Morphology generation for Swiss German dialects

Scherrer, Yves

How to cite

SCHERRER, Yves. Morphology generation for Swiss German dialects. In: Systems and Frameworks for Computational Morphology: (Second International Workshop, SFCM 2011). Mahlow, C. ; Piotrowski, M. (Ed.). Zurich. [s.l.] : Springer, 2011. p. 130–140. (Communications in Computer and Information Science) doi: 10.1007/978-3-642-23138-4_9

This publication URL: https://archive-ouverte.unige.ch/unige:22778

Publication DOI: 10.1007/978-3-642-23138-4_9

Morphology generation for Swiss German dialects

Yves Scherrer

LATL
Université de Genève
Genève, Switzerland
yves.scherrer@unige.ch

Abstract. Most work in natural language processing is geared towards written, standardized language varieties. In this paper, we present a morphology generator that is able to handle continuous linguistic variation, as it is encountered in the dialect landscape of German-speaking Switzerland. The generator derives inflected dialect forms from Standard German input. Besides generation of inflectional affixes, this system also deals with the phonetic adaptation of cognate stems and with lexical substitution of non-cognate stems. Most of its rules are parametrized by probability maps extracted from a dialectological atlas, thereby providing a large dialectal coverage.

Keywords: Swiss German, dialects, cognate words, morphological generation

1 Introduction

Most work in natural language processing is geared towards written, standardized language varieties. This focus is generally justified on practical grounds of data availability and socio-economical relevance, but does not always reflect the linguistic reality of sub-standard varieties. In this paper, we present a morphology generator that is able to handle continuous linguistic variation, as it is encountered in various dialect landscapes. The work presented here is applied to Swiss German dialects; these dialects are well documented by dialectological research and are among the most vital ones in Europe in terms of social acceptance and media exposure.

The task of Swiss German word generation can be formulated as follows: *Given a Standard German root and a set of morphosyntactic features, generate all inflected forms that are valid in the different Swiss German dialects.*

Our approach can be qualified as *cross-lingual* and *multi-dialectal*. It is cross-lingual in the sense that the language variety of the input root (Standard German) is different from the language variety of the output forms (Swiss German). It is multi-dialectal because it aims to generate all forms that occur in the different dialects of German-speaking Switzerland, relying on existing dialectological resources. Hence, the proposed system is more than just a morphological generator: it is a word translation engine that relies on the numerous structural similarities between Standard German and Swiss German.

In the following section, we will briefly describe some linguistic characteristics of Swiss German dialects. In section 3, the general system architecture will be described and illustrated with examples. Section 4 will present some problematic cases that arise from the specific multi-dialectal conception of our model. We show some coverage figures in section 5, and we conclude in section 6.

2 Swiss German dialects

The German-speaking area of Switzerland encompasses the Northeastern two thirds of the Swiss territory. Likewise, about two thirds of the Swiss population declare (any variety of) German as their first language.

It is usually admitted that the sociolinguistic configuration of German-speaking Switzerland is a model case of *diglossia*, i.e. an environment in which two linguistic varieties are used complementarily in functionally different contexts. In German-speaking Switzerland, dialects are used in speech, while Standard German is used nearly exclusively in written contexts.

Despite the preference for spoken dialect use, written dialect use has become popular in electronic media like blogs, SMS, e-mail and chatrooms. The Alemannic Wikipedia contains about 6000 articles, among which many are written in a Swiss German dialect.[1] However, all this data is very heterogeneous in terms of the dialects used, spelling conventions and genres.

The classification of Swiss German dialects is commonly based on administrative and topographical criteria. Although these non-linguistic borders have influenced dialects to various degrees, the resulting classification does not always match the linguistic reality. Our approach does not presuppose any dialect classification. We conceive of the Swiss German dialect area as a continuum in which certain phenomena show more clear-cut borders than others. The nature of dialect borders is to be inferred from the data.[2]

Swiss German has been subject to dialectological research since the beginning of the 20th century. One of the major contributions is the *Sprachatlas der deutschen Schweiz* (SDS), a linguistic atlas that covers phonetic, morphological and lexical differences. Data collection and publication were carried out between 1939 and 1997 [8]. There also exist grammars and lexicons for specific dialects (for instance, [4,11,6,5]), as well as general presentations of Swiss German [10,12].

