

Ueberwasser, Simone (2013): Non-standard data in Swiss text messages with a special focus on dialectal forms. In: Zampieri, Marcos; Diwersy, Sascha (Hrsg.): *Non-standard Data Sources in Corpus-based Research*. (=ZSM-Studien, Schriften des Zentrums Sprachenvielfalt und Mehrsprachigkeit der Universität zu Köln 5. Hrsg: Bongartz, Christiane M.; Riehl, Claudia M.). Aachen: Shaker Verlag, 7-24.

# Non-standard data in Swiss text messages with a special focus on dialectal forms

Simone Ueberwasser  
Deutsches Seminar  
University of Zurich

## Abstract

The investigation in this paper is based on the Swiss SMS corpus. In a first step, the use of the two varieties *Swiss German Dialects* and *Standard Swiss German* are investigated as to who uses which variety and why informants decide to switch varieties. In a second part, the focus is on the spelling of the Swiss German dialects, a group of varieties without spelling norms. This lack of norms leads to an abundance of different spelling forms for the same signifié and accordingly to restrictions when working with the corpus. An additional layer of equivalent terms in Standard German will be created in the corpus to assist researchers.

## 1 Introduction

“A Glossary of Netspeak and Textspeak”: This title of a book by David Crystal [7] implies that there is such a thing as a special language used on the internet and in text messages<sup>1</sup>. In a more recent book by the telling title “txtng: the gr8 db8” [8] about SMS only, the impression that the

---

<sup>1</sup>The terms text message and SMS will be used synonymously in this text.

language used in text messages is a world apart with its own norms and regulations is even enforced. But does this point of view hold true? Isn't it more likely that existing habits of language use are being transferred to this type of text production? One example for such a point of view is certainly the use of dialects<sup>2</sup> in Swiss German SMS, which, as will be shown on the basis of real data, reflects the ordinary everyday language use. As part of an initiative, which started in Belgium<sup>3</sup>, a Swiss SMS corpus<sup>4</sup> featuring nearly 26,000 authentic SMS in different languages and varieties was put together. The 17,962 German SMS (any variety) in the corpus were written by 977 persons, 610 of which are female. Of these SMS, 10,716 are in fact written in SGD, thus offering a unique insight into the use of SGD and into sociolinguistic aspects of its use. On the downside, the extremely nonstandard character of these dialectal SMS poses a major problem when working with the corpus, both, to the research and to an intended automated data processing.

In this paper, I will thus first investigate why informants in the Swiss SMS corpus use one or the other variety. This question will be answered based on demographic data about the informants and on their own statements about their language use. While the demographic data such as age, sex or education was collected systematically, comments about language use were written as edited statements by the informants when answering the question "how do you write text messages?". Accordingly, they can only be evaluated qualitatively. In a second part, spelling strategies for SGD, a variety without codified spelling norms, will be laid out and a short insight into the Swiss SMS team's approach to standardizing the corpus will be given<sup>5</sup>.

---

<sup>2</sup>Following Siebenhaar ([18]:487) the term *dialect* or the short form *SGD* will be used as an umbrella term for all Swiss German dialects hereafter, in spite of their linguistic differences.

<sup>3</sup>C.f. [www.sms4science.org](http://www.sms4science.org).

<sup>4</sup>For more information about the Swiss SMS corpus, please refer to [www.sms4science.ch](http://www.sms4science.ch) as well as [12].

<sup>5</sup>For more information on this process c.f. Ruef/Ueberwasser in this volume.

## 2 Diglossia in Switzerland

Looking at any text book or linguistic dictionary (e.g. [6]) that covers diglossia, the linguistic situation in the German speaking part of Switzerland is most likely mentioned, because it is one of Ferguson's four examples when he first describes his concept. Ferguson defines diglossia as follows:

- (1) DIGLOSSIA is a relatively stable language situation in which, in addition to the primary dialects of the language [...], there is a very divergent, highly codified [...] superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation ([14]:34-35).

