

# Lexicography of a Non-State Language: The Case of Burgenland Romani

Barbara Schrammel and Astrid Rader

University of Graz, Austria

*Burgenland Romani (henceforth BR) is spoken in Burgenland, the easternmost province of Austria. Until recently BR was an exclusively oral language. However, active language use of BR has almost totally ceased in the second half of the 20th century. The self-organisation of the group from the 1990s onwards led to a new appreciation of the language, which is now accepted as the primary identity marker. This new interest in their own language and culture entails the desire for the revival, maintenance and spread of BR. One aspect of language planning in BR concerns the functional expansion of the language into acrolectal domains where it has never been used before.*

*BR is lexicographically documented in two different media, i.e. in ROMLEX (henceforth RL), which is an extendible multi-dialectal lexical database with a freely accessible web-interface (<http://romani.uni-graz.at/romlex/>) and a print dictionary. RL is intended as a tool for comprehensive lexical documentation of BR. At the same time, it is a practical, low-threshold tool for text producers. The print dictionary, on the other hand, primarily serves an emblematic purpose. Given the differing purposes of RL and the print dictionary, different strategies are used in lexicographic decision-making. Roughly speaking, RL favours an inclusive descriptive approach while the print dictionary is rather restrictive and follows normative principles. The paper discusses decisions taken with respect to orthography, lemma selection and meaning for RL and the print dictionary, respectively. We are highlighting lexicographic phenomena, such as increased polysemy, generic usage of terms and heavy borrowing, which are typical of the functional expansion process of stateless minority languages.*

## 1. The speech community and language use

Burgenland Romani (henceforth BR)<sup>1</sup> is a Romani variety spoken in the province of Burgenland, the easternmost province of Austria. Until recently BR was an exclusively oral language. The Burgenland Roma have been living in the area of today's Burgenland since the 15<sup>th</sup>/16<sup>th</sup> centuries. This region has always been multilingual with the languages Hungarian, Croatian, German and Romani. Up to the 20<sup>th</sup> century speakers of BR were subsequently typically plurilingual, whereas nowadays they are merely bilingual with German. Numerous borrowings in BR, especially from Hungarian and German, bear witness to this multilingual setting.

In the 18<sup>th</sup> century the rulers of the Hapsburg monarchy issued a number of laws in order to assimilate the Burgenland Roma into mainstream society. These laws included measures such as the prohibition of nomadic life, forbidding the use of their language, as well as forced mixed marriages and the raising of Roma children outside their families. Despite all these measures the authorities did not succeed in preventing the Roma from preserving and passing on their culture and language. Until the Nazi era, Burgenland Romani was the primary language used in the social microcosmos (family, intra-group communication). Of the approximately 6,000-8,000 Burgenland Roma before 1938, only a few hundred survived the Nazi concentration camps<sup>2</sup>. The genocide by the Nazis resulted in the interruption of the transmission of culture and language of the Burgenland Roma. There are two main reasons for this interruption. First, the extinction of almost the entire generation of 'elders', who traditionally had an important part in transmitting the language and cultural traditions, led to an irrecoverable loss of knowledge. Second, the traumatising experiences during the Nazi era

---

<sup>1</sup> The speakers refer to their language as *Roman*.

<sup>2</sup> It is problematic to determine the exact population of the Burgenland Roma before and after the war. Since the label 'Gypsy' has been attributed to different groups of people (for different reasons) before and after WWII the available figures on group size are especially problematic (see discussion in Baumgartner/Freund 2004: 43ff).

resulted in self-imposed assimilation and self-denial of Roma cultural heritage. Consequently, with few exceptions, Burgenland Romani was no longer the primary language of socialisation.

Today, an estimated 2,500 Burgenland Roma live in Austria (Halwachs 1998: 14). Nowadays active language use of BR has almost totally ceased. However, the self-organisation of the group from the 1990s onwards led to a new appreciation of the cultural heritage, especially the language, which is now accepted as the primary identity marker (Halwachs 2001: 217). This new interest in their own language and culture entails the desire for the revival, maintenance and spread of BR. In 1993 a project for the codification of BR in cooperation with linguists from the University of Graz started. The overall aim of codification and grammatical description, on the part of the BR speakers, was to provide a basis for the future teaching of BR in schools. In the same year the Burgenland Roma were acknowledged as an official Austrian minority ('österreichische Volksgruppe') and therefore the legislation concerning Austrian minority languages ('österreichische Volksgruppensprachen') is applicable to BR. This means that the members of the group are entitled to financial support for the translations of official documents, the preparation of teaching material and media in BR. The officialisation of BR (see table 1) entails the functional expansion of BR into acrolectal domains where it has never been used before. Language planning strategies and measures, which will be discussed in more detail below, have successfully led to regular language courses for different age levels, a wealth of official documents and texts, liturgical texts as well as periodicals and radio broadcasts in BR. These outcomes can be seen as an indicator for the rising emblematic value of the language. However, every day oral use of the language is barely affected by these measures.

## **2. Language planning in Burgenland Romani**

In the following we describe language planning policies and measures applied for BR in the last 15 years using the terminology and concepts of Hornberger's (2006) integrative framework of language policy and planning. Due to the fact that BR is a very small, regionally confined, stateless minority language not all of the aims identified by Hornberger are equally applicable to BR (see table 1).

