Dieter W. HALWACHS / Barbara SCHRAMMEL / Astrid RADER

ROMLEX
THE LEXICAL DATABASE OF ROMANI VARIETIES

ROMLEX is not a Romani dictionary in the usual sense, it is a lexical database. It contains data that are representative of the variation in the lexicon of all Romani dialects, and offers almost complete coverage of the basic lexicon of the Romani language. At present, data are available online covering 25 different Romani dialects, see: http://romani.uni-graz.at/romlex. These entries resp. glossaries are accompanied by translations into English and, depending on the Romani dialect, into other European languages as well.

After a preparatory phase from 1998 to 2001 which was exclusively dedicated to Romani varieties spoken in Austria, and two working phases from 2001 to 2007 ROMLEX enters the next level of development which is outlined in this article.

1 Research context

There is a long history of linguistic study of Romani in philology. The focus of this era of occupation with Romani was historical reconstruction, especially the origin of Romani and the affiliation of Romani to a language family (Rüdiger 1782, Pott 1844-5, Miklosich 1872-80, Turner 1926), descriptive grammars (Paspati 1870, von Sowa 1887, Finck 1903, Sampson 1926), as well as (mainly comparative) lexicographic studies (Pott 1845, Miklosich 1876, 1877, Sampson 1926).

The second half of the twentieth century saw the rise of modern Romani linguistics as a field of general and applied linguistics. This is reflected by a large number of descriptive grammars (for example Gjederman & Ljungberg 1963, Pobożniak 1964, Wentzel 1980, Boretzky 1993, Holzinger 1993, Matras 1994, Igla 1996, Halwachs 1998, Cech & Heinschink 1999, Tenser 2005), but also by linguistic analyses of certain aspects of the structure of Romani (Matras 1995, Matras & Bakker & Kyuchukov 1997, Elšík & Matras 2000, Schrammel & Halwachs & Ambrosch 2005) and by sociolinguistic studies (Friedman 2003, 2005, Halwachs 2003, 2005, Halwachs & Heinschink 2000, Matras 2004). Furthermore this period sees a considerable number of lexicographic studies (Wolf 1960, Valtonen 1972, Calvet 1982, Vekerdi 1983, Uhlik 1983, Demeter & Demeter & Tcherenkov 1990, Hübschmannová et al. 1991, Boretzky & Igla 1994, Soravia & Fochi 1995, Halwachs et al. 2002). These studies vary to a great extent in coverage and format. Alongside comparative works, the most elaborate being Boretzky & Igla (1994), there are also dictionaries on individual varieties and glossaries, which only display vocabulary of very restricted semantic domains (e.g. vocabulary exclusively compiled from fairy tales such as Calvet 1982). The extent and elaboration of grammatical information provided for individual lexical entries also varies considerably in these works.

These developments in the linguistic study of Romani were paralleled by a beginning self-organisation and emancipation of the Roma community, which led to an increasing text production activity. It marks the first period in which Roma widely use their language for text production that lies far beyond basilectal texts like personal letters and the like. The new text types include: official political documents, newspaper reports, autobiographical reports, fairy tale collections and so on. Due to the lack of any written form of the language up to this point, the need arises to at least minimally codify individual varieties, in order to guarantee comprehensibility of the texts produced.

Such codifications often were established in co-operation of linguists and the speech community in the framework of language projects. One such project is the "Romani Projekt Graz" at the Department of Linguistics of the University of Graz, which was crucial in developing ROMLEX.

Within the framework of this project linguists have been engaged in documenting the oral traditions of the larger Romani-speaking communities of Austria (Sinti, Burgenland-Roma, Lovara, Kalderaš, Gurbet, Arli) and in compiling extensive grammatical and lexical descriptions of these varieties for the past decade. The codification, i.e. implementation of a consistent writing system, of some of these varieties was a by-product of this documentation process.