On all levels of linguistic analysis, there are differences between Standard German and Swiss German, as well as among the various Swiss German dialects. The examples given in the following sections will illustrate some of these differences.

[1] See `http://als.wikipedia.org`. Besides Swiss German, the Alemannic dialect group encompasses Alsatian, South-West German Alemannic as well as the Vorarlberg dialects of Austria.

[2] Nonetheless, we will refer to political entities for convenience when describing interdialectal differences in the following sections of this paper.

3 General system architecture

As sketched out above, the morphological generator for Swiss German takes as input a Standard German root and a set of features that determine the inflected form to be generated. These features include part-of-speech tags, morphological tags like gender, number, person, as well as lexical information like inflection class.

Full dialect forms are generated in two steps. First, a dialect root is obtained by applying phonetic and lexical transformations to the Standard German root[3]. Second, the inflected dialectal form is obtained by adding affixes to the dialectal root, according to the feature set given in the input.

For example, the Standard German verbal root *such-* 'to search' will trigger the following root transformation rules for Graubünden dialect:

- *u* → *u* (not *ü*)
- *u* → *ue*
- *e* → *a* (in diphthong)

The result of the first step is thus the dialect root *suach-*. In order to generate the 3rd person plural form, the feature set `VVFIN-3.Pl.Pres.Ind` will trigger the affixation rule with the Graubünden dialect suffix *-end*.[4] The result of the second step is thus the inflected form *suachend*.

In most common settings of morphology generation, the first step is not required. Morphology generators are usually conceived as monolingual tools, where the language variety of the input is the same as the language variety of the output. This contrasts with our setting, where the input root is not identical with the root of the output form: it can undergo phonetic and lexical transformations depending on the dialect. Hence, additional transformations have to be executed before even starting affix generation. This is a consequence of our cross-lingual approach.

Our aim of multi-dialectal coverage leads to further complications. In most cases, one Standard German root will yield several dialect roots, each of which is valid in a different region. Likewise, dialectally different affixes can be added to each dialect root. Therefore, one generation query will usually yield a long list of candidate forms. However, all candidates are associated with maps, extracted from the SDS atlas, that describe the geographic area in which they are valid (see Figure 1). These maps allow to prune the candidate lists according to a specific target dialect.

Moreover, the first and second steps cannot always be clearly separated. In the case of irregular inflection or suppletive forms, the most efficient approach is to combine the two steps. Section 4.2 will show an example of this.

The system presented here is implemented in the form of a database which contains different types of transformation rules, which are applied in cascade with the help of Python scripts. While this approach allows for easy debugging

[3] Here, we simply define the root of a word as identical to its citation form, except for verbs, where specific infinitive affixes are stripped off.

[4] We use the STTS tag set as defined for Standard German [15].

of the rule base, it is not as efficient as an implementation with finite-state transducers. Moreover, transducers could be used in both directions, for analysis and generation. Such an implementation is planned for future work.

Our work does not currently use machine learning techniques. There are two main reasons for this methodological choice. First, the dialectological atlas used as primary resource already contains linguistically interpreted data: the legend of each map specifies the rule and its conditions of applicability in a relatively explicit way. By using hand-written rules, we can fully take advantage of these data. Second, machine learning approaches[5] require a lot of training data, which are notoriously hard to find for small-scale language varieties like Swiss German dialects. The problem is exacerbated with our multi-dialect approach, where distinct training corpora would be required for each dialect.

In the following subsections, we will present the architecture of the database that contains the transformation rules.