The emphasis in this definition lies on the dichotomy between a codified variety used in formal situations and a dialectal one used for ordinary conversations. Ferguson thus does not imply that one variety is used in spoken language production and the other one for written texts. In fact the samples he gives to exemplify the situation include communication forms that use the standard variety in spoken language production (e.g. in sermons and speeches in parliament) but also some that use the dialectal variety in writing (e.g. in captions on political cartoons) ([14]:28). It was only Kolde [15], who introduced the concept of a medial diglossia for Switzerland, i.e. the idea that the German speaking Swiss in most cases use their dialects when speaking and the Standard German variety (StG)<sup>6</sup> when writing. At the time when Kolde suggested this idea, it was certainly valid from the perspective of written language, as Siebenhaar confirms: "Swiss dialects

---

<sup>6</sup>The Standard German varieties used in Switzerland, Austria and Germany are different on the phonetic, lexical and grammatical level (c.f. e.g. [1]). Since the following analyses relies on text messages written by native speakers of SGD, it can be assumed that all texts written in the standard variety use the Swiss German Standard. Accordingly, no differentiation will be made.

rarely appeared in written form before the 1980s, apart from the genre of dialect literature ([18]: 481)". However, as can be seen from Ferguson's examples cited above, StG was already used in spoken text types in the 1950s, and Kolde's concept came under an increasing pressure ever since. Starting in the 1980s, "dialect has gradually been finding its way into channels of personal written communication such as personal letters, notes and e-mails ([18]: 483)". Facing the number of dialectal SMS in the Swiss SMS corpus, this is certainly true for texting, too. However, since there are also native speakers of SGD who write SMS in StG, as will be shown, the question remains whether the use of one or the other variety is triggered by pure coincidence or whether there are any specific factors that influence the choice.

## **2.1 Prestige of the dialect**

One factor that can be excluded as triggering the use of StG is prestige, even though Ferguson states that "[i]n all the defining languages the speakers regard [the standard] as superior to [the dialect] [...]([14]:29)". In spite of Ferguson making this statement for "all the defining languages", he only gives examples from areas other than Switzerland. More importantly, more recent studies, which focus more specifically on the situation in Switzerland, agree on contradicting Ferguson on this point. Rash e.g. writes: "[A]s a diglossic community, [German speaking Switzerland] is unique among [German speaking] countries in that all germanophone Swiss speak a regional dialect as their first language, and no social, economic or educational factors cause the [Swiss German] dialects to be accorded less prestige than the standard language ([17]:50-51)". Consequently, she calls Ferguson's assumption about the low prestige of the dialect "[...] an assessment which does not fit the situation in Switzerland at all well ([17]:49)". Ender ([13]:274), quoting several Swiss researchers, goes one step further in confirming that the dialect indeed has more prestige than the standard language, a point of view supported by many speakers of SGD. The following SMS can be seen as an example for the prestige that the dialect has in the eyes of the author:

- |  |   |
|--|---|
| <p>(2) <i>Guete obig herr [LastName] Ich weiss nöd ob sie das wüisset aber ich (und die vom handball) chönd de sam. nöd cho ...mfg</i></p> | <p>‘Good evening Mr [Last-Name] I don’t know whether you [v-form] know this, but I (and the ones playing handball) cannot come this Saturday [abbreviated] ... with kind regards [abbreviated]’</p> |
|--|---|

(2) is the SMS of a 15-year-old who writes in a voice of deference, as can be seen by the use of the v-form for *you* instead of the more intimate t-form. Additionally, he uses a formal term of address (*Herr*, ‘Mr’) as well as a greeting formula normally seen in business: *MfG* for *Mit freundlichen Grüssen* (‘with kind regards’) rather than the formula that is much more common in SMS but also more intimate *LG* for *Liebe Grüße* (‘love’). Yet, in spite of all these tokens of deference, the SMS is still written in the dialect. Obviously the writer does not consider the use of dialect to be lacking deference or prestige.

Another indicator for the prestige held by the dialect is of course the sheer amount of SMS written in the dialect. Of the 11,758 SMS written by people who registered themselves as native speakers of SGD, 7,391 were written in the dialect, which equals to more than 60%. If the dialect didn’t hold prestige, certainly people would be more reluctant to use it.