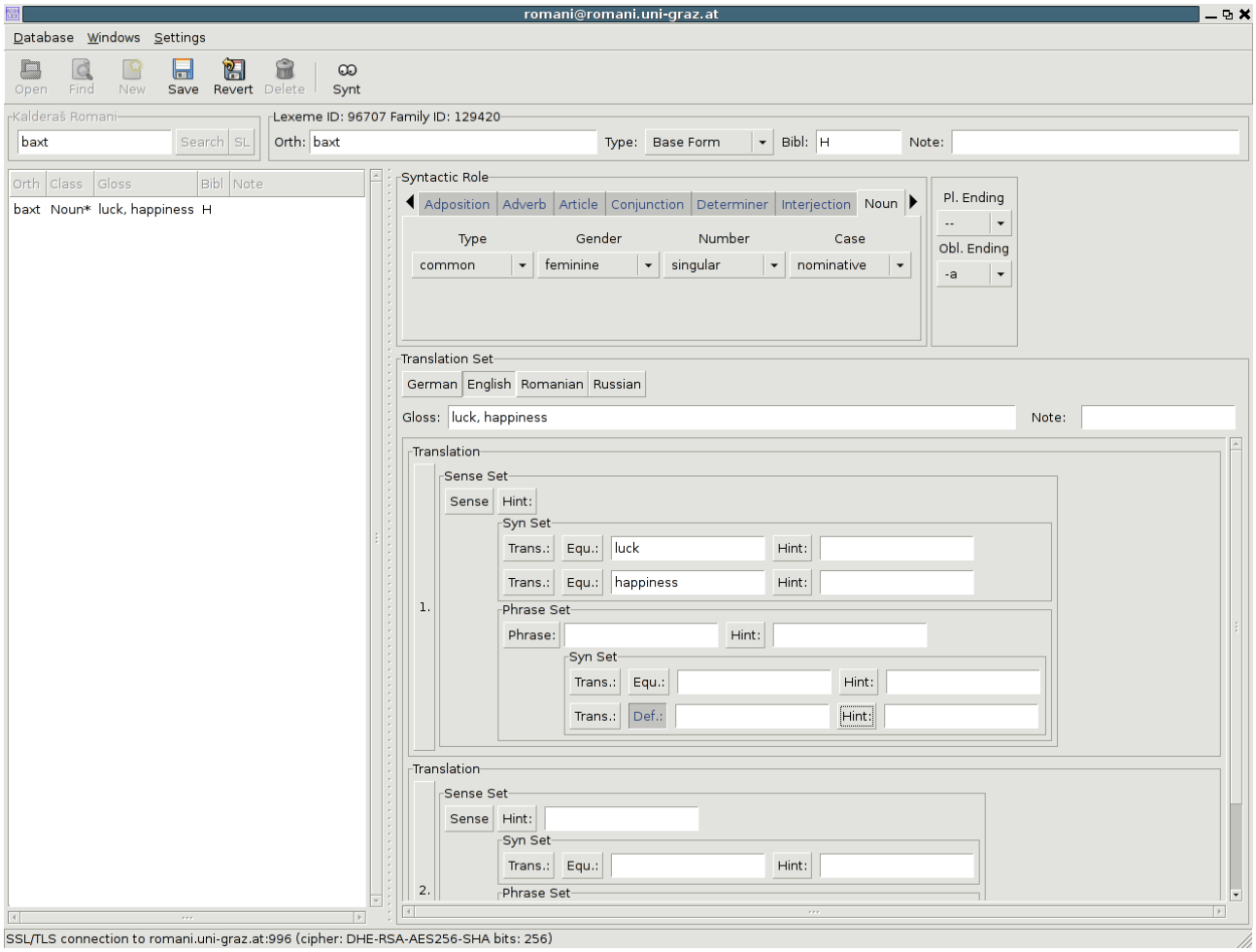
The existence of the Heinschink Collection of Romani oral tradition, going back several decades and covering a great range of Romani varieties from all parts of Europe and the Mediterranean, at the Phonogrammarchiv at the Austrian Academy of Sciences, has provided Austrian linguists with the world's most extensive comparative source of data on Romani. The Heinschink Collection was archived and documented with two grants of the FWF (P 7802-SPR "Sicherung, Dokumentation und Erschließung der Sammlung Heinschink" from July 1990 until August 1992 und P 9083-HIS "Sicherung, Dokumentation und Erschließung der Sammlung Heinschink, Teil 2" from September 1992 until August 1995).

This favourable working setting has resulted in the production of numerous descriptive working papers, journal articles and several monographs and collections outlining the structures of the Romani varieties of Austria (for example Halwachs 1998, Cech & Heinschink 1998, Cech & Heinschink 1999, Halwachs & Menz 1999, Cech & Fennesz-Juhasz & Heinschink 1998, Halwachs & Ambrosch & Wogg 1998, and more). Work on Romani in this context was also pivotal in the development of ROMLEX: the diversity displayed by the Austrian Romani varieties and the Heinschink collection in combination with the often expressed wish of speakers for a dictionary of their language, set a thinking process in motion, which aimed at developing strategies for a type of dictionary that would meet these special conditions. This led to a new concept of lexical documentation, which would reflect intra-variety and inter-variety variation. This finally, was the birth of ROMLEX, which was designed as a lexical documentation tool in the form of a custom-made database, not only for the Austrian Romani context, but for an unrestricted number of Romani varieties. The choice for this format was justified by the many advantages it has compared to a paper version: First, an electronic tool is in principle open for extensions at any time, both with respect to the number of varieties included and the number of individual lexical entries. Second, a database makes it possible to document the diversity of the Romani lexicon without introducing artificial hierarchical structures between lexical variants. Third, a database enables complex cross-variety searches (provided the already available individual lexical entries in ROMLEX will be linked according to a set of parameters), which are much more comfortable to handle, when compared to the use of a very extensive paper volume with cross-references for this purpose.

After three previous project phases (preparatory phase from 1998 to 2000 which was limited to Austrian Romani varieties, phase 1 from 2001 to 2004, funded by the Open Society Institute (OSI), and phase 2 from 2005 to 2007, funded by the OSI and – with respect to the Romani varieties spoken in Austria – by the Austrian Federal Chancellery's "Volksgruppenförderung") at this stage the development and programming of the database is completed. With respect to contents, at present ROMLEX is a collection of 25 electronic glossaries, edited in a unified writing system, with basic morphosyntactic categorisation and translations into various European languages (for details see http://romani.uni-graz.at/romlex/stats.cgi).