3.1 Variables and variants

The structure of our database relies on the dialectological distinction between the concepts of *variable* and *variant* [2, pages 49ff]. A variable is any linguistic phenomenon whose realisation varies along the geographical axis. The different realisations are called variants.[6]

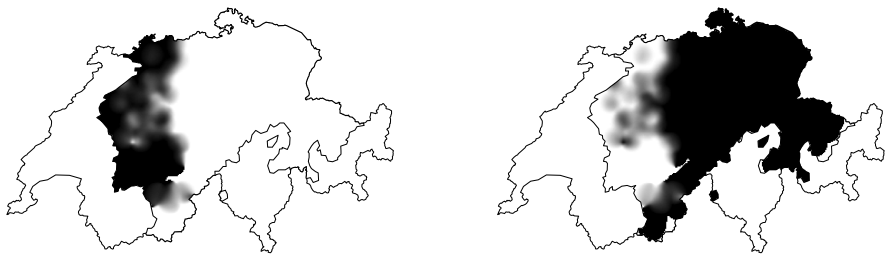

Fig. 1. Maps defining the distribution of the weak nominative singular adjective suffixes. Black surfaces represent high probabilities, white surfaces represent low probabilities. The *i*-suffix appears in Western dialects (black surface in the left map), the null-suffix appears in Eastern dialects (black surface in the right map). These maps have been digitized from the SDS map III/254.

For example, the suffix of weak nominative singular adjectives is a variable in Swiss German dialects. Its variants are the null-suffix in Eastern dialects (e.g.

[5] For a recent overview of unsupervised learning techniques of morphology, see [7].

[6] The distinction between variables and variants is a consequence of the multi-dialect generation approach. In a (deterministic) single-dialect system, there is a one-to-one mapping between variables and variants, which makes the distinction irrelevant.

di schwarz Chatz 'the black cat'), and the *-i* suffix in Western dialects (e.g. *di schwarz**i** Chatz*).

Each variant is associated with a probability map that shows its geographic extension. The maps are extracted from the SDS atlas [8]. The SDS maps contain discrete values at a limited number of inquiry points. These values are converted to a continuous surface by interpolation, such that the grey scale value at each pixel of the surface represents the probability of a variant at that pixel. For any variable, the maps of all its variants are complementary, i.e. the sum of the probability values of all maps at each pixel equals to 1. More details about the interpolation method can be found in [13] and [14]. Figure 1 shows the probability maps for the example given above.

3.2 Phonetic transformations

Most Swiss German words are cognates of Standard German words. Hence, regular phonetic[7] transformations allow us to derive many Swiss German word roots from their Standard German counterparts. Phonetic transformation rules are stored in two database tables, `phoneticVariables` and `phoneticVariants`. Example entries are given in Table 1.[8]

`phoneticVariables:`

`rulename`	`regex`	`priority`	`datafolder`
2-120-nd	[aeiouäöü](nd)$	101	2-120

`phoneticVariants:`

`rulename`	`replace`	`mapname`
2-120-nd	ng	dp_ng
2-120-nd	nn	dp_nn
2-120-nd	nt	dp_nt
2-120-nd	nd	dp_nd

Table 1. Detail of the database tables for phonetic transformations. The example shows the rule *2-120-nd* that transforms post-vocalic word-final *nd* into one of the four variants *ng* (Bern), *nn* (Fribourg), *nt* (Wallis), or *nd* (other dialects).

Each phonetic transformation variable is characterized by a name (`rulename`), a regular expression that allows to identify its contexts of application (`regex`), an

[7] In the presence of multiple dialects, it is somewhat difficult to distinguish phonetic from phonological phenomena. A specific sound difference may have phonemic value in one dialect, but not in another. Hence, the same sound law would be classified under phonetics in one dialect, and under phonology in another. In this paper, we use *phonetic transformation* as a generic term for both types of transformations.

[8] This table, as well as the following ones, is to be read as follows: The first line describes the structure of the table. The entries below the double line show examples.

integer determining the rule order (`priority`), and a file system path in which the corresponding maps are to be found (`datafolder`).

Each variable has one or more variants, which are linked by the `rulename` attribute. Variants are defined by the string which replaces the group matched by the regular expression (`replace`) and the file name of the corresponding map (`mapname`). Currently, 135 phonetic variables are implemented. They correspond to 314 variants.