## 2.2 Dialect as the variety of ordinary conversation

Rather than the variety of low prestige, SGD is the “[...] medium of ordinary conversation ([14]:35).” It is acquired as a mother tongue (c.f. [17]: 49), while StG is only learned at school later on and has a taste of formality, pressure to perform and distance. The dialect, on the other hand, goes along with a cozy feeling of identity and belonging (c.f. [3])<sup>7</sup>. This

<sup>7</sup>This attitude of the German speaking Swiss towards their dialect leads some linguists, who take an ethnolinguistic point of view, to speak of a situation of bilingualism rather than diglossia, c.f. [19].

point of view is also supported by informants of the Swiss SMS corpus. One of them, e.g., answered the question as to how SMS are written with:

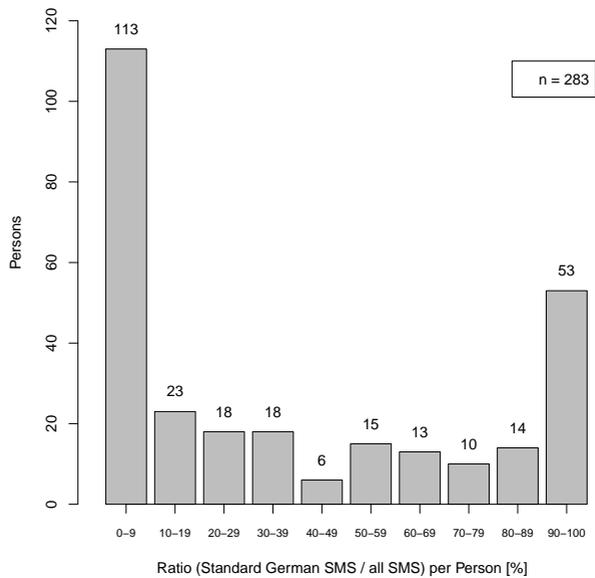
- (3) [...] *natürlich auf Mundart, was in meinen Augen viel persönlicher ist und ja auch meine Muttersprache ist.*      ‘[...] of course in the dialect, which is much more intimate in my eyes and after all it is my mother tongue.’

The informant who made statement (3) shows both, the feeling of proximity when using the dialect but also the distance to StG, because it is not one’s mother tongue. If this feeling of proximity towards SGD and of distance towards StG holds true for most German speaking Swiss, we can expect a high percentage of personal communication, both, in spoken and in written, to take place in SGD, because it is the level of intimacy of the communication that triggers the selection of a specific variety rather than the medium used (i.e. spoken vs. written language). Indeed we see a first indication of this situation when looking at how few people do actually use StG in their SMS.

Figure 1 summarizes the SMS per person and shows for each person the percentage of SMS written in StG. Because people who sent in only one or two SMS would distort the statistics, only people who sent in 10 SMS or more are considered. The chart shows two peaks for those people who write nearly all SMS in dialect (i.e. who write less than ten percent of their SMS in StG) and for those who write nearly all SMS in StG. The percentage of people who switch regularly can be considered as statistic noise. Looking at extreme values, this point of view is confirmed: 56 people write only in dialect and 30 use only the standard language. Coming back to what was said above, we can formulate as follows: Those people who registered SGD as their mother tongue choose either this very variety when writing SMS or else they fall back to the variety which they were taught to consider the norm for written communication at school, i.e. StG, while habitual code switching is rare<sup>8</sup>. However, the comparison between

---

<sup>8</sup>Even though code switching within individual SMS is very frequent but beyond the focus of this study.



**Figure 1:** *Percentage of SMS written in StG for people who wrote ten SMS or more.*

those who prefer the dialect and those who prefer StG shows a clear preference for the dialect.

### 2.3 Why a variety is chosen

The shown preference for consistency in language selection on the one hand and for the dialect on the other hand opens the floor for two different types of questions: a) Which demographic or linguistic factors trigger people to choose one or the other variety? b) If somebody prefers one variety for nearly all his communication via SMS, what makes him or her choose

the other variety? Attempting to find answers to the latter question, we will first look at an SMS that was written by a habitual writer of Standard German, who uses the dialect only for the SMS shown in (4).