As mentioned above, the self-organisation of the Burgenland Roma from the 1990s onwards led to BR becoming the group's primary identity marker. This is remarkable since every-day language use had been in decline ever since WWII and the language was severely threatened with extinction in the 1990s. On the other hand, the community perceived BR as the most overt remnant of their cultural heritage, while being aware that BR is close to extinction. In light of this, it is understandable why one of the most important aims of the newly founded Roma organisation was language maintenance and revival. Since there were only very few competent speakers of BR left by the 1990s, the traditional method of oral transmission of the language was irrevocably interrupted. This means that the maintenance of BR could only be guaranteed through (re)acquisition of the language by the younger generations through formal instruction based on written materials. The precondition to reach this aim was the codification of BR.

Section 10. Lexicography of lesser used or non-state languages

| TYPES                        |   | POLICY PLANNING APPROACH         |  | CULTIVATION PLANNING APPROACH           |  |
|------------------------------|---|----------------------------------|--|---|--|
| extralinguistic aims         | STATUS PLANNING<br>uses of language       | SE<br>LE<br>CT<br>ION            | Officialization<br><i>Nationalization</i><br><i>Standardization</i><br><i>Proscription</i> | IM<br>PL<br>E<br>M<br>EN                | Revival<br>Maintenance<br><i>Spread</i><br><i>Interlingual communication: inter- &amp; intranational</i>   |
|                              | ACQUISITION PLANNING<br>users of language |                                  | Group<br>Education/School<br>Literature<br>Religion<br>Mass Media<br><i>Work</i>           |   | TA<br>TI<br>ON   |
| linguistic aims <sup>3</sup> | CORPUS PLANNING<br>about language         | CO<br>DI<br>FI<br>CA<br>TI<br>ON | Standardization of corpus<br>Graphization<br><i>Standardization of auxiliary code</i>      | EL<br>A<br>B<br>O<br>RA<br>TI<br>O<br>N | Modernization (new functions)<br>Lexical<br>Terminology unification<br><i>Modernization (new functions)</i><br><i>Stylistic</i><br><i>Renovation (new forms, old functions)</i><br><i>Purification</i><br><i>Reform</i><br><i>Stylistic simplification</i> |

*italics ..... not applicable in this context*

Table 1. Integrative framework of language policy and planning (based on Hornberger 2006:29)

The codification of BR, comprising graphisation and standardisation of corpus, began in 1993. The grammatical and lexicographic descriptions of BR were based on spoken and subsequently transliterated narratives from the remaining competent speakers as well as older records of the language<sup>4</sup>. All BR texts up to the year 1996 are considered part of this codification corpus. The output of the codification process is a comprehensive grammatical description of BR, a glossary comprising about 6,000 words and teaching materials. In order to make (re)acquisition of BR attractive to the community, it was necessary to provide contexts for using BR. Since it is quite difficult to trigger a shift from German to BR in intra-family communication from the outside, the BR language planners tried to find new attractive contexts in which BR could be used, such as literature, religion, and mass media. Due to the high prestige of printed literature in our society, literature in BR gives the community a feeling of equality of their language with other languages with a written literary tradition. Since religion traditionally has a high value among Burgenland Roma, it was natural to include religious contexts into the language planning activities: there are numerous translations of bible texts and prayers into BR and the language has been used in services, e.g. for intercessions and baptisms.

The most significant part of the functional expansion of BR, which includes the most extensive text production, is the use of BR in mass media. There are two bi-lingual

<sup>3</sup> Hornberger (2006: 29) distinguishes between *linguistic aims* attributed to the *policy planning approach* and *semi-linguistic aims* attributed to the *cultivation planning approach*. In Hornberger's model the focus with respect to the latter is on language functions. Since our focus is on the actual lexical outcomes of language planning measures, we subsume both aspects of the *corpus planning type* as *linguistic aims*.

<sup>4</sup> These older records comprise BR texts collected and written down by Knobloch in the 1940s and narratives recorded by Heinschink from the 1960s onwards.

(BR/German) periodicals and one mono-lingual periodical (for children) in BR, the language is used for weekly radio broadcasts and news on the internet and there are computer games for language learning. All these media contribute to the modernisation of BR, especially with respect to its lexicon. Implementing language computer games in BR had the effect of introducing technical vocabulary into BR, such as terms for menu navigation, computer commands and error messages. Periodicals and radio broadcasts in BR contribute a wealth of terminology in the fields of politics, history, social problems, art, sports etc. All these texts constitute the elaboration corpus of BR. Due to the general decline in competence in BR, the creation of words for new concepts quite often leads to parallel forms. Terminology unification is required but hard to realise, since the acceptance of the respective forms should be determined via feedback from the whole community. However, periodicals and radio broadcasts constitute one-way-communication – the reactions of community members to these lexical creations are scarce.

A further linguistic effect of using BR in new domains is the emergence of a new register of BR which is the result of a change in the influence of the contact language. While older records of BR reflect the impact of the two most important contact languages, Hungarian (in the first half of the 20<sup>th</sup> century) and dialectal German, lexical expansions as well as emerging grammatical structures in the functionally expanded variety are heavily influenced by formal standard German. This change of the main contact varieties results in two varieties of BR – one prevalently oral, the other prevalently written – which differ with respect to lexicon, syntax and morphosyntax.

In the following we describe the composition of the BR lexicon, followed by a discussion of the lexical documentation of BR in two media: the online database ROMLEX (<http://romani.uni-graz.at/romlex/>) and a print dictionary.

