Quantitative Approaches to Linguistic Variation in IRC: Implications for Qualitative Research

Beat Siebenhaar
University of Leipzig, Germany

Abstract

Qualitative analysis of code choice, code switching, and language style in Internet Relay Chat (IRC) can shed light on functional-pragmatic aspects of the use of different linguistic varieties. However, in a qualitative analysis, the status of varieties within a channel or for a single chatter can only be guessed at. Moreover, qualitative research on linguistic variation in IRC often fails to generalize its findings due to a restricted database or a restricted view of a database. This article introduces an approach that allows for embedding of qualitative research within a quantitative research design. The quantitative method presented here enables general statements to be made about the use of varieties or the usage of certain chatters in a chat channel. The approach is exemplified with data from Swiss IRC channels, in which Swiss German dialects and standard German are used side by side. A large corpus is analyzed for static and dynamic aspects of dialect share. It is argued that this quantitative approach can provide a background for qualitative analysis and facilitate the selection process of relevant data required for qualitative analysis.

Introduction

In many Internet Relay Chat (IRC) channels, more than one language or linguistic variety is used. It is evident that code-switching occurs, as noted in many publications on IRC, including Androutsopoulos and Hinnenkamp (2001), Lam (2004), Paolillo (2001), Warschauer, El Said, and Zohry (2002), and McLellan (2005), who reviews Malayan literature on code-switching and IRC. With the exception of Paolillo (2001), these publications follow an example-based, qualitative approach. However, qualitative research often fails to generalize its findings due to a restricted database. This article introduces an approach that allows embedding of qualitative research within a quantitative research design, in order to enable generalization of the findings. The approach is exemplified with data from IRC channels in German-speaking Switzerland.

German-speaking Switzerland has been discussed as a classic example of diglossia by Ferguson (1959). However, the contemporary linguistic situation in Switzerland has changed since the time of Ferguson's publication. The use of the two varieties, namely standard German and Swiss German, coincides with the distinction between written and spoken language, but it does not correlate with the distinction between high and low prestige varieties, as in many cases of diglossia. In fact, Swiss-German dialects are considered to be highly prestigious in most domains by Swiss-German speakers, regardless of their socio-economic class. Thus the diglossic situation in Switzerland can be considered to be a case of *medial diglossia* (Kolde, 1981, p. 68), which entails several exceptions.[1] Standard German is the required spoken variety in school instruction, in certain TV programs, and in interactions with non-dialect speakers.

Standard German is also generally used for writing, while Swiss-German dialects are reserved for spoken communication. However, non-standardized dialect writing has established itself for personal communication purposes within the last 20 years. This shift to using dialect rather than standard German for written communication is reinforced in computer-mediated communication (CMC).

Far from being uniform, Swiss German dialects show a great amount of variation, which is documented in the eight volumes of the *Sprachatlas der deutschen Schweiz* (1962-1997) and numerous other publications (see the bibliographies in Sonderegger, 1962 and Börlin, 1987). On the basis of the *Sprachatlas der deutschen Schweiz* (1962-1997), Hotzenköcherle (1984) divides German-speaking Switzerland into nine linguistic areas,[2] which are subdivided into three to seven subregions each. However, people from all over German-speaking Switzerland who speak their vernacular can still be localized on a smaller scale (Christen, 1998), a finding that demonstrates the vitality and stability of local dialects. As is common practice in German linguistics, I use the term 'dialect' to refer both to a specific local variety and to the sum of local varieties in contrast to standard German. Referring to, e.g., a Bernese or Zurich dialect is therefore a simplification corresponding roughly to Hotzenköcherle's subregions.

The co-existence of standard German and Swiss-German dialects is reflected in IRC writing. However, their distribution in IRC corresponds more closely to oral than written practices. An initial glance at Swiss-German IRC suggests that the majority of communication is not conducted in standard German but in dialect. Dialect writing is not uniform, and norms for dialect orthography appear to be nonexistent. Such conventions for dialect standardization that do exist, e.g., in scholarly works, are not known to chatters.[3] As a result, each chatter tends to employ his/her own written dialect conventions; these are influenced by standard German orthographic principles and, at the same time, by tendencies to mark differences from standard German orthography.

For example, in Zurich German 'fir tree' is /'tanə/, while in Bern German it is /'tannə/ with a geminate. The standard German spelling is *Tanne*. The spoken dialect differences would suggest a Zurich spelling *Tane* and a Bernese spelling *Tanne*. However, standard German orthography, where double consonants indicate that the preceding vowel is short, is so influential that *Tanne* is also found in the IRC channel #zurich. Furthermore, dialectal schwa /ə/ is often represented by *ä*, a spelling that can be seen as a dialect marker (Christen, 2004; Siebenhaar, 2003). Both tendencies can lead to a Zurich spelling *Tannä* for the pronunciation /'tanə/, while *Tane* and *Tanne* also can be found. Moreover, contradictory tendencies to maintain morphological regularity and to reflect phonetic particularities are found. Overall, the use made of different variants leads to the impression that chatters' choices are arbitrary.

Previous research suggests that the informal situation of IRC favors an informal language style (e.g., Grondelaers, Geeraerts, Speelman, & Tummers, 2001 for Dutch; Hentschel, 1998 for Serbian; Runkehl, Schlobinski, & Siever, 1998 for German; Warschauer et al., 2002 for Arabic; Werry, 1996 for English). In northern German IRC channels, this informality is expressed through some non-standard forms. In most Swiss

IRC channels, it is expressed by choosing dialect rather than the standard variety (see Christen, Tophinke, & Ziegler, 2005). Moreover, since a personal trait of each chatter is his or her recognizable dialect, which allows quite an exact localization, non-standardized dialect writing allows for linguistic individualization. In this context, standard writing also becomes a part of one's chat identity. Furthermore, because an overwhelming majority of chatters read and write both varieties, the varieties can be specialized for different communicative tasks.