- (4) *Das chaibe Maitli lot mis Händy zu jedere Tages- und Nachtzyt im Hosesack klingle. Aber I ha de Plausch dra. Bonne nuit mon étoile. Lu* ‘This blooming [Double entente intended] girly makes my cell phone ring all day long in my pocket. But I take delight in it. Good night my starlet [in French]. [Signature]’

The SMS in 4 is written by a 76-year-old man who sent in four more SMS which were all written in StG. Whereas it is impossible for most SMS to make a statement about the addressee, the language used in this one gives a hint about who it was written to. The formulation *Das chaibe Maitli* would not be used to address a teenager or an adult, but only to children. Seemingly, the child already knows some French, but foreign language teaching starts at a rather young age in Switzerland, so that is no contradiction. Whether the author of the SMS is the grandfather or a friend of the family cannot be said, but he certainly adjusts his language use to the addressee. By using the dialect, he simulates a spoken conversation and thereby expresses closeness and familiarity (c.f.[4]:104).

At the other end of the scale, (5) is the only SMS in StG by a 19-year-old woman who otherwise sent in 91 SMS in SGD.

- (5) *Hoi elsa,danke fürs abmelden.Wünsche jannine demfall gute besserung.Lg leonore* ‘Hi elsa, thanks for canceling. In that case I wish Jannine get well. Love Leonore’

In (5), the formality of the situation seems to be the trigger for the change in variety. Somebody canceling an appointment for somebody else

would normally take place with teachers, coaches etc. and here, this person of respect is answering in the otherwise avoided StG, even though the language use is rather colloquial in all other aspects.

The two SMS shown under (4) and (5) can be seen as examples for the two most influential factors for the choice of variety (c.f. e.g. [18]: 499, [4]:104): the addressee and the formality of the situation. This is also confirmed by informants, who often give intimacy and belonging as reasons for writing in SGD, while StG, according to many informants, is used for addressees who are not likely to understand SGD, such as foreigners, and for situations or addressees which seem more formal, such as persons not well known, business situations and customers. Some young people name their parents as people to whom they write in StG. Whether this is out of respect or because of the difference of age, cannot be said.

Situational code switching and adjustment to the addressee has been found before (e.g. [18], [13]). However, in text messages, other factors seem to play a role, too, e.g. the mobile phone's option to recognize and correct text. Many writers of SGD in fact report to having this feature disabled<sup>9</sup>, while others 'taught' their mobile phones SGD, i.e. appended their phone's vocabulary with SGD expressions. A rather amazing factor that is mentioned regularly as triggering the use of SGD is shortness. Because of the many enclitic forms in SGD, this variety seemingly allows for shorter SMS, a fact that the informants point out as being important to them because they do not want their text messages to be split at 160 characters. In spite of shortness, informants state that they want to be understood. For some, this is a reason to switch to StG, while others state that they adjust their writing strategy for the dialect to be less regional and thus easier to understand. Another important factor to be mentioned is comfort. There are many informants who write in SGD because this is the way they think and talk, thus, they seem to feel that they would have to translate their thought when using StG. However, comfort is also a reason mentioned for using StG, because familiarized spelling conventions are easier to apply than graphical representations of an otherwise spoken variety, which have

---

<sup>9</sup>Of those informants who did send in SMS in SGD and did answer the question about the use of T9, 112 answered with yes while 672 do not use this spell checker.

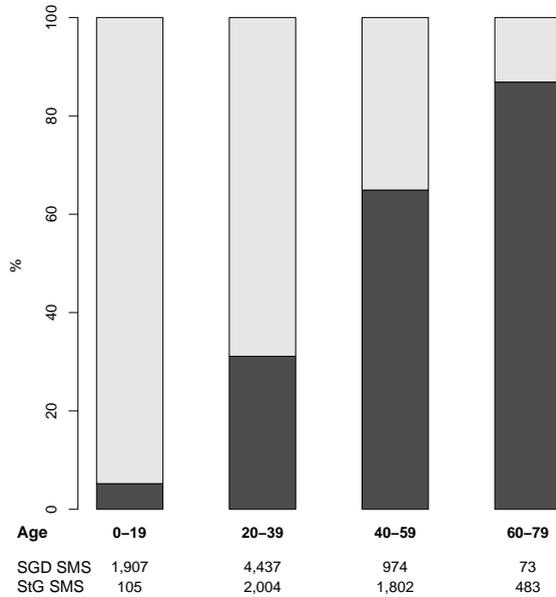
to be invented on the spot. Coming back to selection of variety based on the addressee, a last factor to be mentioned for switching to StG is the age of the addressee. Apparently, according to the informants, older people are more likely to be addressed in StG than younger ones but also more likely to use StG, as will be shown in the next chapter.