### 3. The Lexicon of Burgenland Romani

Like all Romani varieties BR is characterised by several layers of loanwords. The pre-European component of the lexicon consists of lexemes of Indo-Aryan origin and early loans from Persian, Armenian, and Byzantinian-Greek. The number of pre-European words is rather low: the dictionary of Boretzky/Igla (1994) holds about 700 lexemes of Indian origin, about 70 of Persian origin, 40 of Armenian, and about 230 lexemes of Greek origin. In the individual Romani varieties we typically find about 600 pre-European lexemes. The shared European part of the lexicon consists mostly of lexemes from the languages of the Balkans. The more recent lexical layers depend on the individual history of Roma groups, in the case of BR the recent lexical layer of the codification corpus is characterised by lexemes from Hungarian and dialectal German. The elaboration corpus, on the other hand, is characterised by the influence of standard German.

#### 3.1. The codification corpus

In the BR codification corpus there are several strategies to verbalise new concepts. There is productive morphology to integrate foreign lexical material. All borrowed verb stems are integrated by the suffix *-in-* and form a separate conjugation class: e.g. *mulat-in-* ‘celebrate’ ← hun. *mulat* ‘celebrate, make merry’, *bliatinel* ‘bleed’ ← ger.dial. *bluten* ‘bleed’. The integration markers also function as gender markers with nouns. Nonetheless, nouns can also be integrated without any overt integration morphology in both genders. The selection of a particular marker is triggered by the gender of the German word, the gender of Hungarian loans is determined by sexus or by the ending of the respective Hungarian noun: e.g. *farajn-Ø*

(m.) ‘association’ ← ger. *Verein* (m.) ‘association’, *rafang-o* (m.) ‘chimney’ ← ger.dial. *Rauchfang* (m.) ‘chimney’, *šrafčigeri* (m.) ‘screwdriver’, ← ger.dial. *Schraubenzieher* (m.) ‘screwdriver’, *form-a* (f.) ‘form’ ← ger. *Form* (f.) ‘form’, *roas-Ø* (f.) ‘journey’ ← ger.dial. *Reise* (f.) ‘journey’, *bika-Ø* (f.) ‘bull’ ← hun. *bika* ‘bull’. The last example illustrates an instance where the form of the borrowed word (*bika*) overrides the sexus of the designated concept (bull) and determines gender assignation in BR (words ending in *-a* typically have feminine gender in BR).

The importance of the word form for gender assignation has to do with the fact that Hungarian has no gender. With the change to German as the primary contact language, the German gender of borrowed nouns becomes the main determining factor. Adjectives are integrated with *-(n)o* and enter the declinable adjective class of BR, e.g. *seleno* ‘green’ ← slaw. *zelen*, *roštašno* ‘rusty’ ← hun. *rozsdás*, *frišno* ‘fresh’ ← ger. *frisch*. Recent German loan adjectives typically are integrated with *-i* (in attributive use, see example (1 and 2))/*Ø* (in predicative use, see example (3)) and are indeclinable in BR: e.g.

- |     |                           |                          |                         |                      |                     |
|-----|---------------------------|--------------------------|-------------------------|----------------------|---------------------|
| (1) | BR <i>i</i><br>det.nom.f  | <i>grau-i</i><br>grey-im | <i>mačka</i><br>cat.f   |                      | ‘the grey cat’      |
| (2) | BR <i>o</i><br>det.nom.pl | <i>grau-i</i><br>grey-im | <i>mačk-i</i><br>cat-pl |                      | ‘the grey cats’     |
| (3) | BR <i>o</i><br>det.nom.pl | <i>mačk-i</i><br>cat-pl  | <i>grau</i><br>grey     | <i>hi</i><br>be.3.sg | ‘the cats are grey’ |

While integration with *-i* is consistent with Bakker’s (1997: 13f.) model of athematic morphology, Halwachs (1998: 111) argues that the forms in *-i* are a result of the copying of German adjectives, which in their dialectal form end in *-i*. Besides borrowed adjectives, there are many adjectives that have borrowed stems, but were derived in BR from loans of other word classes. There are a number of adjectival suffixes with clear semantic contents (e.g. composition/texture, ethnicity). In BR there is only one suffix that still has a clear semantic content, i.e. the suffix *-itiko* indicates ethnic and geographic provenience, the only counter-example being (*minden*)*felitiko* ‘various’.

Another important strategy to verbalise new concepts in BR is calquing on the model of German words. With nouns the most important German model is composition, which in BR is rendered with an adjectival genitive (modifier) in combination with a noun (head). The use of an adjectival genitive to modify a noun is a common Romani phenomenon, with both the head and the modifier keeping their nominal categories, i.e. gender and number. These are displayed in the respective definite articles, as shown in example (4) from Kalderaš Romani (KR).

- |     |                          |                            |                       |                                       |          |
|-----|--------------------------|----------------------------|-----------------------|---------------------------------------|----------|
| (4) | KR <i>o</i><br>det.nom.m | <i>rakhlo</i><br>son.nom.m | <i>e</i><br>det.obl.m | <i>car-os-k-o</i><br>king-obl.m-gen-m | ‘prince’ |
|-----|--------------------------|----------------------------|-----------------------|---------------------------------------|----------|

On the model of German compounds in BR this construction underwent some changes: First, corresponding to the German model the word order is always modifier-head; second, the genitive does not trigger congruence anymore (see example 5). Besides these calqued structures, there are still more conservative variants of the construction, in which the genitive keeps its independence and triggers congruence (see example 6):

|     |                            |  |                            |                    |
|-----|----------------------------|--|----------------------------|--------------------|
| (5) | BR <i>o</i><br>det.nom.m   | <i>dil-es-ker-o</i><br>fool-obl.m.-gen-m | <i>kher</i><br>house.nom.m | ‘loony bin’        |
| (6) | BR <i>le</i><br>det.obl.m. | <i>dil-es-ker-o</i><br>fool-obl.m-gen-m  | <i>kher</i><br>house.nom.m | ‘the fool’s house’ |

The most important German model for calquing verbs are particle verbs. The German particles, which are derived from spatial adverbs, are rendered in BR with the corresponding spatial adverbs, e.g. *tel pisin-* literally under-write ‘to sign’ ← ger. *unterschreiben*. Alongside borrowing of verbal stems, calquing German particle verbs is the most productive procedure to express new verbal meanings in BR.