The writing system used in ROMLEX is basically phonemic, with a number of exceptions concerning the inclusion of important phonetic characteristics of individual Romani varieties. In encoding the single characters we follow and use the UNICODE standard. The writing system is displayed at http://romani.uni-graz.at/romlex/wsphonemic.xml. Grammatical information is coded according to the conventions of morphosyntactic categorisation developed by the Isle-Eagles-Multext projects, selected and slightly modified as appropriate for Romani. A simplified outline of the use of these conventions for Romani is shown at http://romani.uni-graz.at/romlex/msd.cgi. The

following screenshot gives an impression how morphosyntactic information and translations are displayed in the database:



Lexical coverage of the individual glossaries differs to a great extent with respect to the number of lexical entries contained (compare the numbers in the table below) and with respect to lexical domains that are reflected in the glossaries. This is due to the fact that in the previous working phases ROMLEX glossaries were based on very different sources, such as more or less elaborated dictionaries, corpuses from non-scientific text production, or fairy tale collections. This means that while some of the 25 Romani varieties display vocabulary from a wide range of lexical domains (e.g. Burgenland Romani or Kalderaš Romani), others are more restricted and might display vocabulary typical of a specific text sample. Thus, while the individual 25 varieties are already homogenised with respect to codification, i.e. the writing system implemented, lexical coverage in ROMLEX at present is very heterogeneous and makes cross-variety comparison a difficult matter. It will be one of the core work tasks of the next project phase to homogenise lexical coverage (at least) for a chosen number of Romani varieties in ROMLEX.

Still, ROMLEX is the most comprehensive documentation of the Romani lexicon to date. It is accessible online via a web-interface. The web-interface allows searching for individual words in the available glossaries, both starting from a translation language or from Romani. A detailed description can be found at http://romani.uni-graz.at/romlex/lex.xml.

2 Innovative aspects and impact

The important impact of the future ROMLEX is that it will represent the most comprehensive, comparable and accessible source of lexicographic information on Romani to date. This will be

reached by several measures: the selection of 10 representative Romani varieties, the homogenisation of the lexical coverage of these varieties, linking procedures with respect to content structure both within and across varieties. As a by-product of these linking procedures, lexemes in ROMLEX will be etymologised.

Both the work tasks to be accomplished during the next working phase and the final outcome of the project, the new ROMLEX, are innovative on several levels.

One innovative aspect concerns the homogenisation of the lexical coverage of the glossaries of the 10 selected varieties. The approach to systematically gather lexical material for a number of thematically defined domains for so many varieties has never before been taken in Romani lexicography. Through this approach ROMLEX provides lexicographic data, which is homogenised with respect to lexical coverage and orthography, but allows for the greatest heterogeneity with respect to individual lexical entries. This comparability of data, which is unique in Romani lexicography, is the pre-condition for systematic comparison across varieties.

The actual comparison of the data will be made possible through the re-organisation of the content structure of ROMLEX, i.e. by intra-variety linking and inter-variety linking.

By intra-variety linking individual lexemes are organised hierarchically into clusters of morphologically and etymologically related lexemes. Thus, ROMLEX will not only allow to search for individual lexical entries within a variety (as it does at the present stage), but to search for entire lexeme clusters. The specific relations between members of a lexeme cluster are tagged in the database. At the top of such a lexeme cluster stands the "parent", which is the morphologically simplest (but not necessarily morphologically simple) lexeme available in the respective dialect. It is not suggested, however, that the parent is the primary variant of a number of alternations. It simply is the use of a concrete lexeme for an abstract function, i.e. to be a nod, which holds together a lexeme cluster, and at the same time is the point of contact for inter-variety linking, which is another innovative feature of the content structure of ROMLEX.

Inter-variety linking is accomplished by linking etymological cognates that appear at the top of lexeme clusters, i.e. parent forms, in the individual varieties through an "anchor". The anchor is the most conservative among the attested parent forms. Anchors are tagged for word class and have English translations, which reflect their basic meanings. Furthermore etymological information is attached to the anchors. This second linking measure allows comparing systematically the lexeme-clusters, which have been established by intra-variety linking, across varieties in a very comfortable, and time-saving, way.

The innovative creative insights of the future content and content structure of ROMLEX are as follows:

- Already at the outset of the working phase ROMLEX contains lexicographic information on Romani that goes beyond print-published dictionaries so far to date. As mentioned above, the structure of ROMLEX does not limit the amount of lexicographic information to be contained in the database in the future, and it is a declared aim of ROMLEX to make ten selected varieties even more comprehensive and especially comparable during this working phase. Systematically comparable lexicographic data for so many varieties is unique in the history of Romani lexicography.
- By intra- and inter-variety linkage ROMLEX will be structured in a way so that it has the functionality of a comparative dictionary, presenting information both on the morphology of individual lexemes as well as their etymologies. As such, ROMLEX sets new standards with respect to content, comprehensiveness and accessibility of lexicographic information on Romani.
- Therefore, the format, content structure and content of ROMLEX provides the possibility to conduct linguistic studies in the most comfortable way imaginable, saving both time and work effort, without sacrificing the highest scientific standards.

- Among other things, the new ROMLEX enables to search not only for individual lexemes, but for entire lexeme clusters. This allows lexico-semantic and morphological studies, both on the level of individual varieties and cross-variety perspective. These studies can address questions of historical change, dialect dispersion and language contact.

The innovative creative insights with respect to the field of language documentation and codification:

- ROMLEX represents a new model of language documentation and codification without standardisation, which suffices the demands of a diasporic, heterogeneous, stateless language.
- Furthermore, ROMLEX is an example of sensible use of up-to-date technology in the humanities. In using these technologies the aim of ROMLEX is not to merely produce a digital version of a dictionary, but to create an added scientific value, which sets a new standard of lexicographic documentation. Therefore, ROMLEX illustrates the potential of such technologies for linguistic research in particular, and the humanities in general.