## 2.4 Who chooses which variety

Beat Siebenhaar, like the informants of the Swiss SMS corpus, expects a correlation between age and use of the dialect ([18]:492). However, in his chat data he finds a U-shaped curve for the dialect ratio in respect to age. The situation in the Swiss SMS corpus is much closer to the expectation, as can be seen in figure 2, which shows the ratio between SMS in SGD and those in StG summarized by age groups. Only SMS written by native speakers of SGD are considered.

Even though the data represented in figure 2 is highly significant as a whole ( $\chi^2 = 334.7$ ,  $df = 9$ ,  $p < 0.001$ ), the last column for informants who are 60 and older has to be taken with a pinch of salt because there are only 50 people in this group (as compared to, e.g., 269 in the youngest group), but still the figures are telling. Different approaches can be taken when trying to interpret these figures. Older people were certainly socialized before the ‘dialectal-turn’ in the 1980s as presented above. Younger people, on the other hand, are likely to be more active authors of texts in the new media and thus better accustomed to the dialect. A third factor might be the professional or educational importance of writing in StG. A 15-year-old boy, e.g., writes in his comment that he used to write SMS in StG but has recently switched over to the dialect. It seems that the teaching of basic education, which takes place in StG, was too dominant in his young life to allow for a modified language use. As he grew into puberty, other factors became more important. For older people, on the other hand, the need to perform in the world of business, and accordingly the need to perform in StG, might have become a routine that is kept up when writing SMS.

While age plays an eminent role in the use of varieties for SMS, another factor that is relevant in Siebenhaar’s data does not. He sees an



**Figure 2:** *Percentage and absolute figures for SMS written in SGD and StG respectively by mother tongue speakers of SGD.*

increased use of the dialect in a chat named #bern and explains it with people from Bern being more affiliated with their dialect due to cultural circumstances ([18]:492). While I fully agree with his argumentation, the same distribution cannot be found in the Swiss SMS corpus. Here, people who live in the canton of Bern are actually slightly less likely than the average to use the dialect. This difference in findings can be explained by the different approach to labeling people as coming from Bern. A chat named #bern attracts people who care about Bern no matter where they live but probably they speak the according dialect, while in the SMS corpus, people are labeled as coming from Bern if they registered an according postal

code. These people, however, might only work and live in the capital but originate from other regions. Thus, the difference in findings between this study and Beat Siebenhaaar's is most likely due to the difference between people living in a certain area and people originating there. Furthermore, participating in a chat group creates a strong in-group feeling and if the name of the group corresponds to a geographical entity, participants are likely to use the corresponding dialect more extensively than they would do in another communication form, if only to enforce the feeling of belonging.

A last factor to be considered is the informants' sex. Of those people who have SGD as their mother tongue,  $\frac{2}{3}$  are female and they are responsible for  $\frac{2}{3}$  of the SMS. Like their male counterparts they only write approximately  $\frac{1}{3}$  of their SMS in StG. However, when looking at people's sex in proportion to their age, the figures become more telling. In the two age groups who prefer SGD in figure 2 (i.e. people younger than 40), women are stronger users of SGD than men, while in the other two groups, women are more likely than men to use StG. Thus, women seem have a forerunner role in the use of either variety according to their age group.

So far the focus of this study has been on differences in the use of SGD and StG. While the spelling norms for the latter are clearly defined and mostly followed in the Swiss SMS corpus, this is not the case for SGD. The next chapter will thus focus on the spelling of SGD and the impact of the lack of norms on corpus based research.

## **3 The lack of a writing norm**

### **3.1 Spelling the Swiss German dialect**

SGD diverges from StG on all linguistic levels, i.e. phonetically, morphologically, syntactically, lexically and in pragmatics. Furthermore, it does not have a spelling norm. While all other factors are at best a challenge when working with a corpus, the lack of a spelling norm is a hindrance. The following representation forms of the StG word *viel* ('much') were only found by intuition and not systematically:

(6) *viel, vil, viiu, viil, vel, veu, fiu, fil, vieu*<sup>10</sup>

In spite of the many different variants shown in (6), there is still some consistency in the graphical representation of SGD, as Müller ([16]) shows for the representation of consonants, which is often close to the StG equivalent in her data. Dürscheid and Stark ([11]) on their side find consistency in the representation of lexical morphemes. Additionally, individual informants are often consistent within themselves. In some cases informants systematically use one specific letter or group thereof to stand for a sound that is represented by a different graph in StG and they comment on this behavior in the corpus' meta-data, e.g. “[...] und schreibe die sch zb. bei stern aus” (‘[...] and [I] use <sch> [to represent /ʃ/] as in *Stern*’). This rule is then consistently followed by this informant, resulting in tokens such as *schpontan* for *spontan* (‘spontaneously’) or *dia geilschta* for *die Geilsten* (‘the greatest’). However, spelling in the corpus as a whole is still very inconsistent for different reasons. Firstly, as Müller ([16]:175) points out, the graphs used to present vowels diverge much more from the StG forms because the phonetic vowel system of the two varieties is different. Secondly, SGD is an umbrella term that covers many dialects which are different among themselves. These differences are normally also expressed in the written representation. An example can be seen in (6), where the last letter represents either the sound /l/ or the vocalized form /u/, thus demonstrating a regional distribution. The third reason for the inconsistency is the informants’ urge to be creative, as can be seen in forms such as *koohl* or *kuuhl*, which both stand for *cool*.

This inconsistency is cumbersome when working with the corpus in two ways. On the one hand, every lexical item has to be looked up by means of a regular expression to cover for as many imaginable spelling variants as possible. This, of course becomes even more complex than in the examples shown under (6), when searching for inflected forms. On the other hand, it restrains any kind of automated data processing, such as part-of-speech tagging. It is therefore necessary to add a parallel second layer with a standardized spelling to the corpus. When it comes to searching the

---

<sup>10</sup>No inflected forms considered

corpus, this layer will be treated in much the same way as in a multilingual parallel corpus.

### 3.2 Standardizing the Swiss German dialects

In theory, two different approaches seem possible when standardizing the SGD data in a corpus: a) standardize towards StG or b) standardize within the SGD. The latter is an impracticable task as it would mean to set the dialect of one region as the standard (as Ferguson ([14]:27) did), or to define a new standard, an attempt that has failed before (c.f. [16]:157). Any attempt to create a written norm for SGD would be seen as a political inter-ception with their national (or, more precisely: with their inter-regional) independence by the Swiss (c.f. [5]:122), or, as Rash ([17]: 80) puts it: “Any debate over the introduction of an artificially created written Swiss German is, in a sense, hypothetical, as most observers agree that it will never exist.” Thus, the only possible way to go is to standardize towards StG, and that is what the Swiss SMS team is doing. They use the labour of payed students to attribute StG equivalents to every SGD token following certain rules.

The first rule is not to mingle with syntax. Accordingly, when standardizing the example a) in (7) they will work towards b) rather than towards c) and thus leave the ellipses empty (which has to be filled in proper StG as shown under c) but is normally left empty in SGD as shown in a), unless special emphasis is put on the here omitted personal pronoun), and they will also not invert the two pronouns *ihm* and *es* as should be done in StG (c.f. c)). This allows researchers to find syntactic structures typical for SGD in spite of the standardization.

- (7) a) *Häsch* Ø *em s gsait?* ‘Did you tell it to him?’  
b) *Hast* Ø *ihm es gesagt?*  
c) *Hast du es ihm gesagt?*

All other rules can be summarized under four credos: a) The standardized token’s spelling should be as close as possible to the dialectal token as it is found in the corpus. b) No lemmata should be invented. c) The semantic

content should not be changed. d) Consistency stands above any other rule.

As has to be expected, these credos are not comprehensive. When standardizing some tokens, the credos do not offer a applicable solution, for other tokens, the credos contradict each other. Accordingly, more detailed rules were defined for the students to follow and individual solutions were documented in databases. The most important ones are:

- (8) a. Clitical tokens are split into their components. In (7 a), e.g., the first three tokens are more likely to be spelled as *hüschems* rather than in the way presented here. In the standardized level, however, the separated form as shown under (7 b) would be use<sup>11</sup>.
- b. Every lemma found in a codex<sup>12</sup> can be used as a standardized form, even if the semantics diverge between StG and SGD (e.g. *schaffen* means *to work* in SGD but *to create* in StG, but will still be used to represent *to work*). In some cases the “Idiotikon”<sup>13</sup> is used to find historic lemmata linked to SGD tokens that are still used in StG albeit with a different meaning.
- c. For tokens that are syntactically relevant in SGD but do not have equivalents in StG, the most frequently used variant in the corpus is used (e.g. *go*, *ga*, *goga* etc. are standardized as *go*). Since there is not even a handful of this type of recurring particles in SGD, this interference seems acceptable.
- d. Case is adjusted to StG spelling norms, i.e. nouns are spelled in upper-case.
- e. Prepositions are not adjusted to StG. In the phrase *I drive to Zurich* the preposition is realized as *nach* in StG and as *uf* (StG: *auf*) in SGD. It will consequently be standardized as *auf*, thus

---

<sup>11</sup>Since the original layer and the standardized one will be aligned in the corpus browser, clitical forms will be discoverable by searching for multi-word tokens in the standardized layer.

<sup>12</sup>Mainly [9], [10] and [2].

<sup>13</sup>A major codex that collects SGD expressions from all regions, c.f. <http://www.idiotikon.ch/>

keeping the SGD lemma but spelling it in the StG form. The same goes, of course, for all word forms and not only for prepositions.

- f. If a token cannot be reliably interpreted, it is marked as *unclear*. E.g. the form *kuzih*, that is found at the end of some SMS can be a long form of *Kuss* ('kiss') or a short form of the first name *Markus*. Accordingly, it is marked as *unclear*.
- g. Individual tokens from languages other than German are marked.

Only a short list of the rules that were given to the students who standardize the corpus is shown under (8). More rules exist on a systematic level but also on the level of individual tokens. Furthermore, in order to ensure consistency, the students work with a proprietary server-based tool that looks up tokens already standardized by any of the students and presents them as suggestions<sup>14</sup>. This should—hopefully—not lead to a biased interpretation by the students but instead support consistency.

Once this standardized level is ready, it will be added to the corpus as a parallel level, thus allowing for more spelling variants to be found than were presented here. More importantly, however, this level will be used for further processing of the data. It will serve as the basis for part-of-speech tagging and possibly other annotations, thus transforming the Swiss SMS corpus into a little jewel for research on colloquial everyday language use in Switzerland.

## 4 Conclusions

At least in the German speaking part of Switzerland, the language use in SMS is nothing new and nothing apart. Instead, the informants of the Swiss SMS corpus thoughtfully adapt their language use to the situation and to the addressee. The foremost factor for the selection of a variety is the addressee, because the author of an SMS wants to be understood and to be friendly by adjusting to the addressee's preferences. For younger

---

<sup>14</sup>For more information c.f. Ruef/Ueberwasser in this volume.

people, the use of SGD is the norm from which they only divert depending on the addressee or the formality of the situation. Older people, on the other hand, stick to the rules they learned at school and consequently use StG for all written texts, also for SMS, while only switching to SGD for special addressees (such as small children) and special situations (such as situations of familiarity). In general, the use of SGD is seen as more intimate but also as easier, because the use of StG is often felt as a task of translation, especially by younger informants. People, who are more prone to use StG, on the other hand consider the need to invent spelling options as a burden that is more demanding than the 'translation' to StG. In general, the dialect is seen as the medium of ordinary conversation and because SMS is mostly a communication form used for ordinary conversations, SGD is the logical choice when writing SMS.

While the use of one or the other variety is nearly a dichotomy, the ways of spelling SGD show a broad variability. Informants more often rely on StG conventions but also feel the need to manifest phonic shibboleths of their own dialect in the graphic form, especially when it comes to vowels. An example is the representation of the /ʏ/ in *Glück* as <ö>, because it is pronounced as /ø/ in the canton of Aarau (c.f. [16]:171). Accordingly, graphic representations of one specific phonic value can diverge to a great extent, thus posing major problems when working with the corpus. The Swiss SMS team has thus decided to invest time and money into a complete manual standardization of their corpus towards an additional layer in StG. This layer should mirror the original SMS layer as well as possible, a task that is achieved by using equivalents which are as close as possible to the dialectal token and by not changing the original syntax.

## References

- [1] Ulrich Ammon. *Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten*. de Gruyter, Berlin, New York, 1995.