Beyond borrowing and calquing, semantic extension and internal derivation are two additional processes of word formation in BR.

### 3.2. The elaboration corpus

There are profound linguistic differences between the codification corpus and the elaboration corpus concerning lexicon, morphology and syntax. This is due to the paradoxical situation of BR: with respect to everyday language use, BR is on the edge of extinction; with respect to written text production, there is a pressing need to create lexical material to express new concepts. While the above described strategies are still in use, the process of creating new lexical material is now being clearly affected by the declining language competence. Difficulties in lexical recall and insecurity in lexical and morphological judgements lead to a number of phenomena, among them competing lexical forms, generic usage of terms, increased polysemy and heavy borrowing that jeopardises the consistency of the BR declension system. Most of these phenomena have also been observed in language attrition situations (see the discussion for N|uu in Sands, Miller & Brugman 2007).

#### 3.2.1. Competing lexical forms

One effect of difficulties in lexical recall is the creation of new lexical forms for concepts that have already been expressed in BR. Competing forms can occur through borrowing, calquing and internal derivation.

For the concept *shoulder* BR has the Indo-Aryan word *phiko*, the Slavic *lapicka* (borrowed via Hungarian *lapocka*), and the German *šulteri* (from German *Schulter*). As there is no longer active competence in Hungarian among most speakers of BR, the first two words are considered as the ‘true BR’ words for *shoulder*, even though the German based form is by far the most frequent form in all recent records of the language.

We also find three words for the concept *appointment*: *arvakeripe* (a nominalisation of *ar vakerele* ‘to arrange, to settle, to fix’, literally *out-speak*, calqued on German *ausreden* ‘to arrange, to settle, to fix’), *telvakeripe* (a nominalisation of *tel vakerele* ‘to arrange, to settle, to fix’, literally *down-speak*, calqued on German *absprechen* ‘to arrange, to settle, to fix’), and *arkeripe* (a nominalisation of *ar kerel* ‘to arrange, to settle, to fix’, literally *out-make*, calqued on German *ausmachen* ‘to arrange, to settle, to fix’)<sup>5</sup>. At present it is impossible to judge whether all forms for this concept will be retained, or if one of the synonyms will begin to dominate.

Finally, in BR there are two abstract nouns, *džumipe* and *džuminipe*, which are related to the

<sup>5</sup> Note that all three calques are based on one particular context-sensitive meaning of the German model words.

loan verb *džuminel* ‘to press’. The lexemes are used for calques on German nouns containing *-druck*, *-drücken* ‘pressure, print, crush, suppression’, both with a concrete and an abstract meaning. The old form *džumipe* is the most frequent form, occurring in older as well as recent records of the language. It can hardly be considered a derivation from *džuminel* since it lacks *-in-*, the loan verb integration marker. The form *džuminipe* on the other hand is a regular derivation from *džuminel*, occurring exclusively in very few texts in the elaboration corpus as well as in elicitation situations with speakers with metalinguistic awareness. This insinuates that the latter is consciously derived from the verb *džuminel* on those occasions where the original form is not accessible to the speaker.

### 3.2.2. Generic usage of terms

While it is common for Romani that words are replaced by borrowings from the respective current contact language, the maintenance of three synonyms from three different lexical layers (*phiko* / *lapicka* / *šulteri* ‘shoulder’, see above) is unusual. Further enquiry with older BR speakers reveals a semantic differentiation: *phiko* designates the upper part of the shoulder, *lapicka* refers to the omoplate, *šulteri* means ‘shoulder’. However, this distinction could not be verified in actual text production. In documented language use *šulteri* is by far the most frequent form and has the same semantic scope as German *Schulter* ‘shoulder’. Occurrences of the other two lexemes show that their meaning has also been adapted to the German model. This means that under strong German influence, the differentiated two older lexemes became associated with the more generic meaning of German *Schulter*. Since the semantic differentiation of terms is now obsolete, the process of replacing older forms by more recent borrowings is currently taking place, as can be seen by the high frequency of *šulteri* and the gradual disappearance of *phiko* and *lapicka*.

### 3.2.3. Increased polysemy

In BR increased polysemy is the result of the wish to fill lexical gaps without using any borrowed material. It occurs especially with lexemes from older lexical layers. There seem to be two variant processes that lead to increased polysemy in BR: First, lexemes that designate basic concepts (e.g. basic qualities such as good or bad, or basic verbs such as to do/make or to go) acquire new meanings, because they mirror the distribution of the German equivalent. An example of this process is the inherited word *lačo*: the basic meaning of the word in the codification corpus is ‘good’. In data from the elaboration corpus and in our elicitation the word *lačo* acquires – among others – the meanings ‘well-behaved’, ‘immaculate’, ‘regular’, ‘right’, ‘able’, ‘ripe’ and ‘fertile’. These meanings of *lačo* are highly context-sensitive, if not restricted to one single context. In German each of these meanings can be expressed by one or several different words, which can all be replaced by German *gut*, which is the equivalent of *lačo*.