3 Methodology

3.1 Variety selection

The starting point for the coming working phase are the 25 electronic glossaries of individual Romani varieties contained in ROMLEX today. The lexical coverage of these 25 glossaries is very heterogeneous, due to the fact that they are based on quite different sources. While for various reasons, it is not feasible to homogenise the lexical coverage for each variety, e.g. because some varieties are extinct (Welsh Romani) or documentation is very scarce (Gurvari Romani), we will select 10 Romani varieties of the 25 varieties available in the present ROMLEX for the future working agenda. The restriction to 10 varieties also results from the necessity to undertake extensive lexicographic research (both analysis of existing material and elicitation of new material with native speakers). At the same time, the selection of varieties is guided by criteria, which assure that the selection is representative of the manifold of structural diversity of Romani. The main criteria are linguistic relevance and geographic spread, but for reasons of practicability also availability of data and feasibility of fieldwork are taken into account.

- Linguistic relevance refers to different structural types of varieties: In modern Romani dialectology (e.g. Bakker & Matras 1997, Elšík 2000) different structural types are reflected by a division into several dialect groups of equal ranking. This division has become a popular reference grid in Romani linguistics. The groups are Vlax, Balkan, Central and Northern, with a later (Matras 2002) differentiation into Northwestern, Northeastern and a number of isolated dialects. The individual groups are defined by individual sets of features that are characteristic of the group, and often also by geographic proximity. While group affiliation is not problematic for varieties that are prototypical representatives of a group, these divisions might come short in assigning a group affiliation for varieties that are transitional between the individual groups. This has to do with the interpretation of the groups within a genetic framework of dialectology, which has been challenged by Matras (2005). Nonetheless, for variety selection in the present context the above division provides a useful guideline, since it reflects the most prominent structural types of Romani varieties.
- Geographic spread in the context of variety selection on the one hand refers to taking into account Romani varieties from a wide range of geographic locations in order to include the highest possible rate of inter-variety variation. On the other hand it applies to taking into account Romani varieties, which are spread out over a large, not necessarily contiguous, geographic area, in order to include the highest possible rate of intra-variety variation.

Taking into account these two criteria in combination with availability of data, which mostly refers to work on Romani in Austria in the last decade, and feasibility of fieldwork, which refers to already established contacts with native speakers of particular varieties, the following list of 10 Romani varieties is suggested:

Variety	Description	N°*	Fieldwork possibilities
Kalderaš-R.	North-Vlax / European++ [i.a. Austria]/ well documented	9.824	good contacts with native speakers
Lovara-R.	North-Vlax / European+ [i.a. Austria]/ well documented	4.955	good contacts with native speakers
Gurbet-R.	South-Vlax / European+ [i.a. Austria]/ partly documented	9.825	contacts with native speakers
Ursari-R.	Balkan / European / poorly documented	2.636	good contacts with native speakers
Arli-R.**	Balkan / European++ [i.a. Austria] / documented	7.210	contacts with native speakers
Burgudži-R.	Balkan / European [i.a. Austria] / documented	3.971	contacts with native speakers
East Slovak-R.	Northern Central / regional / well documented	8.914	good contacts with native speakers
Burgenland-R.***	Southern Central / local [East-Austria] / well documented	6.255	good contacts with native speakers
North Russian-R.	Northeastern / regional+ / well documented	4.602	few contacts with native speakers
Sinte-R.	Northwestern / European+ [i.a.Austria] / well documented	7.453	few contacts with native speakers

* N° = current number of entries in ROMLEX
** 7.210 = sum of 3 varieties which are currently included separately (Kosovo-Arli-, Macedonian-Arli-, Sofia-Erli-R.)
*** geographically most Southern Central varieties are local, have only a few speakers and are threatened by extinction

This selection of Romani varieties reflects the manifold of structural variation of Romani and gathers representatives of all dialect groups. Included in this selection are the six larger Romani varieties spoken in Austria and Bugurdži, a Romani variety spoken by a small speech community in the Vienna area, first of all, because these varieties are especially relevant in the context of an Austrian research project. Secondly, the distribution of Romani dialects spoken in Austria today reflects the heterogeneity of Romani varieties in Western European countries, which results from differences in the socio-historical background and immigration time of the single groups (cf. Halwachs 2005). As indicated, the ten individual glossaries significantly differ in numbers. There are a lot of printed sources for most of the varieties, which were collected and only partly evaluated during the first two working phases of ROMLEX. During these previous activities also good contacts to native speakers of almost all ten selected varieties were established. This makes large-scale fieldwork as intended in the current project proposal feasible within the time-frame of the project.

3.2 Homogenisation of lexical coverage

The basic idea of homogenisation of lexical coverage is to compile a wordlist of approximately 5000 lexical entries, which reflect the basic vocabulary of Romani, and which will be the core of the glossaries of each of the 10 selected Romani varieties.

