *Regionalsprache.de (REDE). Forschungsplattform zu den modernen Regionalsprachen des Deutschen. Teil 6: REDE SprachGIS – Das forschungszentrierte sprachgeographische Informationssystem von Regionalsprache.de*. edited by Dennis Bock et al. Forschungszentrum Deutscher Sprachatlas. Marburg. Accessed at <https://www.regionalsprache.de> (18.01.2017)

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>



8. Appendix

8.1 Example of a Questionnaire and its Snippets

Besonders wichtig: Bei Hauptwörtern das Geschlecht (der, die, das) angeben.

85/13. Ist Ihnen der Ausdruck **Fiederling** für a) Mensch mit schwächtiger Figur; b) Getreide mit wenig Körnern oder in anderer Bedeutung bekannt? Bezeichnen Sie einen schwächtigen Menschen als **Fiederer**? Bitte jeweils genaue Aussprache und Bedeutungsangabe.

nein

85/14. Wird Fleisch von Kälbern, Schweinen, Schafen als **junges** oder **kleines Fleisch**, das Fleisch von Stieren, Ochsen, Kühen als **altes** oder **großes Fleisch** bezeichnet? Bitte Satzbeispiele.

nein

85/15. Verwenden Sie den Ausdruck **vom Fleisch fallen** für „mager werden“? Bitte Satzbeispiel.

keins: Des is ja direkt vom Fleisch gfallen.

85/16. Kennen Sie den Ausdruck **Baumfleisch** für ein aus Knödeln und gekochten Zwetschgen bestehendes Gericht, das am Gründonnerstag gegessen wird, bzw. wurde? Bitte genaue Sachangaben, Satzbeispiel.

nein

85/17. Sind Ihnen die Ausdrücke **Flachtl, Flackel** o. ä. für „Fläschchen; kleines Trinkgefäß“ bekannt? Bitte Aussprache, Satzbeispiel, Sachangaben, evtl. mit Skizze.

nein

85/18. Kennen Sie die Redewendung **er hat Kniefleisch** für „er ist neidig“? Bitte Satzbeispiel.

nein

85/19. Ist Ihnen der Ausdruck **Kenderfleisch** o. ä. für „Geräuchertes“ bekannt? Bitte genaue Aussprache und Satzbeispiel.

nein

85/20. Sagen Sie **Luderfleisch, Schadenfleisch** für „schlechtes, minderwertiges Fleisch“? Welche Bezeichnungen sind Ihnen sonst für minderwertiges, etwa für die Freibank bestimmtes Fleisch bekannt, z. B. Ringelfleisch?

*„billiges Fleisch“ gibt es, da wird eine Kuh ausk'haust.
(ausgehauen.)*

85/21. Kennen Sie den Ausdruck **Riemenfleisch** für „Schlegelfleisch vom Rind“? Bitte Aussprache und Satzbeispiel.

nein

85/22. Ist Ihnen der Ausdruck **Schupfleisch** o. ä. für „Fleisch zum Würstmachen“ oder in anderer Bedeutung bekannt? Bitte Aussprache, Satzbeispiel und Sachangaben.

nein

85/23. Sagen Sie **fleischenen, fleischeinen** o. ä. für „nach Fleisch riechen“? Bitte genaue Aussprache und Satzbeispiel.

*Dös riecht so fleischon.
Dös schaut so fleischon aus.*

85/24. Kennen Sie den Ausdruck **der fleischt sich an** für „er nimmt zu“? Was bedeutet es, wenn man von einer Kuh sagt: **sie ist angefleischt**? Bitte jeweils Satzbeispiele.

nein

85/25. Ist Ihnen der Ausdruck **Flascher** (Floascher o. ä.) für einen schlecht wirtschaftenden Bauern bekannt? Bitte Aussprache und Satzbeispiel.

nein

85/26. Kennen Sie die Ausdrücke **auf z'Fleiß, auf b'Fleiß, mit Fleiß, mit b'Fleiß** o. ä. für „absichtlich“? Bitte genaue Aussprache und Satzbeispiele.