- [2] Ulrich Ammon, Hans Bickel, Jakob Ebner, Ruth Esterhammer, and Markus Gasser. *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol*. de Gruyter, Berlin, New York, 2004.
- [3] Brigitte Aschwanden. »Wär wot chätä?« Zum Sprachverhalten deutschschweizerischer Chatter. *Networx*, (24), 2001.
- [4] Bettina Braun. Jugendliche Identitäten in SMS-Texten. In Christa Dürscheid and Jürgen Spitzmüller, editors, *Zwischentöne. Zur Sprache der Jugend in der Deutschschweiz*, pages 101–114. Verlag NZZ, Zürich, 2006.
- [5] Michael Clyne and Joshua A. Fishman. German as a pluricentric language. In Michael Clyne, editor, *Pluricentric Languages, Contributions to the Sociology of Language*, pages 117–147. de Gruyter, Berlin, New York, 1992.
- [6] David Crystal. *A Dictionary of Linguistics & Phonetics*. Blackwell, Malden etc., 2003.
- [7] David Crystal. *A Glossary of Netspeak and Textspeak*. Edinburgh University Press, Edinburgh, 2004.
- [8] David Crystal. *txtng: the gr8 db8*. Oxford University Press, Oxford, 2008.
- [9] Duden. *Die deutsche Rechtschreibung*. Bibliographisches Institut & F. A. Brockhaus, Mannheim u.a., 25., völlig neu bearbeitete und erweiterte Auflage, 2009.
- [10] Duden. *Schweizerhochdeutsch. Wörterbuch der Standardsprache in der deutschen Schweiz*. Duden Verlag, Mannheim, 2012.
- [11] Christa Dürscheid and Elisabeth Stark. Anything goes? SMS, phonographisches Schreiben und Morphemkonstanz. In *Die Schnittstelle von Morphologie und geschriebener Sprache. Tagungsband DGfS 2011*, in prep.

- [12] Christa Dürscheid and Elisabeth Stark. sms4science: An International Corpus-Based Texting Project and the Specific Challenges for Multilingual Switzerland. In Crispin Thurlow and Kristine Mroczek, editors, *Digital Discourse: Language in the New Media.*, pages 299–320. Oxford University Press, New York, London, 2011.
- [13] Andrea Ender and Irma Kaiser. Zum Stellenwert von Dialekt und Standard im österreichischen und Schweizer Alltag – Ergebnisse einer Umfrage. *Zeitschrift für germanistische Linguistik*, (37/2):266–295, 2009.
- [14] Charles A. Ferguson. Diglossia. In Thom Huebner, editor, *Sociolinguistic Perspectives. Papers on Language in Society, 1959 – 1994*, Oxford Studies in Sociolinguistics, pages 25–39. Oxford University Press, Oxford, 1959/1996.
- [15] Gottfried Kolde. *Sprachkontakte in gemischtsprachigen Städten. Vergleichende Untersuchungen über Voraussetzungen und Formen sprachlicher Interaktion verschiedensprachiger Jugendlicher in den Schweizer Städten Biel/Bienne und Fribourg/Freiburg i.Ue.* Wiesbaden. *Zeitschrift für Dialektologie und Linguistik*. Beiheft. Franz Steiner Verlag, Stuttgart, 1981.
- [16] Christina Margrit Müller. Dialektverschriftung im Spannungsfeld zwischen standardnah und lautnah. Eine korpuslinguistische Untersuchung der Rubrik Dein SMS in der Aargauer Zeitung. In Helen Christen, Franz Patocka, and Evelyn Ziegler, editors, *Struktur, Gebrauch und Wahrnehmung von Dialekt. Beiträge zum 3. Kongress der Internationalen Gesellschaft für Dialektologie des Deutschen (IGDD), Zürich, 7.-9. September 2009*, pages 154–178. Praesens, Wien, 2011.
- [17] Felicity Rash. *The German Language in Switzerland: Multilingualism, Diglossia and Variation*. Linguistic and Cultural Studies. Peter Lang, Bern, 1998.

- [18] Beat Siebenhaar. Code choice and code-switching in Swiss-German Internet Relay Chat rooms. *Journal of Sociolinguistics*, (10/4):481–506, 2006.
- [19] Iwar Werlen. Mediale Diglossie oder asymmetrische Zweisprachigkeit? Mundart und Hochsprache in der deutschen Schweiz. *Babylonia*, 1(1):22–35, 1998.