3.3 Lexical replacement

Cognate words can be derived from Standard German stems with the help of the phonetic rules described above. However, there are cases where dialects use stems with a different etymological origin. In other cases, the dialectal form does have some vague phonetic resemblance with the Standard German counterpart, but which would be difficult to capture with a phonetic rule. Therefore, we introduce lexical replacement rules, which are again defined in a `lexicalVariables` table and a `lexicalVariants` table. Examples are given in Table 2.

`lexicalVariables:`

`rulename`	`lemma`	`tag`	`feats`	`datafolder`
nichts	nichts	PIS		4-171-nichts
immer	immer	ADV		6-026-immer

`lexicalVariants:`

`rulename`	`form`	`mapname`	`addFeats`	`phonRulesAfter`
nichts	nüüt	dp_nüüt		üü-ii
nichts	nüt	dp_nüt		ü-i
nichts	nünt	dp_nünt		
nichts	nütz	dp_nütz		
nichts	nix	dp_nix		
immer	immer	dp_immer		
immer	geng	dp_geng		
immer	all	dp_all		

Table 2. Detail of the database tables for lexical substitution. The first example shows the dialectal variants of Standard German *nichts* 'nothing', whose idiosyncratic behavior is difficult to capture with phonetic rules. Its first two entries generate the forms *nüüt* and *nüt*, which are further transformed to *niit* and *nit* with the help of the regular phonetic rules *üü-ii* and *ü-i* in some dialects. The second example refers to the translation of Standard German *immer* 'always', where completely different lexemes are used throughout the Swiss German dialect area.

The `lexicalVariables` table allows to specify the lemma of a word and its part-of-speech tag, but can also contain finer-grained morphological information (field `feats`). The latter is mainly used for irregular inflection patterns (see section 4.2).

The `lexicalVariants` table specifies a stem that completely replaces the Standard German stem. It also allows to modify morphological features in the field `addFeats`. This functionality is used for example to change the gender tag when a masculine noun lexeme is replaced by a feminine noun lexeme.

The field `phonRulesAfter` allows to specify a subset of phonetic rules to be applied after lexical substitution. This functionality is illustrated in the first example (Table 2), where two additional variants *niit* and *nit* exist. The distinction between *ü*-based and *i*-based variants is completely regular and already accounted for by a phonetic rule. Hence, this phonetic rule is added to the corresponding variants.

Currently, the database contains 260 lexical variables and 559 lexical variants. Most of them are high-frequency adverbs, pronouns and irregular verbs.

3.4 Affix generation

The tables `affixVariables` and `affixVariants` define the inflectional affixes for regular noun, verb, adjective and pronoun inflection. While most rules deal with simple suffixation, more complex affixation types are also supported. Table 3 shows examples of noun plural formation.

`affixVariables:`

rulename	tags	morpho	datafolder
n0-uml	NN	Pl,NCl_uml,NCl_uml_er,NCl_uml_e	
n1-e	NN	Pl,NCl_e,NCl_uml_e	
n1-er	NN	Pl,NCl_uml_er	
n1-ene	NN	Pl,NCl_ene	3-187-ene

`affixVariants:`

rulename	regexFind	suffix	specialInfix	mapname
n0-uml			umlaut	
n1-e	([^i])e?$	\1e∼		
n1-er	(er)?$	er		
n1-ene	(i)$	ine		dp_ine
n1-ene	(i)$	ene		dp_ene
n1-ene	(i)$	eni		dp_eni

Table 3. Detail of the affix generation tables, showing selected rules for noun plurals. Rule *n0-uml* adds umlaut to the stem vowel independently of the dialect chosen. Rules *n1-e* and *n1-er* add suffixes depending on the phonetic environment specified in `regexFind`. Rule *n1-ene* illustrates the use of dialect-dependent suffixes, with map information given in the fields `datafolder` and `mapname`. All rules depend on inflection class information as specified by the *NCl* tags. The symbol ∼ is substituted by *n* when the following word starts with a vowel, and dropped otherwise.