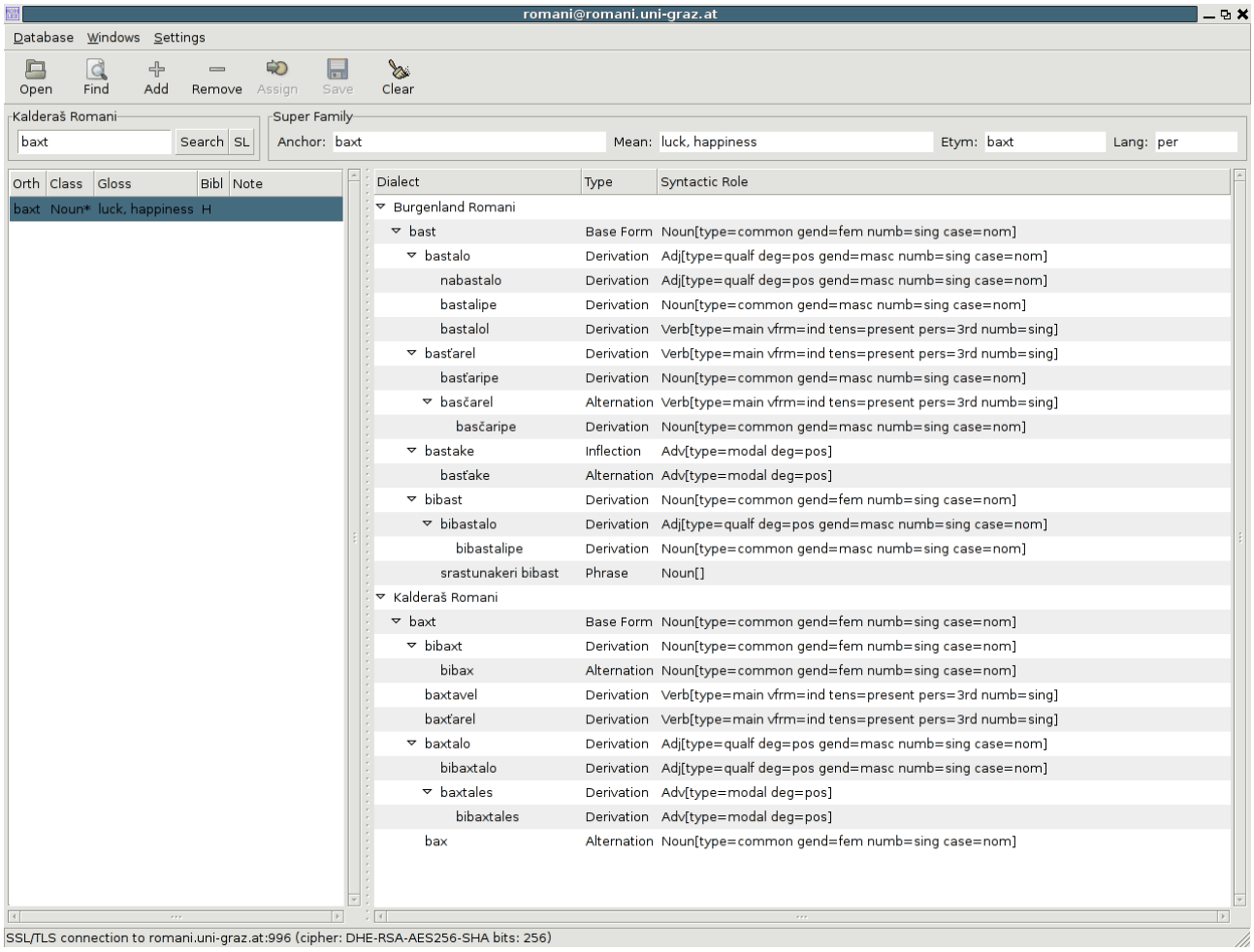
The term 'basic vocabulary' does not refer to a fixed inventory of lexical items in lexicography and linguistics in general. While it is intuitively clear, which basic domains should be included in a compilation of basic vocabulary, there is no well-defined list of all basic domains that qualify for basic vocabulary. Looking at lists of basic semantic domains in lexicographic works (Hoffmann & Hoffmann 2003, McGregor 2005, Muhr 2000), one finds that there is a core of semantic domains, which always appears in such lists, these include entities like human beings, kin terms, body parts and body products, animal and animal products, food, everyday technology and artefacts, plants and plant products, the world (including weather and time), social life; events like bodily activities and states, human activities, motion, position, thinking / speaking / perception, production of noises (non-speech), grasping / grabbing / holding; furthermore lexemes referring to quantities, qualities and anchoring in time and space as well as a limited set of function words (pronouns, determiners, basic conjunctions, adpositions). In addition culture specific values, beliefs and practices are

considered basic vocabulary. Some of the domains just enumerated are part of the pre-European lexicon of Romani (Matras 2002: 20ff), which forms the starting point for a wordlist of basic vocabulary in the present context. This list will be expanded with lexemes from the above enumerated basic semantic domains, with special attention to items that are important to Roma culture. The existing glossaries of the ten selected Romani varieties will then be checked against this elaborated list, missing lexemes will be amended through lexicographic research (analysis of existing materials, fieldwork with native speakers). Electronic search tools, which make this work tasks more effective, are to be developed in the course of the project.

The final outcome of this work task is the actual expansion of the ten selected varieties, so that each of them includes the approx. 5000 lexemes contained in the elaborated basic vocabulary list, in addition to lexemes from other domains, as already documented in the now existing glossaries of the ten varieties.

3.3 Linking processes and etymology

A central work task of the future working phase is the re-organisation of the content structure of ROMLEX by intra-variety linking and inter-variety linking, which will be accomplished for the ten selected dialects. The following screenshot of a test run gives an impression of the future linking structure of ROMLEX:



Both intra-variety linking and inter-variety linking require a well-founded knowledge of the grammar of Romani. This work task is to be completed by two full-time academic researchers. All linking measures have to be done "by hand" via the database frontend.