*Sie hät eahm mit b'Flaiß
af 'n Kopf afe'k'haust.*

85/27. Ist Ihnen der Ausdruck **fenzeln, fenscheln** o. ä. für „schmeicheln, zärtlich sprechen, lächeln“ oder in anderer Bedeutung bekannt? Bitte Aussprache, genaue Bedeutung und Satzbeispiel.

nein

85/28. Ist Ihnen **fenschen** (flea'schn o. ä.) für a) weinen, b) das Gesicht verziehen, c) die Zähne fletschen oder in anderer Bedeutung bekannt? Bitte Satzbeispiel.

*Gestern hät o den ganzen
Tag gflechtscht (gerämt).*

Figure 2: second page of one copy of questionnaire 85

85/13. Ist Ihnen der Ausdruck **Fiederling** für a) Mensch mit schwächlicher Figur; b) Getreide mit wenig Körnern oder in anderer Bedeutung bekannt? Bezeichnen Sie einen schwächlichen Menschen als **Fiederer**? Bitte jeweils genaue Aussprache und Bedeutungsangabe.

nein

85/14. Wird Fleisch von Kälbern, Schweinen, Schafen als **junges** oder **kleines Fleisch**, das Fleisch von Stieren, Ochsen, Kühen als **alles** oder **großes Fleisch** bezeichnet? Bitte Satzbeispiele.

nein

85/15. Verwenden Sie den Ausdruck **vom Fleisch fallen** für „mager werden“? Bitte Satzbeispiel.

Kils: Der is ja direkt vom Fleisch gfallen.

85/16. Kennen Sie den Ausdruck **Baumfleisch** für ein aus Knödeln und gekochten Zweischnen bestehendes Gericht, das am Gründonnerstag gegessen wird, bzw. wurde? Bitte genaue Sachangaben, Satzbeispiel.

nein

85/17. Sind Ihnen die Ausdrücke **Flachtl**, **Flackel** o. ä. für „Fläschchen, kleines Trinkgefäß“ bekannt? Bitte Aussprache, Satzbeispiel, Sachangaben, evtl. mit Skizze.

nein

85/18. Kennen Sie die Redewendung **er hat Kniefleisch** für „er ist neidig“? Bitte Satzbeispiel.

nein

85/19. Ist Ihnen der Ausdruck **Kenderfleisch** o. ä. für „Geräucheretes“ bekannt? Bitte genaue Aussprache und Satzbeispiel.

nein

85/20. Sagen Sie **Luderfleisch**, **Schadenfleisch** für „schlechtes, minderwertiges Fleisch“? Welche Bezeichnungen sind Ihnen sonst für minderwertiges, etwa für die Freibank bestimmtes Fleisch bekannt, z. B. Ringelfleisch?

*„billiges Fleisch“ gibt es, da wird eine Kuh ausk'haust.
(auszshamm.)*

85/21. Kennen Sie den Ausdruck **Riemenfleisch** für „Schleppfleisch vom Rind“? Bitte Aussprache und Satzbeispiel.

nein

85/22. Ist Ihnen der Ausdruck **Schupffleisch** o. ä. für „Fleisch zum Würstmachen“ oder in anderer Bedeutung bekannt? Bitte Aussprache, Satzbeispiel und Sachangaben.

nein

85/23. Sagen Sie **fleischenen**, **fleischenen** o. ä. für „nach Fleisch riechen“? Bitte genaue Aussprache und Satzbeispiel.