There are three ways of adding an affix to a stem. The simplest one is to add a suffix. In this case, the `suffix` column contains the suffix, and the `regexFind` and `specialInfix` columns remain empty. The second possibility allows to remove some material before adding the suffix: the material to be removed is specified by the regular expression in `regexFind`. If using regular expressions is not practical or not powerful enough, there is a third possibility: one can specify a particular affixation function (written in Python) in the field `specialInfix`. This functionality is used for example to add umlaut (first row in Table 3).

In the example of Table 3, the different suffixes are selected by the *NCl* feature which has been attributed to the nouns on the basis of their Standard German inflection class. However, these features can be changed. For example, Swiss German dialects tend to use umlaut plural marking more often than Standard German (e.g. *Hunde* — *Hünd* 'dogs', *Pullis* — *Pülli* 'sweaters').

Currently, the database contains 82 variables and 165 variants. It covers the inflectional paradigms of adjectives, nouns, regular verbs, determiners, and preposition-determiner combinations.

4 Problematic cases

In this section, we describe some specific problems that arise on the one hand from the multi-dialect generation approach, and on the other hand from linguistic particularities of the Swiss German dialects.

4.1 Lexical restrictions of phonetic transformations

Dialect evolution sometimes yields unpredictable results. Some rules apply to certain words but not to others, without there being a clearly identifiable cause. For example, the stem vowel in the words *Gras* 'grass', *sparen* 'to save', *Arbeit* 'work' and *Axt* 'axe' is changed to *ä* in some Northeastern dialects. It is difficult to generalize a phonetic context that might trigger these transformations. Therefore, we chose to use a "whitelist" which enumerates all the lemmas that undergo this transformation.

In other cases, the cause of a specific evolution is known, but is difficult to detect for practical reasons. For example, the two Middle High German vowels *û* and *ou* have fallen together in Modern Standard German *au*, but have remained distinct in Swiss German (*uu* and *au/ou*, respectively). In a model based on Standard German input, it is thus impossible to predict the correct Swiss German form. No phonetic cue tells us that Standard German *Haus* 'house' should become *Huus*, but that Standard German *Baum* 'tree' should remain *Baum*. Again, we use a whitelist to enumerate the lemmas in either class.

Currently, the whitelist contains a total of 13,000 lemmas associated to 39 rules. It has been compiled by a native dialect speaker, on the basis of the Derewo lemma list [9]. For other rules, a blacklist (words that are excluded from the rule) was more practical. It contains 450 lemmas associated to 3 rules.

4.2 Short verbs

Short verbs can be defined as verbs that have a monosyllabic infinitive form. While (written) Standard German only has two short verbs (*sein* and *tun*), Swiss German dialects have a dozen of them. These verbs are characterized by short, irregular forms and rather obscure morpheme boundaries between stem and affix. As an additional difficulty, according to SDS data, there are Northwestern dialects in which short verbs are inflected like regular verbs.

`lexicalVariables:`

`rulename`	`lemma`	`tag`	`feats`	`datafolder`
gehen-pl	gehen	VVFIN	Pl,Pres,Ind	3-058-gehen-pl

`lexicalVariants:`

`rulename`	`form`	`mapname`	`addFeats`	`phonRulesAfter`
gehen-pl	gö	dp_gö	KurzV	kurzV-pl-vokal ö-e
gehen-pl	gang	dp_gang	RegV	
gehen-pl	göng	dp_göng	RegV	

Table 4. Extract of the lexical substitution tables, showing the relevant entries for generating plural forms for the short verb *gehen*.

In section 3.3, we have presented the tables for lexical substitution. For cases of suppletive morphology, we also use these tables. The field `feats` allows us to restrict the rules to certain inflected forms. The example presented in Table 4 shows the different plural forms of the verb *gehen* 'to go'. The first variant generates a short stem *gö* and adds the feature `KurzV`, which will trigger an affixation rule common to all short verbs, yielding *göö, göi, gönd, gön* etc. The second and third variants generate long stems and add the feature `RegV`, which will trigger regular verb inflection rules, yielding *gange* or *gönge*.