- Intra-variety linking organises individual lexemes hierarchically into clusters of morphologically and etymologically related lexemes. The structures resulting from dialect-internal hierarchical linking represent clusters of related word forms, with internal specifications of the relations holding between individual word forms. Such clusters are extremely valuable in Romani dialectology in general and cross-variety studies of derivational processes in Romani in particular. From a lexicographic viewpoint such clusters can help to identify gaps in individual glossaries and to modify elicitation tools accordingly. Furthermore, hierarchical lexeme clusters bring a practical advantage for future editing of printed dictionaries based on ROMLEX. Since links between two elements are specified, the ordering of elements in the printed dictionary could be automated (e.g. derivations could be sorted automatically under the head word in a dictionary).

  At the top of such a lexeme cluster stands the "parent", which is the morphologically simplest (but not necessarily morphologically simple) lexeme available in the respective variety. A "parent" is a form that can't be derived from any other lexeme in the variety. Derived forms can function as parent if the bases they are derived from are obsolete in the variety at question. The Burgenland Romani word *prast-a-l* (historically *prast-av-el*) 'to ride a horse', for instance, functions as a parent, since the historical base form *prast-el* – which is attested in other varieties – is not attested in today's Burgenland Romani. The parent is the link out of the variety to the anchor.

  The specification of the relations holding between individual word forms of lexeme clusters are intended to reflect processes at work in Romani word formation. While the background of extensive work on Romani suggests some obvious types of relations (e.g. derivation, which covers various types of derivational processes; alternation, which refers to phonetic variants; verb phrase, which refers to light verb constructions with verbs such as *del* 'to give'), it has to be emphasised that a comprehensive list of the relevant types of relations for Romani can only be established through empirical work, i.e. actual linking, in which type definitions are tested, evaluated and then if necessary adapted or amended. In the course of this work it will be crucial to establish principles of linking, which warrant that word forms, which could be linked to several elements, can be dealt with in a unified way. These linking principles will account for uniform, strictly hierarchically intra-variety linking.

  A number of preliminary linking principles, some of which are displayed by the screenshot on the previous page, have been established in the course of the conceptual development of the new content structure of ROMLEX. So far, these linking principles have been applied only to a limited extent in preliminary linking test runs in previous working phases. These test runs suggest that the set of linking principles will need to be elaborated and modified content wise in the course of large-scale intra-variety linking of a whole Romani variety. Also, only large-scale linking will make it possible to test the technical feasibility of the linking principles established so far within the database. Therefore, both test runs to see how operational linguistically the established linking principles are in actual linking, and technical test runs to evaluate the implementation of these linking principles in the database are important work tasks at the outset of intra-variety linking.

- Inter-variety linking is accomplished by linking etymological cognates that appear at the top of lexeme clusters, i.e. parent forms, in the individual varieties through an "anchor" which functions as base entry. The anchor is the most conservative among the attested parent forms. Reconstructed forms are not used in ROMLEX. Anchors are tagged for word class and have English translations, which reflect their basic meanings. Furthermore etymological information is attached to the anchors.

  Even though the anchor is an actually existing form, in its function as anchor it is stripped of its affiliation to a particular variety. Thus, taking the most conservative of the attested parent forms does not mean that we suggest that this form is in any way more valuable or more important than the other attested forms. It has to be emphasised, that "the most conservative

form" can also be a recent loan from a current contact language. The aim of anchoring is by no means to create a unified "hyper-variety" of Romani, which could be interpreted as a standard variety of Romani, through the backdoor. Anchors function as two-fold technical nods for a multi-variety dictionary of Romani: they are links between the individual Romani varieties, and they are links between the individual Romani varieties (= variety specific entries) and the etymologies. Each anchor has exactly one link to an etymology and links to exactly one (or no) parent in each of the individual Romani varieties. In the absence of a standard variety of Romani that could function as a link between the individual varieties, anchors are the only way to reach the aim of a multi-variety dictionary of Romani. However, the functions of anchors could also be fulfilled by a random variable. The choice for an actually existing lexeme as anchor is a matter of convenience, since the forms at question are contained in the database and can be elected as anchors by a simple tagging procedure. In addition, an actually existing form has the advantage that the interpretation of etymological information for lexeme clusters in the individual varieties does not depend on computational reading.

Etymologies are based on a thorough survey of the existing literature, including not just dictionaries that have listed etymologies (e.g. Miklosich 1872-80, Turner 1926, Sampson 1926, Vekerdi 1983, Manuš et al. 1997, Valtonen 1972), but also numerous articles that suggest etymologies for individual items or groups of words (e.g. Boretzky 1995, Tálos 1999). Etymologies are tagged with references, in order to provide a comprehensive picture of the etymological discussion, where applicable. Furthermore, the project will seek cooperation with historical linguists, who specialise on Indo-European languages, in order to clarify and check unclear or disputable etymologies.

The etymology for inherited, i.e. Indo-Aryan, lexemes should go back to the oldest equivalent word form possible. For the etymology of loan words (Asiatic & European) only the "contact form", i.e. the form of the word at the time of borrowing, is relevant, and not the oldest possible form of the word. In other words: the etymology of loan words within the contact language of origin is irrelevant.

The enabling technology for all these linking processes and the inclusion of etymological information is already available in the ROMLEX database. Electronic search tools, which make inter-variety linking more effective are to be developed in the course of the project phase described.

4 Cooperation

ROMLEX has already benefited from intense cooperation in its earlier phases, both at the national and at international level: At the national level, first of all with the "Romani Projekt Graz", which provided both lexical data, knowledge on language documentation in this particular context as well as infrastructure, and with the "Phonogrammarchiv" of the Austrian Academy of Science, which provided the project with valuable corpora (especially the "Sammlung Heinschink") from which lexical material was extracted. It is intended to continue intensive cooperation with both organisations in the next project phase.