A closer view of various IRC channels reveals that the use of standard German and Swiss German dialects not only varies across individuals but over time, leading to a fluctuation of the dominant variety within a particular channel. Depending on both the current and general linguistic situation of a channel, the language for each individual contribution is chosen against a different linguistic background. In this article, I focus on a quantitative analysis of these background aspects that should be considered for a thorough qualitative analysis of IRC in situations where two varieties co-exist. This analysis is exemplified with data from Swiss-German IRC channels.

Method

Two different approaches are used for the analysis: On the one hand, a static quantitative approach is selected to establish the dominant language variety of a channel, to compare the different channels, and to establish the linguistic preferences of individual chatters. On the other hand, a windowed or moving average analysis is used to focus on situations with fluctuating dominance of the varieties. These diverse analyses allow for a general view on the use of different language varieties in a specific IRC space. In so doing, they provide the contextual background for further qualitative insights.

For this analysis, the varieties are identified at the word level.[4] A binary distinction is made between standard German on the one hand and the sum of all Swiss German dialects on the other. For this study of IRC in German-speaking Switzerland, a word list was compiled and used as a comparison basis for a computer program that went through the entire corpus. The list contains 70 standard German words[5] with corresponding Swiss-German variants from all dialects. Words consist of auxiliaries, highly frequent verbs, indefinite pronouns, some adjectives, prepositions, and a few nouns. Since the result should be quantitative in nature, a high frequency of the words in the corpus is a basic requirement. However, even highly frequent words must be unambiguously assigned to one of the two varieties, if they are to be considered. This is not possible if a word is identical in both standard German and Swiss-German dialects, or if an unrelated word with identical spelling exists in the other variety.

A number of more specific problems may arise in compiling the word list. First, words that are the same in standard German and a dialect cannot be used. For instance, the standard German pronoun *er* 'he' /eːɐ̯/ corresponds to /æːr/ in western Swiss German dialects, which is usually written *är*. This would be a perfectly unambiguous relation, except that the form of the same word in the eastern dialects is /ɛːr/, which is normally

written *er*. It therefore cannot be included. Moreover, as standard spelling influences chat usage in the corpus, chatters who speak a western dialect may also use the spelling *er*.

Second, there can be several corresponding Swiss German variants for a single standard German word. An example is the standard German *musst* '(you) must', which has several dialectal equivalents, including *muesch/muäsch/musch/muescht/muest*. As long as we have a 1:n relation, the correspondence can be assured with an exhaustive list. However, if a standard German word corresponds to an unrelated dialectal word form, this word cannot be used. For example, standard German *aus* 'out' corresponds to *us/uus/uss* in all Swiss-German dialects. However, in western dialects where the *l* is vocalized in the geminate, the spelling *aus* can also represent standard German *alles* 'everything'. In this case, it cannot be decided on the word level if an instance of *aus* corresponds to standard German *aus* 'out' or to western Swiss German *aus* 'everything'. Consequently, both 'out' and 'everything' must be excluded from the list.

Third, standard German spelling learned in school may influence spontaneous dialectal writing, which is usually oriented to pronunciation. An example is the spelling distinction *e* and *ä*, which corresponds to a phonological contrast between /æ/ and /ɛ/ or /ɛ/ and /e/. Standard German *Leben* ('life; we/they live') and *lebe* ('I live') are pronounced as [læb̥ə / læːb̥ə / lɛb̥ə / lɛːb̥ə / lɛəb̥ə / lab̥ə] in the Swiss German dialects. The representations with an /ɛ/ sound are only found in dialects without the /æ/ sound (Moulton, 1960). In these dialects /ɛ/ is phonologically opposed to /e/. Therefore, a representation with the grapheme *ä* (or *ae*) is obvious, and it is the most frequent representation in the corpus. Nevertheless, *e* spellings can also be found in dialect messages; this has to be seen as an influence of standard German orthography. The confusion of the *e/ä* distinction renders impossible a dialect/standard distinction of words on the basis of the distribution of these graphemes alone.

The fourth problem is more subtle. Different grammatical constructions may bias the results, as is the case with distinctive past tense markings in the two varieties. The simple past tense in written standard German has been replaced by the present perfect tense in the southern German dialects, including Swiss German. As a result, *ich war* ('I was') and *i bi gsii* ('I have been') are corresponding forms. The different number of words in the two varieties may bias the result of the variety detection algorithm, as more Swiss German words will be detected than standard German words.

Fifth, typing errors and spelling mistakes may also blur results. For example, the word for 'good, well' is *gut* in standard German and *guet,* with a diphthong, in Swiss German dialects. Nevertheless, we find several occurrences of the spelling *gut* in otherwise entirely dialect messages written by chatters who use only dialect in all other messages. These instances are considered to be cases of likely misspellings. The possibility of calculation errors due to such misspellings cannot be excluded.