*Pö's riecht so fleischp'n.
Dös schaut so fleischp'n aus.*

85/24. Kennen Sie den Ausdruck **der fleischt sich an** für „er nimmt zu“? Was bedeutet es, wenn man von einer Kuh sagt: **sie ist angefleischt**? Bitte jeweils Satzbeispiele.

nein

85/25. Ist Ihnen der Ausdruck **Flaischer** (Floascher o. ä.) für einen schlecht wirtschaftenden Bauern bekannt? Bitte Aussprache und Satzbeispiel.

nein

85/26. Kennen Sie die Ausdrücke **auf z'Fleiß**, **auf b'Fleiß**, **mit Fleiß**, **mit b'Fleiß** o. ä. für „absichtlich“? Bitte genaue Aussprache und Satzbeispiele.

Sie hät eahm mit b'Fleiß auf 'n Kopf äfe'k'haust.

85/27. Ist Ihnen der Ausdruck **fenzeln**, **fenscheln** o. ä. für „schmeicheln, zärtlich sprechen, lächeln“ oder in anderer Bedeutung bekannt? Bitte Aussprache, genaue Bedeutung und Satzbeispiel.

nein

85/28. Ist Ihnen **fensch** (flea'schn o. ä.) für a) weinen, b) das Gesicht verziehen, c) die Zähne fletschen oder in anderer Bedeutung bekannt? Bitte Satzbeispiel.

Gestern hät o den ganzen Tag gfleimscht (geweint).

Figure 3: semi-automatically created snippets of one copy of questionnaire 85

8.2 Screenshot of Search Results (Bavarian Dictionary)

Figure 4: main view of the search results for questionnaire 54. Due to limited space, only half of it is visible in this screenshot

8.3 Screenshot of Search Results (Franconian Dictionary, Table View)

Ihre Suche nach Grundform *Haus** ergab 55 Treffer.

Stamm	Grundform	Bedeutung	Grammatik	Originaltext	Umschrift	Ort	Planquadrat	Kommentar	Gewähsperson	Kommentar	Bearbeiter	GP	Bogen	Frage	Bild
**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**
Arrest	Hausarrest	Freiplatz beim Fangenspiel	S NomSg	Hausarrest	Hausarrest	Bühler	V26,7					1	5	31	
Bock	Hausbock	Insekt, Holzschilling	Sm NomSg	Hausbock	Hausbock	Flachlanden	d30,1	Holzschilling				2	114	7	
Ecke	Hauseck	Außenkante des Hauses (in Redensart)	Sn NomSg	der ründ's Hauseck o	der ründet das Hauseck an	Forth	a34,8			Gesamtbedeutung: urinieren		1	69	35	
Geld	Hausherausgeld	Geldzahlung des Erben an Eltern bei Hofübergabe	Sn NomSg	das Hauseraus geld	das Hausherausgeld	Heinrichthal	V23,1					1	7	47	
halten	Haushalten	führen eines Haushalts	Sn AkkSg	versorcht dös Haushälten	versorgt das Haushalten	Edlesdorf	T37,2	jedoch ohne das Erbe im Auge zu haben				1	8	34	
halten	haushalten	sparen, Geld nicht verschwenden	Vt PP	haushalten haouts und haouts za wos bracht	haushalten hat er und hat es zu was gebracht	Vordorf	V38,5			#evtl verschr		1	56	33	
halten	haushälterisch	sparsam, geizig	Adj peid	haushälterisch	haushälterisch	Rothensburg o.d.T.	d28,4					1	69	2	
halten	haushälterich	sparsam, geizig	Adj peid	haushäl(er)lich	haushälterich	Hürbel a. R.	d29,9					1	69	2	
haun	Haupan	Fällkerbe in Baumstamm schlagen	Sm NomSg	Ha-upan	Haupan	Röttenbach	g33,1			zweites a ist unterstrichen		1	1	50	
Haus	Haus	Anrede unter Verliebten	Sn NomSg	aldis Haus	altes Haus	Schwemmelbach	V27,1					1	116	11	

Figure 5: search results for the base form “Haus” (house), extended by the asterisk * to encompass compounds in the results