Second, abstract nouns ending in *-ipe* are prone to the acquisition of new meanings. The derivation of abstract nouns in *-ipe* is one of the most productive processes in BR. Since the functional expansion into acrolectal domains involves a wide range of abstract concepts, polysemy is particularly widespread in derivations in *-ipe*, as one single derivation is used for all abstract concepts related to the meanings of the derivational base. The word *kamipe* < *kamel* ‘to love, to want, to owe’ is attested with such meanings as ‘wish’, ‘intention’, ‘love’, ‘debt’, ‘obligation’, ‘love affair’, ‘attachment’ and ‘passion’, the word *khetanipe* ‘being together’ < *khetan(e)* ‘together’ acquires the meanings ‘partnership’, ‘relationship’ and ‘consensus’, *maripe* < *mar-* ‘to beat’ originally meant ‘brawl’. In data from the elaboration corpus and elicitation *maripe* is now used to mean ‘violence’.

While increased polysemy has been discussed in the context of language attrition for other small languages (e.g. Sands, Miller & Brugman 2007: 61f., Fabunmi & Salawu 2005: 404), increased polysemy in BR is a result of the need to express an increasing number of new concepts and therefore has to be seen in the context of language revival.

### 3.2.4. Heavy borrowing

If no suitable BR word is available or if the speaker does not mind using foreign linguistic material, lexical gaps are filled by borrowing. Due to the above described integration morphology of Romani, borrowings are easily integrated without inflicting damage to the structural integrity of the language. However, the increasing insecurity in lexical / morphological judgements results in a rising number of loan words that are integrated under violation of the consistency of word formation rules. Inserting German words without any morphological integration is a valid strategy both for masculine and feminine nouns as long as they end with a consonant or if their final vowel fits the respective gender specific loan word declension class.

According to the rule that German neuter becomes masculine in BR, the neuter word *Thema* ‘topic’ is assigned masculine gender. Since the word is inserted without morphological integration, its ending *-a* is homophonous with the integration marker *-a* that assigns a lexical item to the feminine loan word declension class. This indicates a fundamental change with respect to weighting of the criteria for gender assignment in BR. While phonetic shape was once the principal criterion (see *bika* (f.) ‘bull’), now the gender in the donor language seems to be the only criterion. Even more problematic is the integration of masculine loan words with *-a*: The word *o diploma*<sup>6</sup> ‘diploma’ ← ger. *Diplom* (n.) is integrated with *-a*, but following the rule that the loan in BR has the same gender as in German, or as in the case of neuter masculine gender, it is masculine in BR. The integration marker *-a*, which is reserved for feminine loans, and the masculine gender of the word are in contradiction and make it impossible to assign the word to a particular inflection class. Similarly, non-integrated words ending in *-e* (e.g. *mite* (f.) ← ger. *Miete* (f.) ‘rent’, *pile* (f.) ← ger. *Pille* (f.) ‘pill’) cannot be assigned to a particular declension class, since there is no suitable class.

Another example of insecurity in lexical judgement concerns the use of the morpheme *-itiko* as default integration marker for German adjectives ending in *-isch* in recent data, e.g. *erotitiko* ‘erotic’ (from German *erotisch*). As mentioned above, in the codification corpus *-itiko* is reserved for adjectives designating a geographic or ethnic provenience. Due to insecurity in lexical judgements speakers often cannot say whether a given lexical form is acceptable or not which poses a practical problem to a lexicographer working in an ongoing lexical expansion process.

## 4. Lexicography of BR

BR is lexicographically documented in ROMLEX (henceforth RL), which is an extendible multi-dialectal lexical database with a freely accessible web-interface (<http://romani.uni-graz.at/romlex/>). The BR data contained in RL at present stem from the codification corpus (recordings of oral texts from the 1960s and from 1993-1996) and lexicographic elicitation. It is being continuously expanded with data from the so-called elaboration corpus, which is a growing collection of published written texts (law texts, declarations, bible texts, contributions to periodicals etc.), most of which are translations from German. These texts

<sup>6</sup> The choice of the integration marker *-a* for this word might be triggered by the internationally used term *diploma*.



have a high emblematic value for the speech community. Thus, RL unites lexical data from both the functionally restricted and the expanded varieties of BR.

The idea for a print dictionary of BR comes from the authorities of the province of Burgenland who wish to archive their regional minority languages and have offered financial support for this project. Likewise, it was an urgent goal of the speech community already at the outset of the codification process in 1993 to have a printed bilingual (BR-German, German-BR) reference dictionary of their language. The realisation of the project is being conducted at the University of Graz in cooperation with a local NGO that represents the Burgenland Roma.

There are different requirements for RL and the print dictionary of BR. RL is intended as a tool for comprehensive lexical documentation of BR. At the same time, it is a practical, low-threshold tool for text producers and provides up-to-date materials of BR. This is especially important at the stage of corpus elaboration. BR materials are being continuously updated in RL meaning that text producers who use RL are likely to use existing lexemes instead of creating new terminology. From a language planning perspective this means that RL offers the opportunity to unify terminology during the process of lexical modernisation. The print dictionary, on the other hand, primarily serves an emblematic purpose. Besides a grammar of the language and teaching materials, a print dictionary is another prestigious object which demonstrates the equality of BR with other languages. From a status planning perspective, the print dictionary marks an important step in the officialisation of BR. With respect to Hornberger's (2006) integrative model of language policy and planning (see table 1), RL is affiliated to cultivation planning and corpus planning and assists the elaboration of BR. The print dictionary, on the other hand, is affiliated to policy and status planning.