Finally, in the channels #hiphop and #teentalk, 'leet speak' (*13375p34k*)—a visual encryption code where letters are replaced by similar looking ciphers (h2g2, 2002)—is sometimes used. While preserving readability by humans who know how the code works, 'leet speak' prevents the automatic recognition of the linguistic variety of given words (Perea, Duñabeitia, & Carreiras, 2008). Leet speak is mainly used in nouns and formulae of phatic communication, and leet speak words are often written according to English pronunciation. However, they are not very common in the mentioned list. Therefore, the errors introduced by the use of leet speak are minimal.

These possible errors notwithstanding, most of these issues can be dealt with, thus validating the automatic allocation process. About 10% of all words used in the corpus can be assigned to either variant—a Swiss German dialect or standard German—by implementing this procedure. This 10% seems to be a rather low percentage, and one may ask if that is enough to validate the procedure. A sample manual classification of the data supports this approach. The results of a 10-minute extract are given in Table 1.

Level	all words – manually classified		unambiguous words – manually classified		automatically classified	
	Count	%	Count	%	Count	%
Swiss German dialect	287	20	287	57	91	52
Standard German	219	15	219	43	85	48
Standard or SG dialect	616	43				
Greeting	50	4				
Nickname	200	14				
Foreign word	23	2				
Short name	30	2				
Total	1425	100	506	100	176	100

Table 1. Comparison of manual and automated classification of 283 messages (May 3, 2007 22:00-22:10 #flirt40plus)

Table 1 shows that about one-third of all words can be unambiguously attributed on the word level to either standard German or to a Swiss German dialect. Well over one-third have a corresponding form in the other variety, so that a clear classification is not possible without resorting to syntactic criteria or the variety of the surrounding words. The rest are nicknames, greetings (e.g., *bye, cu* engl. 'see you'*, ciao, salut*—with local forms such as *sali, salü, säle*), and abbreviations and elements typical of CMC with features from dialect, standard German, French, English, Italian, etc. (*LOL, g, haha, hdg* = Swiss German *ha di gärn* 'I love you'). Comparing the manual and the automated classification reveals an error of 5%. An additional test using part of a log file where no standard German forms were automatically detected (165 messages, May 3, 2007 22:20-22:30 #bern) was successful, as the manual scan did not disclose a single error. Hence, the automatic process is stable enough to be used for the calculation of dialect ratio.

The dialect-to-standard ratios computed in this manner were attributed to different channels in order to provide an overview of regional and age-defined channels. They were also attributed to individual chatters to characterize chatters' language choice(s) and to different message types and time slots with the aim of tracing the dialect/standard distribution over time.

Data

Thirteen IRC channels running on the server bluewin.ch have been logged since 2002. The channels are accessible via a specific IRC client or java applet in any browser. The 13 channels were recorded for 200 to 700 hours within one month every year until 2008, and two different kinds of channels were tracked:

a) regional channels named after a town or area; most chatters in these channels are between 15 and 25 years old and come from the region indicated in the chat channel name (#bern, #basel, #zuerich, #aargau, #wallis, #graubuenden), and

b) supraregional channels named after a special interest or age group (#teentalk, #hiphop, #flirt20plus, #flirt30plus, #flirt40plus, #flirt50plus, #flirt60plus).

Altogether, nine million messages containing 41.7 million words were recorded.

Tables 2 and 3 below outline the activity of the channels of each type; empty cells indicate years in which a given channel was not recorded. The numbers represent messages per hour for the entire recorded period. This means that the changing intensity of activities over the course of the day was not considered. Nevertheless, it is evident that most channels experienced their most active period in 2005-2006. In 2008, there are only six regularly active chatters in #hiphop and #wallis who contribute more than 50 messages, with most other messages coming from casual chatters.

Year	teentalk	hiphop	flirt20plus	flirt30plus	flirt40plus	flirt50plus	flirt60plus
2002							
2003					884.9		
2004		31.4	256.9	676	843.4	558.4	133.3
2005	838.7	18.2	253.6	782	891.1	607.6	227
2006	969	6.3	307.3	710	850.8	687	262.5
2007	710.4	1.7	302.1	354.4	762.5	624	127
2008	679	2	289.8	550.7	862.2	659.9	136.6

Table 2. Activity: Messages per hour in the supraregional IRC channels by year

Year	aargau	bern	graubuenden	wallis	zuerich	basel
2002		217.1	107.9	144.3	169.9	
2003	128.6	289.7			310.4	
2004	101.9	217.8	158.4	102.7	289.3	
2005	162.8	178.4	255	89.2	295.6	247.2
2006	27.7	195.6	161.4	99.9	256.4	182.5
2007	14.7	146.1	29	4.9	92.6	52.2
2008	9.2	124.3	15.1	3.2	73.1	17.7

Table 3. Activity: Messages per hour in the regional IRC channels by year

Results

This section presents some quantitative results based on the method discussed above. These results allow for a general interpretation of the linguistic situation in Swiss IRC, while simultaneously providing the contextual background for analyzing individual chatters' behavior.