At the international level during the previous phases collaboration has been under way with various partners and partner institutions: The most important of them is the "Manchester Romani Project" [http://www.llc.manchester.ac.uk/Research/Projects/romani/] at the School of Languages, Linguistics and Cultures at the University of Manchester. The "Manchester Romani Project" is the partner project of the "Romani Projekt Graz" [http://romani.uni-graz.at/romani/]. Other important partners are i.a. the Institute of Linguistics and Finno-Ugric Studies at the Charles University in Prague and the Institute of Anthropology, Archaeology and Linguistics at Aarhus University. It is intended to continue regular exchange with these partners and others in the framework of the present proposal. We will furthermore intensify and establish new contacts with scholars working in the field of language documentation.

References

Bakker, Peter & Matras, Yaron. 1997. Introduction. In: Matras & Bakker & Kyuchukov. eds.: vii-xxx.

Boretzky, Norbert. 1993. *Bugurdži: Deskriptiver und historischer Abriß eines Romani-Dialekts*. Wiesbaden: Harrassowitz.

Boretzky, Norbert. 1995. Armenisches im Zigeunerischen (Romani und Lomavren). *Indogermanische Forschungen* 100: 137-155.

Boretzky, Norbert & Igla, Birgit. 1994. *Wörterbuch Romani-Deutsch-Englisch für den südosteuropäischen Raum: Mit einer Grammatik der Dialektvarianten*. Wiesbaden: Harrassowitz.

Calvet, Georges. 1982. *Lexique Tsigane. Dialecte des Erlides de Sofia*. Paris: Publications Orientalistes de France.

Cech, Petra & Heinschink, Mozes F. 1998. *Basisgrammatik*. *Arbeitsbericht 1 des Projekts "Kodifizierung der Romanes-Variante der österreichischen Lovara"*. Wien: Romano Centro.

Cech, Petra & Heinschink, Mozes F. 1999. *Sepečides-Romani: Grammatik, Texte und Glossar eines türkischen Romani-Dialekts*. Wiesbaden: Harrassowitz.

Cech, Petra & Fennesz-Juhasz, Christiane & Heinschink, Mozes F. eds. 1998. *Texte österreichischer Lovara*. *Arbeitsbericht 2 des Projekts "Kodifizierung der Romanes-Variante der österreichischen Lovara"*. Wien: Romano Centro.

Demeter, Roman S. & Demeter, Pjotr S. & Tcherenkov, Lev N. 1990. *Gypsy-Russian and Russian-Gypsy Dictionary (Kalderaš dialect)*. Moscow: Russkij Jazyk.

Elšík, Viktor. 2000. Dialect variation in Romani personal pronouns. In: Elšík & Matras. eds.: 65-94.

Elšík, Viktor & Matras, Yaron. eds. 2000. *Grammatical relations in Romani: The noun phrase*. Amsterdam: Benjamins.

Finck, Franz Nikolaus. 1903. *Lehrbuch des Dialekts der deutschen Zigeuner.* Marburg: Elwert.

Friedman, Victor. A. 2003. Romani as a minority language, as a standard language, and as a contact language: Comparative legal, sociolinguistic, and structural approaches. In: Fraurud & Hyltenstam. eds.: 103-134.

Friedman, Victor A. 2005. The Romani Language in Macedonia in the third millenium: Progress and Problems. In: Schrammel, B. & Halwachs, D. W. & Ambrosch, G. eds.: 163-173.

Gjerdman, Olof & Ljungberg, Erik. 1963. The language of the Swedish Coppersmith Gipsy Johan Dimitri Taikon: Grammar, texts, vocabulary and English word-index. Uppsala: Lundequist.

Halwachs, Dieter W. 1998. *Amaro vakeripe Roman hi / Unsere Sprache ist Roman: Texte, Glossar und Grammatik der burgenländischen Romani-Variante*. Klagenfurt: Drava.

Halwachs, Dieter W. 2003. The Changing Status of Romani in Europe. In: Hogan-Brun & Wolff. eds.: 192–207.

Halwachs, Dieter W. & Ambrosch, Gerd & Wogg, Michael. eds. 1998. *Märchen und Erzählungen der Burgenland-Roma*. Oberwart: Verein Roma.

Halwachs, Dieter W. & Menz, Florian. eds. 1999. *Die Sprache der Roma: Perspektiven der Romani-Forschung in Österreich im interdisziplinären und internationalen Kontext*. Klagenfurt: Drava.

Halwachs, Dieter W. & Heinschink, Mozes F. 2000. *Language change in progress. The case of Kalderash Romani in Vienna.* Paper presented at the 5th International Conference on Romani Linguistics, Sofia, 14-17 September 2000.

Halwachs, Dieter W. & Ambrosch, Gerd unter Mitarbeit von Deman & Katharina, Glaeser, Ursula & Wogg, Michael. 2002. *Wörterbuch des Burgenland-Romani (Roman)*. *Arbeitsbericht 10 des Roman-Projekts*. Graz: Romani-Projekt.

Halwachs, Dieter W. 2005. Roma and Romani in Austria. *Romani Studies* 5/15-2: 145-173.

Hoffmann, Hans G. & Hoffmann, Marion. 2003. *Großer Lernwortschatz Englisch. 15000 Wörter zu 150 Themen. Erweiterte und aktualisierte Neuausgabe*. Ismaning: Max Hueber Verlag.