Given the differing purposes of RL and the print dictionary, different strategies are used in lexicographic decision-making. Roughly speaking, RL favours an inclusive descriptive approach while the print dictionary is rather restrictive and follows normative principles. Trap-Jensen (2002) argued against the possibility of purely descriptive lexicography since lexicographers always deal with the question of whether to 'include a particular item or not, and if so in what form' (Trap-Jensen 2002: 503). He brings up the issues of lemma selection, orthography and meaning. All of these have to be dealt with both for RL and the print dictionary. The respective decisions inevitably contribute to establishing norms for BR. The following sections illustrate the lexicographic decision making process for RL and the print dictionary of BR, respectively.

#### **4.1. Possibilities, requirements and strategies for ROMLEX**

RL offers the possibility to document the ongoing process of functional expansion of the BR lexicon in a very flexible way. The technical make-up of RL allows for continuous integration of new material. Individual lemmas as well as new meanings for existing lemmas can easily be added to the RL database. RL is designed as a multi-dialectal lexicographic database which is accessible via a web-interface (<http://romani.uni-graz.at/romlex/>). Since there is no common writing system for Romani varieties, the orthography used in RL is a phoneme-based writing system (see <http://romani.uni-graz.at/romlex/wsphonemic.xml>).

The BR data in RL at present reflect the basic vocabulary drawn from the codification corpus of BR. Recently, the vocabulary of BR has been expanded by elicitation of lexemes in order to fill semantic gaps that had previously been identified on the basis of a list of approximately 5,000 lexemes. The next step in documenting the lexicon of BR is the analysis of the

continuously growing elaboration corpus. As already mentioned above, the texts of this corpus differ in lexicon and on the structural level from texts of the codification corpus: the texts of the elaboration corpus are characterised by a significantly higher occurrence of nominalisations, by restructuring of categorial semantics (e.g. ablative ending is used in possessive function) and ad-hoc-verbalisations to designate concepts which belong to domains for which BR had formerly not been used.

#### 4.1.1. Orthography

The use of a particular orthography in any kind of dictionary is always a normative act. The case of BR is problematic, since the writing system used for all Romani varieties in RL does not correspond to the writing system used in the BR community. Text producers in BR use a writing system that has been developed during the codification process of BR. This writing system draws heavily on German orthography. The BR word for ‘children’ is spelled *fačuvča* in RL, but *fatschuvtscha* by BR text producers. The use of a consistent writing system for all Romani varieties enables linguists to conduct cross-dialectal research. Furthermore it allows non-Romani speakers to correlate graphemes with sounds. However, producers of BR texts perceive the use of the RL writing system for BR as a disadvantage and therefore are less willing to use it. The rejection of RL as a practical tool for BR text production impairs the potential of RL for terminology unification. Therefore, a tool has been developed to transpose the RL writing system to the German-based orthography.

#### 4.1.2. Lemma selection

Since it is a core function of RL to document the functional expansion of BR, the lemma selection process for RL must not be too restrictive. Therefore, competing lexical forms are to be included in RL. We are aware that this decision initially means that we cannot guarantee the validity of each individual lexical entry. Through ongoing lexical analysis of the growing elaboration corpus, we will be able to determine the current status (frequency of usage and thereby acceptance) of the respective words. RL caters for all the requirements of the peculiar data set from an elaboration corpus: it is expandable and can be edited anytime, allowing for all necessary modifications. However, lexemes that display an obvious violation of morphological rules (see 2.2.4. for examples) are not included in RL. There are many instances of words that do not fit any declension class both in elicited data and in data from the corpus. Furthermore, loan integration of German adjectives which do not designate ethnic or geographic provenience with *-itiko* occurred several times in data from elicitation. The respective words are discussed with informants who have been involved in the ongoing expansion processes and have meta-linguistic awareness of the integration rules. In addition, problematic formations in combination with rule-conform forms are presented to native speakers without meta-linguistic awareness in order to assess the acceptability of the individual forms. Following the judgement of competent native speakers, the form *erotiši* instead of *erotitiko*, the form *diplom* (m.) instead of *diploma* (m.) and the forms *pila* (f.) and *mita* (f.) instead of *pile* (f.) and *mite* (f.) are included in RL.

#### 4.1.3. Meaning

In the process of functional expansion existing words are used in new contexts and thereby acquire new meanings. As shown in the examples above (see 2.2.2) this can result in increased polysemy. The strategy so far is to include the meanings of the lexemes in the codification corpus and meanings given in elicitation. Elicitation is obviously a suboptimal method to document the full range of meaning and the current use of a word. Svensén (2009:55f.) points out that for bilingual dictionaries this is most successfully achieved by analysing parallel text corpora. We therefore use the parallel elaboration corpus to verify and

complement the meanings and contexts of lexemes contained in RL. Backing up our data with examples from a parallel corpus allows us to deal with increased polysemy in a non-restrictive way. Very context specific meanings can be illustrated by usage phrases.