Dialect Share of the Channels

First, the dialect share of the different channels is calculated. Figure 1 shows the mean dialect share of the different channels by year. In general, the regional channels have a higher share of dialect (69%-95%) than the supraregional channels (43%-92%), which show a greater spread. The dialect share can vary from year to year, although the relative position of the channels is quite stable. The development over the years is not very clear; nevertheless, it seems that the use of dialect reached its peak in 2005 for most channels and has been declining ever since. Exceptions are #bern, #flirt60plus, and #hiphop, which are discussed below.
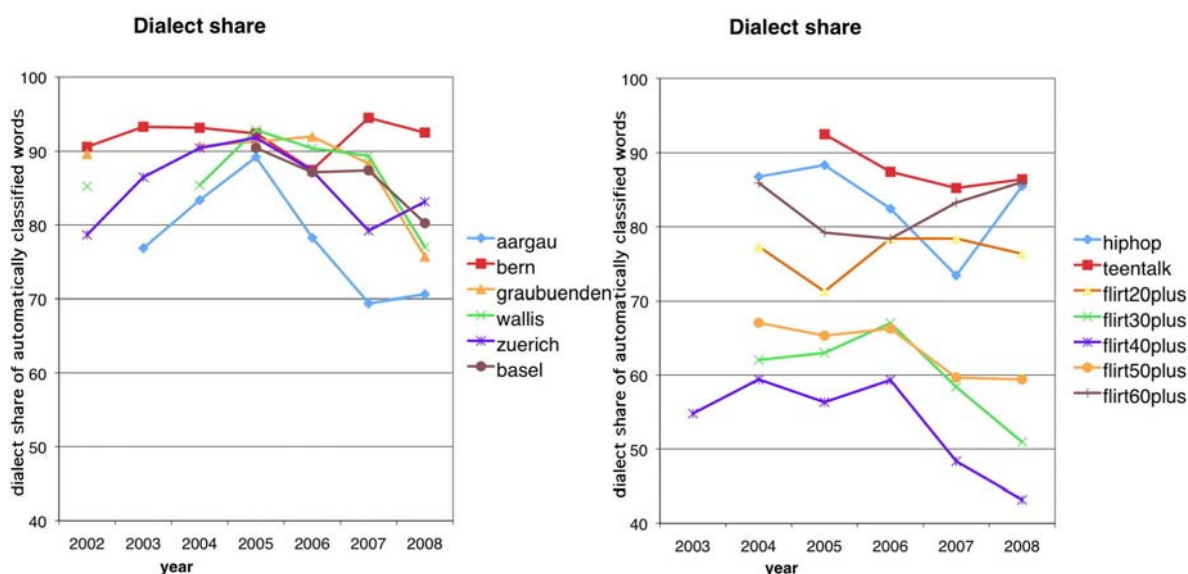

Figure 1. Dialect share of regional (left) and supraregional (right) IRC channels by year

The question then arises: Where do the differences among the channels derive from? The regional channel #bern, which covers the region of the capital Berne, has one of the highest dialect shares recorded. Moreover, dialect shares ascend in #bern after dialect shares descend in other channels. #wallis and #graubuenden, which cover alpine regions, generally have a high dialect share despite a drop in the last year. The channel #aargau, which represents a rural and suburban area between the centers Zurich, Basel, and Berne, always has the lowest dialect share. #zuerich and #basel, which cover most urban centers, occupy a median position. Berne has traditionally had a strong dialectal awareness that includes literature and a wide acceptance of dialect use in current pop music, which sets it apart from the other midland cantons of Zurich and Aargau. The strong position of the Bernese dialect is reflected in the high dialect share in the channel #bern. Moreover, the cantons of Zurich, Basel, and Aargau border Germany and may therefore also attract German chatters who do not have to abandon standard German, even though Aschwanden (2001, p. 61) reports a case of a German chatter who has learned to chat in Swiss German.

The fall of the dialectal share in #hiphop may be a consequence of the radical decrease in channel activity. While in 2004 there were 31 messages per hour on average, the average dropped to two messages per hour in 2008. As a result, this channel is nearly inactive, and its results are no longer representative.
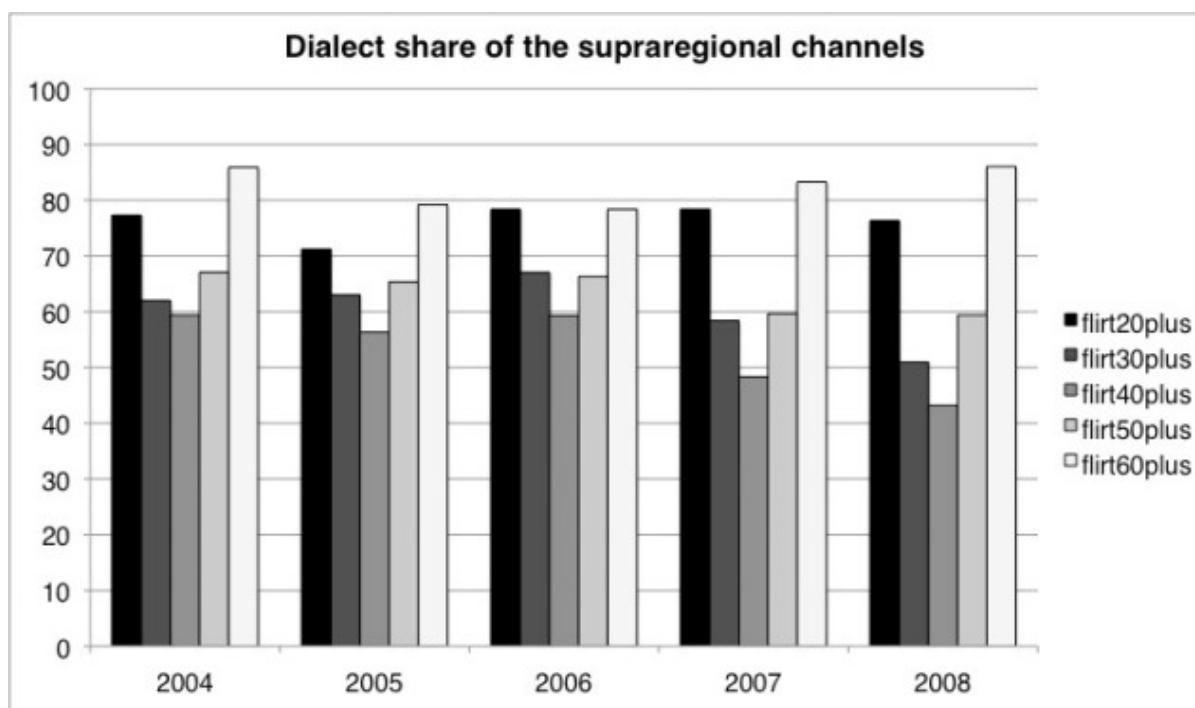