Holzinger, Daniel. 1993. *Das Romanes: Grammatik und Diskursanalyse der Sprache der Sinte*. (= Innsbrucker Beiträge zur Kulturwissenschaft, 85.) Innsbruck: Verlag des Instituts für Sprachwissenschaft der Universität Innsbruck.

Hübschmannová, Milena & Šebková Hana & Žigová, Anna. 1991. *Kapesní slovník romsko český a česko romský*. Praha: Státní pedagogické nakladatelství.

Igla, Birgit. 1996. *Das Romani von Ajia Varvara. Deskriptive und historisch-vergleichende Darstellung eines Zigeunerdialekts*. Wiesbaden: Harrassowitz.

McGregor, William. 2005. Frs. Herman Nekes and Ernest Worms' dictionary of Australian languages, Part III of Australian languages (1953). In: Proceedings of the 2004 Conference of the Australian Linguistic Society. Sidney: University of Sydney.

Mānušs, Leksa (with Neilands, Jānis & Rudevičs, Kārlis). 1997. *Čigānu-latviešu-angļu etimoloģiskā vārdnīca un latviešu-čigānu vārdnīca*. Rīgā: Zvaigzne ABC.

Matras, Yaron. 1994. *Untersuchungen zu Grammatik und Diskurs des Romanes. Dialekt der Kelderaša / Lovara*. Wiesbaden: Harrassowitz.

Matras, Yaron. 2002. *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.

Matras, Yaron. 2004. The role of language in mystifying and de-mystifying Gypsy identity. In: Saul & Tebbut. eds.: 53-78.

Matras, Yaron. 2005. The classification of Romani dialects: A geographic-historical perspective. In: Schrammel, B. & Halwachs, D. W. & Ambrosch, G. eds.: 7-22.

Matras, Yaron. ed. 1995. *Romani in contact: the history and sociology of a language*. Amsterdam: Benjamins.

Matras, Yaron & Bakker, Peter & Kyuchukov, Hristo. eds. 1997. *The typology and dialectology of Romani.* Amsterdam: John Benjamins.

Miklosich, Franz. 1872-1880. *Über die Mundarten und Wanderungen der Zigeuner Europas* I-XII. Wien: Karl Gerold's Sohn.

Muhr, Rudolf. 2000. *Österreichisches Sprachdiplom Deutsch. Lernzielkataloge zu Basisformulierungen, Lexik-Sprechhandlungen, Höflichkeitskonventionen, Diskurs- und Diskursstrukturen, Deutsch als plurizentristische Sprache*. Wien: öbv & hpt.

Paspati, Alexandre G. 1870 [reprint 1973]. *Études sur les Tchinghianés ou Bohémians de l'Empire Ottoman*. Istanbul: Karomela. [Osnabrück: Biblio]

Pobożniak, Tadeusz. 1964. *Grammar of the Lovari dialect.* Kraków: Pa*ƒ*stwowe wydawnictwo naukowe.

Pott, August. 1844-1845. *Die Zigeuner in Europa und Asien. Ethnographisch-linguistische Untersuchung vornehmlich ihrer Herkunft und Sprache*. Halle: Heynemann.

Rüdiger, Johann C. C. 1782. [reprint 1990]. Von der Sprache und Herkunft der Zigeuner aus Indien. In: *Neuester Zuwachs der teutschen, fremden und allgemeinen Sprachkunde in eigenen Aufsätzen*, 1. Stück. Leipzig. [Hamburg: Buske]

Sampson, John. 1926 [reprint. 1968]. *The dialect of the Gypsies of Wales, being the older form of British Romani preserved in the speech of the clan of Abram Wood*. Oxford: Clarendon Press.

Schrammel, Barbara & Halwachs, Dieter W. & Ambrosch, Gerd. eds. 2005. *General and Applied Romani Linguistics. Proceedings of the Sixth International Conference on Romani Linguistics*. München / Newcastle: Lincom Europa.

Soravia, Giulio & Fochi, Camillo. 1995. *Vocabolario sinottico delle lingue zingare parlate in Italia*. Roma: Centro Studi Zingari / Instituto di Glottologia Università di Bologna.

Sowa, Rudolf von. 1887. *Die Mundart der slovakischen Zigeuner.* Göttingen: Vandenhoeck und Ruprechts Verlag.

Tálos, Endre. 1999. Etymologica Zingarica. *Acta Linguistica Hungarica* 46: 215-268.

Tenser, Anton. 2005. *Lithuanian Romani (= LW/M 452).* München / Newcastle: Lincom Europe.

Turner, Ralph L. 1926. The position of Romani in Indo-Aryan. *Journal of the Gypsy Lore Society*, 3/5: 145-189.

Uhlik, Rade. 1983. *Srpskohrvatsko – romsko – engleski rečnik. Romengo alavari*, Sarajevo: Svjetlost, oour Izdavačka djelatnost.

Valtonen, Pertti. 1972. *Suomen mustalaiskielen etymologinen sanakirja*. Helsinki: Soumalaisen kirjallisuuden seura.

Vekerdi, József. 1983. *A magyarországi cigány nyelvjárások szótára*. *(= Tunolmányok, 7)*. Pécs: Janus Pannonius Tudományegyetem Tanárképző Kara.

Wentzel, Tatjana W. 1980. *Die Zigeunersprache (Nordrussischer Dialekt)*. Leipzig: Enzyklopädie.

Wolf, Siegmund A. 1960. *Grosses Wörterbuch der Zigeunersprache*. Mannheim: Bibliographisches Institut.