#### **4.2. Possibilities, requirements and strategies for the print dictionary**

While the possibility of continuous re-evaluation and correction of the data on the basis of the growing elaboration corpus allows the adoption of a non-restrictive strategy for RL, lexicographic decisions for the print dictionary have to be taken more carefully. The print dictionary has a deadline, and once printed, the contents cannot be modified, except in further editions. In our case, the main function of the print dictionary is emblematic and not practical usage. It is therefore crucial that the print dictionary unites lexical items which the majority of the speech community can identify with, thereby preserving its symbolic value in the years to come. Furthermore, it has a considerable prescriptive authority due to its nature as a print medium. The following section discusses how to establish criteria for a print dictionary of BR which guarantee the validity and acceptance of the individual lexical entries, without entirely suppressing evidence from the ongoing expansion process.

The print dictionary is based on RL and will offer information that is already contained in the lexical database but not yet accessible via the web-interface (morphological information, word-families, etymology). It contains the lexical analysis of the codification corpus and vocabulary elicited with two informants to fill previously identified semantic gaps. One of the informants uses BR alongside dialectal German in his social microcosm (family, Roma neighbours), while the other, a highly experienced translator (German to BR), also uses BR in formal domains due to his function as the head of a Roma NGO in Burgenland. Diverging results from the elicitation were checked against the elaboration corpus, in order to assess the respective validity. Identical results were included in the print dictionary provided they do not contradict BR word formation rules.

##### **4.2.1. Orthography**

In the print dictionary BR is written using the German-based writing system as the speech community prefers this orthography. The dictionary will also contain an index that displays the RL writing system parallel to the German-based orthography.

##### **4.2.2. Lemma selection**

In consideration of the mainly emblematic purpose of the print dictionary, lemma selection is a very delicate and important matter. On the one hand the dictionary is intended to document the lexical wealth of BR as a fully fledged language; on the other hand it should only contain words which all members of the community consider as their own language. Recent loans from the current contact language are particularly problematic with respect to lemma selection, since some speakers would not categorise them as BR but as German and would therefore object to their inclusion in a BR print dictionary. The mere existence of a print dictionary is more important than the inclusion of all lexical items occurring in the process of functional expansion. For this reason and because a wide acceptance in the community is desirable, a quite restrictive approach is to be taken.

Therefore we have decided to include all attested BR words of non-German origin and loans that are based on German dialectal forms (as typical for German loans in the codification corpus). For all other lexemes of German origin their frequency of occurrence in the BR corpus is checked. The final decision whether a word is to be included in the print dictionary is taken on the basis of the semantic domain a lexeme belongs to. Loans that belong to

domains of everyday life (e.g. household, work, school, sports) are to be included in the print dictionary, while more specialised vocabulary (technology, politics, social sciences) is treated in a more restrictive way. To give an example: a word like *kompjuteri* ‘computer’ will be included in the print dictionary, as it describes an entity that today is part of everyday life, while a loan like *sterijotip* (m.) ‘stereotype’ will not be included in the print dictionary.

The next problem we face is the question of how many synonyms for each concept in a print dictionary of BR should be included. RL can easily handle many synonymous forms. For the print dictionary, however, it is necessary to reduce the number of forms and to choose forms that are likely to enhance the speakers’ acceptance of the dictionary. With respect to the example quoted above – *phiko* < Indic, *lapicka* < Hungarian, *šulteri* < German, the following strategy is chosen: If the synonyms belong to different historical layers the speakers’ judgement as to what is part of their language has to be the most important criterion. Thus *phiko* and *lapicka* are to be included in the print dictionary while *šulteri* will be dropped despite its high frequency. To put it more generally: Inherited words and words from former contact languages without active competence in the speech community are retained, while synonyms from the current contact language are dropped.

The synonyms *arvakeripe*, *telvakeripe* and *arkeripe* are mapped on different German synonyms for ‘appointment, arrangement’. They are an instance of neologism formation by mapping BR material on German words. As the example above shows, depending on the German model word in the speakers’ mind, we find synonymous formations based on different BR material. Due to the high productivity of the process and the permanent availability of all the models, each form is equally likely to reoccur. Since the material used for code mapping is exclusively BR, speakers perceive all three forms as part of their own language. Subsequently, all three forms will be included in the print dictionary.

Another example of competing lexical forms are synonyms that originate from internal derivation. In BR we find two abstract nouns, *džumipe* and *džuminipe*, which are related to the loan verb *džuminel* ‘to press’. The lexemes are used for calques on German nouns containing *-druck*, *-drücken* ‘pressure, print, crush, suppression’, thus both in a concrete and an abstract meaning. The old form *džumipe* is the most frequent form, occurring in older as well as recent records of the language. It can hardly be considered a derivation from *džuminel* since it lacks *-in-*, the loan verb integration marker. The form *džuminipe* on the other hand is a regular derivation from *džuminel*, occurring exclusively in very few texts in the elaboration corpus as well as in elicitation situations with speakers with meta-linguistic awareness. This insinuates that the latter is consciously derived from the verb *džuminel* on those occasions where the original form is not accessible to the calquing speaker. Due to the decline in everyday language use and the corresponding difficulties in lexical recall, we expect that the frequency of *džuminipe* will increase. In the elaboration corpus both forms are used for calques on German nouns containing *-druck*, *-drücken* ‘pressure, print, crush, suppression’. Thus both forms are included in RL, but only the more widely accepted old form *džumipe* is included in the print dictionary.

#### 4.2.3. Meaning

Complementary to the introduction of many synonyms for one concept is the accumulation of many meanings in one lexeme. Expanding the semantic extension of a given lexeme is a common strategy for expanding the lexicon of functionally restricted minority languages. Here the semantic range of words of high identificatory value is expanded, instead of actual transfer of foreign matter from a contact language. However, in the case of BR one has to

carefully examine the type of semantic expansion at work in a given lexeme.