Figure 2. Dialect share of the supraregional flirt channels by year

In comparing the age groups of the supraregional flirt channels by year (Figure 2), we find a standard-to-dialect ratio in the shape of a U, with the younger and older chatters at the high points of the U and the middle age group at the low point. This shape is typical of a situation of stable variation (Labov, 2001). The middle-aged group

#flirt40plus uses more standard German, i.e., the variety associated with overt prestige for writing, while younger (#flirt20plus) and older (#flirt60plus) chatters use more dialect. Thus, with reference to the apparent-time hypothesis, the distribution does not indicate language change, which would entail a rise or fall from younger to older chatters, but rather variation that remains stable over time. Still, this result is surprising, given that dialect writing in private communication only emerged in the early 1980s.

We can also look at the data from a real-time change perspective. Although the data collected in the last five years do not necessarily allow for a stable real-time interpretation, they reveal certain tendencies. More specifically, the decline of the dialect share for the three middle-age groups and in most of the regional groups in the last three years can be seen in the wider context of official pressure in favor of standard German since the publication of the 2000 PISA study.[6] The groups with a higher dialect share are more resistant to this pressure from the outset and therefore adhere to their dialect use, while the other groups give in to this pressure more easily. This is only a short-term development in the data, however, and needs to be documented in the future.

Dialect Share of Individual Chatters

The data reveal differences not only among the channels but also among chatters on the same channel (see Figure 3).
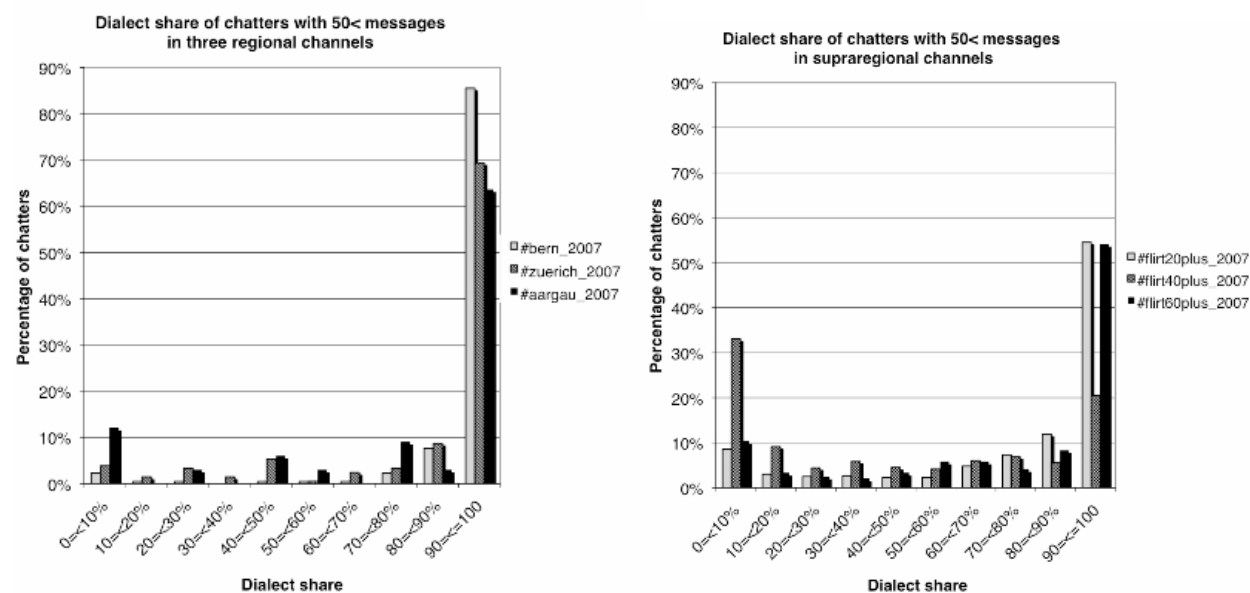

Figure 3. Dialect share in percentage of chatters with more than 50 messages
(N: #bern = 222; #aargau = 33; #zuerich = 208; #flirt20plus = 562;
#flirt40plus = 1132; flirt60plus = 241)

Figure 3 shows the dialect-to-standard ratio of all chatters with more than 50 messages for three regional channels (left graph) and three age-related channels (right graph). The regional channels show a clear preference for the dialect, with little divergence. The

age-related channels show a bimodal distribution with a clear overall preference for dialect, with the exception of #flirt40plus, where one third of the chatters almost exclusively use standard German; in all other channels this value ranges between 2% and 12%. More than half of all chatters in the supraregional channels use both varieties, as compared to the regional channels where this value is much lower. In contrast to the results presented in Siebenhaar (2006), which focused on data collected between 2003 and 2005, the more recent data in this study show an accentuation of the two edges, which means that more chatters adhere to one particular variety. This is quite obvious for #bern, where 86% of chatters hardly use a standard German word. In contrast, one third of all chatters in #flirt40plus only use standard German, and their amount has increased in recent years.