5 Experiments

In this section, we report the results of a simple experiment intended to determine the coverage of the transformation rules.

The most straightforward evaluation method would be to start with a list of annotated Standard German inflected words and to evaluate the Swiss German output generated for several dialects. However, it is hard to obtain reliable acceptability judgements from dialect speakers accustomed to highly variable spellings and pronunciations. Instead, we measured how many words of an existing multi-dialectal Swiss German corpus are analyzed correctly by our system.

The multi-dialect corpus consists of 100 sentences in five dialects, extracted from the Swiss German Wikipedia: Basel (BA), Bern (BE), Eastern Switzerland (OS), Wallis (WS), and Zürich (ZH). The dialect classification was done directly by the Wikipedia writers. These texts were then translated back to Standard German.

As our system is not conceived as an analyzer, we simulate this capability: starting with a Standard German word list, we generate a full form dialect lexicon with the help of our generator. Analyzing a word from the corpus then amounts to looking it up in the full form dialect lexicon. The (morphosyntactically annotated) Standard German word list has been extracted from the leaf nodes of the TIGER treebank [1]. With this approach, we can only recognize dialect words whose Standard German counterparts occur in the TIGER lexicon. As a result, many compound nouns and proper nouns are not recognized, even if the transformation rules would permit it. Because of this restriction, the maximum accuracy of our system lies at about 70% of word types and about 80% of word tokens.

We first obtained coverage figures without geographical filtering. In this scenario, when a Basel dialect word is analyzed, the system may also return derivations that are only valid in the region of Bern. The results are presented in the first row of Table 5. Except for the notoriously difficult Wallis dialect, the figures are fairly consistent across dialects: about 40% of word types and about 60% of word tokens are analyzed correctly.

The second scenario involved geographical filtering, retaining only analyses that obtained a minimal probability[9] of 10% in the most representative city of the respective dialect area.[10] Results are given in the second row of Table 5. With respect to the first scenario, there is only a slight performance drop (about 5% for types as well as tokens) for Basel and Zürich dialects, the three other dialects show performance drops ranging from 8% to 16%. The latter regions are larger and show more internal dialect variation than the former. In addition, the Wikipedia authors of the three latter regions probably use a dialect that diverges from the reference city dialect chosen for our evaluation.

	Types					Tokens				
	BA	BE	OS	WS	ZH	BA	BE	OS	WS	ZH
Without geographical filtering	42%	40%	41%	25%	45%	62%	57%	60%	44%	65%
With geographical filtering	37%	29%	27%	17%	40%	57%	47%	44%	30%	58%

Table 5. Percentages of correctly analyzed dialect words.

Several types of errors are encountered. First, some errors are due to different spelling choices. Indeed, the lack of binding spelling rules for Swiss German dialects makes the task difficult. For example, our system generated *bestaat*

9 Recall that each rule comes with a probability map. When several rules are applied to a word, the respective probability maps are combined by pointwise multiplication.

10 The city of Basel for BA, the city of Bern for BE, St. Gallen for OS, Brig for WS, and Zürich for ZH.

'consists' for Zürich dialect, while the Wikipedia corpus contained *bestaht*. Both variants are pronounced identically; the former conforms to the Dieth spelling guidelines [3], while the latter is closer to Standard German spelling rules. We found that the Wikipedia authors prefer a spelling closer to Standard German especially for long, complex words.

Other errors are due to missing rules. For instance, Standard German *Kirche* 'church' is phonetically transformed to *Chirche*, while a lexical transformation should be used to obtain *Chile* in Zürich dialect. Likewise, some specific inflectional affixes for Wallis dialect have not been implemented, which partially explains the lower scores.

Another type of error is due to diachronic change. Standard German *zeigt* 'shows' yields *zäägt, zaagt, zeigt* in Eastern Swiss German. While all of these forms have been used widely in that region in the 1950s (at the time of the SDS inquiries), they have become marginal today. The most frequently used version today is *zaigt*, which is indeed what we find in the Wikipedia texts.