An example of this process is the inherited word *lačo*: the basic meaning of the word in the codification corpus is ‘good’. In data from the elaboration corpus and in our elicitation the word *lačo* acquires – among others – the meanings ‘well-behaved’, ‘immaculate’, ‘regular’, ‘right’, ‘able’, ‘ripe’ and ‘fertile’. These meanings of *lačo* are highly context-sensitive, if not restricted to one single context. In German each of these meanings can be expressed by one or several different words, which can all be replaced by German *gut*, which is the equivalent of *lačo*. Therefore, these meanings are not given as translations of *lačo* in the print dictionary. Nonetheless, these meanings are illustrated by usage phrases that exemplify further uses of the headword *lačo*.

In the word *maripe* we can observe another type of semantic expansion at work: in the codification corpus the word has the meaning ‘brawl’, deriving from the verb *marel* ‘to beat’. The suffix *-ipe* is productively used in BR in the derivation of abstract nouns from lexical items with concrete meanings. In data from the elaboration corpus and the elicitation, *maripe* is now used to mean ‘violence’. Hence, in this semantic expansion the original abstract dimension is taken one step further, which is congruent with the functional expansion into formal domains. Therefore, this new meaning is listed as a translation of *maripe* in the print dictionary.

While all new meanings of *kamipe* i.e. ‘wish’, ‘intention’, ‘love’, ‘debt’, ‘obligation’, ‘love affair’, ‘attachment’ and ‘passion’ can be related to the meanings of the derivational base *kamel* (‘to love, to want, to owe’), the meaning ‘consensus’ for the word *khetanipe* < *khetan(e)* ‘together’ is quite distant from the original meaning of the word and the derivational base. Therefore, this meaning will not be included in the BR dictionary.

## 5. Conclusion

The different decisions taken for the lexical database and for the print dictionary arise from the different roles they have to assume. The print dictionary gives the authorities the possibility to archive BR the same way as the other regional languages in Burgenland have been. For the speech community, on the other hand, this equalisation of BR is part of the emblematic value of the print dictionary. Due to the authority of the printed word the print dictionary is likely to become the reference work for BR with a homogenising impact concerning spelling, abbreviations, hyphenation etc. A print dictionary is also an important symbol of the officialisation of BR.

The homogenising impact of RL is situated on a different level. As an easy to use resource it is the primary tool for every day usage. Instead of creating new words, text producers select existing forms that are documented in RL, which prohibits the proliferation of new formations i.e. it assists terminology unification in the ongoing lexical modernisation process. The lexical analysis of texts produced in this way, on the other hand, allows the determination of which of the once created and documented lexemes are actually used and accepted by the speech community. The functional expansion of European national languages occurred over several centuries and partly paralleled extralinguistic changes (development of technology, politics etc.). Functional expansion of non-state languages, on the other hand, is demanded – by the speech community itself as well as by decision makers – within a much shorter time span. Taking a two-fold lexicographic approach, as in the case of BR, is the only way to prevent it being too restrictive or too liberal with respect to new formations.

## References

- Bakker, P. (1997). 'Athematic Morphology in Romani: The Borrowing of a Borrowing Pattern'. In Matras, Y; Bakker, P.; Kyuchukov, H. (eds.). *The Typology and Dialectology of Romani*. Amsterdam/Philadelphia: John Benjamins. 1-22.
- Baumgartner, G.; Freund, F. (2004). *Die Burgenland Roma 1945-2000. Eine Darstellung der Volksgruppe auf der Basis archivalischer und statistischer Quellen. Ein Forschungsprojekt des Kulturvereins Österreichischer Roma*. Eisenstadt: Burgenländische Forschungen 88.
- Boretzky, N.; Iglá, B. (1994). *Wörterbuch Romani-Deutsch-Englisch für den südosteuropäischen Raum : mit einer Grammatik der Dialektvarianten*. Wiesbaden: Harrassowitz.
- Fabunmi, F.A.; Salawu, A.S. (2005). 'Is Yorùbá an Endangered Language?'. In *Nordic Journal of African Studies* 14 (3). 391-408.
- Halwachs, D.W. (1998). *Amaro vakeripe Roman hi. Unsere Sprache ist Roman. Texte, Glossar und Grammatik der burgenländischen Romani-Variante*. Klagenfurt/Celovec: Drava Verlag.
- Halwachs, D. W. (2001). 'Das Burgenland-Romani zwischen Anerkennung und Sprachtod'. In *ÖGL* 45 (4). 211-218.
- Hornberger, N.H. (2006). 'Frameworks and Models in Language Policy and Planning'. In Ricento, T. (ed.). *Language Policy. Theory and Method*. London: Blackwell. 24-41.
- Sands, B.; Miller, A; Brugman, J. (2007). 'The Lexicon in Language Attrition: The Case of N|uu'. In Payne, D.; Peña, J. (eds.). *Selected Proceedings of the 37th Annual Conference on African Linguistics*. Somerville: Cascadilla Proceedings Project. 55-65.
- Svensén, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-making*. Cambridge: Cambridge University Press.
- Trap-Jensen, L. (2002). 'Descriptive and Normative Aspects of Lexicographic Decision-Making: The Borderline Cases'. In Braasch, A.; Povlsen, C. (eds.). *Euralex 2002. Proceedings I-II*. København. 503-509.