Dialect Share in Time

Figure 4 shows that the dialect share is not a fixed value, but changes over time. The figures track the value for the evening of May 3, 2007, between 8 p.m. and midnight local time. The dialect share is again gauged with the automated detection algorithm described above. The value is calculated every minute and smoothed within a three-minute window for the sake of clarity. To shed light on the relationship between activity and dialect share, the figure also shows the relative number of messages. The highest value per minute within this four-hour period is set to 1; this value is smoothed with a three-minute window as well. The dialect share is represented in blue, the relative number of messages in red.
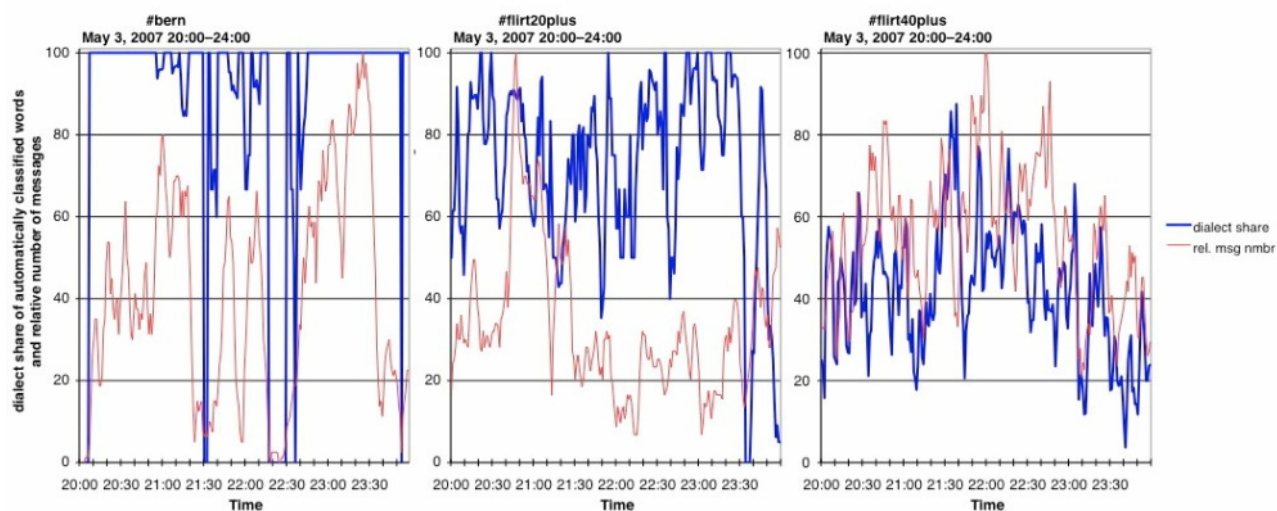

Figure 4. Dialect share (blue) and relative number of messages (red) over time for the evening of May 3, 2007 on #bern, #flirt20plus, and #flirt40plus

The development in these three channels is clearly different. In #bern, the dialect share is at 100% most of the time, with occasional lower activity rates causing the dialect share to drop to a very low value. For the entire month that was recorded in 2007, there is a moderate correlation of activity and dialect share (Pearson product-moment correlation coefficient r=0.41). In #flirt20plus, this correlation is much lower (r=0.31), and

for #flirt40plus it is significant, although almost not remarkable (r=0.14). Nevertheless, these correlations suggest that higher activity rates lead to a higher dialect share. The dialect share of channel #flirt20plus is less stable than that of channel #bern, while channel #flirt40plus has a similar volatility, albeit on a lower level.

Consequences for Qualitative Analysis

These quantitative results can provide a useful background for qualitative analyses. The results shown here clearly illustrate that the selection of analyzed time slots, channels, and chatters can have a major impact on the patterns observed. Selecting a dataset that is too small or limited in other ways may bias results towards one particular variety, while a very large dataset may blur internal differences. These impacts are well known (Hakim, 2000; Kleining & Witt, 2001). The analysis presented in this article demonstrates one solution to these problems. A quantitative analysis can lay the groundwork for selecting a representative data extract for purposes of qualitative analysis. More specifically, the quantitative analysis allows for selecting a part of a log file that is representative for the corpus with regard to the distribution of language varieties. It allows picking out certain chatters with a specific mix of varieties for an individual-centered analysis. When searching for log-file passages containing language alternations for a qualitative analysis of code-switching, the quantitative approach allows broadening a limited selection. Such qualitative analyses become more useful and convincing when they are embedded through quantitative means. On the one hand, possible gaps in the analysis may be discovered and subsequently filled. On the other hand, the status of the analyzed parts within the whole dataset can be described.