6 Conclusion

We have presented a cross-lingual, multi-dialectal approach to word generation for Swiss German dialects.

Cross-lingual word generation only makes sense if a large amount of lexical pairs are cognates, and if the inventory of morphological and lexical features is fairly parallel across both language varieties. Because of the close etymological connection between Modern Standard German and Swiss German, these conditions are met. Cross-lingual generation allows us to rely on existing resources for the source language, which are much more numerous than for the target dialects. Multi-dialectal coverage is achieved by using existing dialectological resources in the form of probability maps. To our knowledge, this is a novel line of research.

Given these particularities and the largely manual creation of the rule base, we obtain honorable coverage figures of about 50% of tokens on several dialects.

The proposed set of transformation rules could be used as a part of a machine translation system between Standard German and the Swiss German dialects. Other potential applications include morphosyntactic analysis of dialect texts in order to enhance information retrieval, and integration in speech recognition and synthesis systems. The latter point is especially interesting given the mainly spoken usage of Swiss German dialects.

In future work, we plan to improve the rules on the basis of a detailed error analysis. Furthermore, a reimplementation with a finite-state toolkit would provide numerous benefits, such as higher speed and bidirectionality.

References

1. Brants, S., Dipper, S., Hansen, S., Lezius, W., Smith, G.: The TIGER Treebank. In: Proceedings of the Workshop on Treebanks and Linguistic Theories. Sozopol (2002)

2. Chambers, J.K., Trudgill, P.: Dialectology. Cambridge University Press, Cambridge, 2 edn. (1998)
3. Dieth, E.: Schwyzertütschi Dialäktschrift. Sauerländer, Aarau, 2 edn. (1986)
4. Fischer, L.: Luzerndeutsche Grammatik. Ein Wegweiser zur guten Mundart. Comenius, Hitzkirch, 2 edn. (1989)
5. Gallmann, H.: Zürichdeutsches Wörterbuch. Verlag Neue Zürcher Zeitung, Zürich, 2 edn. (2009)
6. Häcki Buhofer, A., Gasser, M., Hofer, L.: Das Neue Baseldeutsche Wörterbuch. Christoph Merian Verlag, Basel (2010)
7. Hammarström, H., Borin, L.: Unsupervised learning of morphology. Computational Linguistics 37(2) (2011), to appear
8. Hotzenköcherle, R., Schläpfer, R., Trüb, R., Zinsli, P. (eds.): Sprachatlas der deutschen Schweiz. Francke, Bern (1962-1997)
9. Institut für Deutsche Sprache, Programmbereich Korpuslinguistik: Korpusbasierte Wortformenliste (bzw. Grundformenliste) DEREWO, v-30000g-2007-12-31-0.1, mit Benutzerdokumentation. Mannheim (2007), `http://www.ids-mannheim.de/kl/derewo`
10. Lötscher, A.: Schweizerdeutsch. Geschichte, Dialekte, Gebrauch. Huber, Frauenfeld (1983)
11. Marti, W.: Berndeutsch-Grammatik für die heutige Mundart zwischen Thun und Jura. Francke, Bern, Switzerland (1985)
12. Rash, F.: The German Language in Switzerland. Multilingualism, Diglossia, and Variation. Peter Lang, Bern (1998)
13. Rumpf, J., Pickl, S., Elspaß, S., König, W., Schmidt, V.: Structural analysis of dialect maps using methods from spatial statistics. Zeitschrift für Dialektologie und Linguistik 76(3) (2009)
14. Scherrer, Y., Rambow, O.: Natural language processing for the Swiss German dialect area. In: Proceedings of KONVENS'10. Saarbrücken (2010)
15. Thielen, C., Schiller, A., Teufel, S., Stöckert, C.: Guidelines für das Tagging deutscher Textkorpora mit STTS. Tech. rep., University of Stuttgart and University of Tübingen (1999)