A code-switch from dialect to standard German has differing meanings in #bern, #flirt60plus, and #flirt40plus. A code-switch also acquires a distinctive connotation when it is performed by a chatter who uses dialect in 95% of his/her messages, as compared to a chatter who uses both dialect and standard German. It makes a difference in meaning if different chatters use mainly dialect or if they chat in both varieties. Moreover, the meaning of a particular code-switch depends on whether the interlocutors use predominantly standard German or predominantly dialect. In this sense, a quantitative approach can provide a detailed and contextualized background against which a qualitative analysis may be carried out. This approach not only assists in terms of embedding the qualitative results, but also facilitates the process of selecting the relevant data required for a qualitative analysis. Suppose one is interested in factors favoring a switch from one variety to the other in a bilingual chat room. One's findings suggest that political topics or the presence of specific chatters favor such a switch. With a quantitative analysis of the type presented here, it will easily be possible to find other occurrences of code-switching in the log file in order to verify these findings. Conversely, the qualitative results could serve as the basis for an interactional explanation of the quantitative results. The question of why certain chatters, identified using the quantitative method, use both varieties (see Figure 3) can only be answered through a qualitative analysis of the interaction. However, this qualitative research is not presented here. A longer qualitative analysis of data focusing on code-switching and language choice in Swiss German IRC rooms from 2003 is presented in Siebenhaar

(2005), and a shorter analysis of data from May 2005, with an English translation, can be found in Siebenhaar (2006).

Notes

1. Rash (1998, p. 50) emphasizes these exceptions in the domain of oral communication and proposes the term 'functional diglossia'. Developments in the language of new media within the last 10 years, especially the highly frequent use of dialect in chat communication and text messaging, support this position.

2. Hotzenköcherle (1984) describes only eight regions. The ninth is omitted, as he died before the manuscript was finished.

3. Dieth (1938/1986) and Marti (1972) may be known by dialect researchers or linguists. As I have shown for IRC, those rules only apply to the extent that they correspond to general orthographic principles (Siebenhaar, 2003).

4. A word is defined as a string of characters between two delimiting characters (spaces, punctuation marks, end of line)

5. The list consists of the following standard word forms: *habe, hab, hast, hat, hatte, hattest, hätte, hättest, gehabt* ('have' 1-3 sing. present, past, past subjunctive, past participle); *bist, ist, war, warst, gewesen* ('be', 2-3 sing. present, 1-3 past, past participle); *muss, must, musst* ('must' 1-3 sing. present); *komme, komm, kommst, kommen, kommt, käme, kämest, kämen, kämt* ('come' 1-3 sing. pl. present, 1-3 sing. pl. past subjunctive); *kann, kannst, können, könnt* ('can' 1-3 sing., 1-3 pl. present); *gesagt* ('say' past participle); *willst* ('will' 2 sing. present); *geht, gehts* ('go' 3 sing. present); *machst* ('make' 2 sing. present); *weisst* ('know' 2 sing. present); *gibst, gibt, giebt, gibts, giebts* ('give' 2-3 sing. – including misspellings), *schreib-, (-)schrieb(-)* ('write' 1-3 sing., pl. present, past, past participle); *nicht* ('not'); *nichts* ('nothing'); *jemand* ('somebody, anybody'); *etwas* ('something, anything'); *etwa* ('about'); *gut, gute, guter, guten, gutes* ('good'); *auch* ('too'); *wirklich* ('really'); *weiter* ('further'); *auf, aufs* ('on'); *hinauf, rauf* ('up'); *zusammen* ('together'); *hinab* ('down'); *oben* ('at the top'); *heute* ('today'); *schon* ('already'); *Zeit* ('time'); *Freund* ('friend'), and *Abend* ('evening').

6. The Programme for International Student Assessment (PISA) is an internationally standardized assessment of the scholastic performance of 15-year-old schoolchildren. In 2000, mediocre test results on reading skills led to a heated debate on the issue. As a result, the 26 cantonal ministers of education called for a strengthening of the standard language in Switzerland (EDK, 2003).

References

Androutsopoulos, J., & Hinnenkamp, V. (2001). Code-Switching in der bilingualen Chat-Kommunikation: ein explorativer Blick auf #hellas und #turks. In M. Beißwenger (Ed.), *Chat-Kommunikation* (pp. 367–402). Stuttgart: ibidem.

Aschwanden, B. (2001). *"Wär wot chätä?" Zum Sprachverhalten deutschschweizerischer Chatter*. Online (Networx 24). Retrieved June 12, 2008, from http://www.mediensprache.net/de/networx/docs/networx-24.asp

Bays, H. (1998). Framing and face in Internet exchanges: A socio-cognitive approach. *Linguistik online*, 1/98. Retrieved September 24, 2008, from http://www.linguistik-online.com/bays.htm

Börlin, R. (1987). *Die schweizerdeutsche Mundartforschung 1960–1982. Bibliographisches Handbuch.* Aarau, Frankfurt am Main, Salzburg: Sauerländer (Sprachlandschaft 5).

Christen, H. (1998). *Dialekt im Alltag. Eine empirische Untersuchung zur lokalen Komponente heutiger schweizerdeutscher Varietäten.* Tübingen: Niemeyer.

Christen, H. (2004). Dialekt-Schreiben oder sorry ech hassä Text schribä. In E. Glaser, P. Ott, & R. Schwarzenbach (Eds.), *Alemannisch im Sprachvergleich. Beiträge zur 14. Arbeitstagung für alemannische Dialektologie in Männedorf (Zürich) vom 16. - 18.9.2002* (pp. 71–85). Stuttgart: Franz Steiner (ZDL-Beiheft 129).

Christen, H., Tophinke, D., & Ziegler, E. (2005). Chat und regionale Identität. In S. Krämer-Neubert & N. R. Wolf (Eds.), *Bayerische Dialektologie. Akten der Internationalen Dialektologischen Konferenz 26.–28. Februar 2002* (pp. 425–438). Heidelberg: Winter.

Dieth, E. (1938). *Schwyzertütschi Dialäktschrift.* Zürich: Orell Füssli. [Dieth, E. (1986). *Schwyzertütschi Dialäktschrift. Dieth-Schreibung.* 2nd ed., adapted and edited by C. Schmid-Cadalbert. Aarau: Sauerländer.]

EDK, Schweizerische Konferenz der kantonalen Erziehungsdirektoren (2003). *Aktionsplan "PISA 2000"-Folgemassnahmen.* Retrieved September 30, 2008, from http://www.edudoc.ch/static/web/arbeiten/pisa2000_aktplan_d.pdf

Ferguson, C. (1959). Diglossia. *Word, 15*, 325–340.

Grondelaers, S., Geeraerts, D., Speelman, D., & Tummers, J. (2001). Lexical standardisation in internet conversations. Comparing Belgium and The Netherlands. In J. M. Fontana, L. McNally, M. T. Turell, & E. Vallduví (Eds.), *Proceedings of the First International Conference on Language Variation in Europe* (pp. 90–100). Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, Unitat de Investigació de Variació Lingüística.

h2g2. (2002, August 16). *An explanation of l33t speak.* Retrieved September 25, 2008 from http://www.bbc.co.uk/dna/h2g2/A787917

Hakim, C. (2000). *Research design: Successful designs for social and economic research.* 2nd ed. London: Routledge.

Hentschel, E. (1998). Communication on IRC. *Linguistik online,* 1/98. Retrieved September 24, 2008, http://www.linguistik-online.de/irc.htm

Hotzenköcherle, R. (1984). *Die Sprachlandschaften der Schweiz.* Aarau, Frankfurt am Main, Salzburg: Sauerländer.

Kleining, G., & Witt, H. (2001). Discovery as basic methodology of qualitative and quantitative research. *Forum: Qualitative Social Research*, 2 (1). Art. 16. Retrieved September 30, 2008, from http://www.qualitative-research.net/index.php/fqs/article/view/977/2130

Kolde, G. (1981). *Sprachkontakte in gemischtsprachigen Städten.* Wiesbaden: Franz Steiner (ZDL Beihefte 37).

Labov, W. (2001). *Principles of linguistic change. Volume 2: Social factors.* Oxford, UK: Blackwell.

Lam, W. S. E. (2004). Second language socialization in a bilingual chat room: Global and local considerations. *Language Learning & Technology, 8*(3), 44–65. Retrieved September 24, 2008, from http://llt.msu.edu/vol8num3/pdf/lam.pdf

Marti, W. (1972). *Bärndütschi Schrybwys.* Bern: Francke.

McLellan, J. (2005). *Malay-English language alternation in two Brunei Darussalam on-line discussion forums.* Retrieved September, 27, 2008, from http://adt.curtin.edu.au/theses/available/adt-WCU20060213.092157/

Moulton, W. G. (1960). The short vowel systems of northern Switzerland. *Word,* 16, 155-182.

Paolillo, J. C. (2001). Language variation on Internet Relay Chat: A social network approach. *Journal of Sociolinguistics, 5*(2), 108–213.

Perea, M., Duñabeitia, J., & Carreiras, M. (2008). Observations R34D1NG W0RD5 W1TH NUMB3R5. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 237–241.

Rash, F. J. (1998). *The German language in Switzerland: Multilingualism, diglossia and variation.* Bern: Lang.

Runkehl, J., Schlobinski, P., & Siever, T. (1998). Sprache und Kommunikation im Internet. *Muttersprache. Vierteljahresschrift für deutsche Sprache* 2, 97-109.

Siebenhaar, B. (2003). Sprachgeographische Aspekte der Morphologie und Verschriftung in schweizerdeutschen Chats. *Linguistik online, 15*, 125-139. Retrieved June 12, 2008, from http://www.linguistik-online.com/15_03/siebenhaar.pdf

Siebenhaar, B. (2005). *Varietätenwahl und Code-Switching in Deutschschweizer Chatkanälen. Quantitative und qualitative Analysen.* Online (Networx 43). Retrieved June 12, 2008, from http://www.mediensprache.net/de/networx/docs/networx-43.asp

Siebenhaar, B. (2006). Code choice and code-switching in Swiss-German Internet Relay Chat rooms. *Journal of Sociolinguistics, 10*(4), 481-506.

Sonderegger, S. (1962). *Die schweizerdeutsche Mundartforschung 1800–1959. Bibliographisches Handbuch mit Inhaltsangaben.* Frauenfeld: Huber & Co. (Beiträge zur schweizerdeutschen Mundartforschung 12).

Sprachatlas der deutschen Schweiz (SDS). Begründet von Heinrich Baumgartner und Rudolf Hotzenköcherle. In Zusammenarbeit mit Konrad Lobeck, Robert Schläpfer, Rudolf Trüb und unter Mitwirkung von Paul Zinsli herausgegeben von Rudolf Hotzenköcherle. 1962-1997. Bern, vol. VII and VIII Basel: Francke.

Warschauer, M., El Said, G. R., & Zohry, A. (2002). Language choice online: Globalization and identity in Egypt. *Journal of Computer-Mediated Communication, 7*(4). Retrieved September 22, 2008, from http://jcmc.indiana.edu/vol7/issue4/warschauer.html

Werry, C. C. (1996). Linguistic and interactional features of Internet Relay Chat. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (pp. 47-64). Amsterdam /Philadelphia: John Benjamins.

Biographical Note

Beat Siebenhaar [siebenhaar@uni-leipzig.de] is a professor of German linguistics in the German Department at the University of Leipzig. His research interests include all aspects of linguistic variation, from classical dialectology to sociolinguistics, from language change to computer-mediated communication, and from prosody to style.