Quantitative approaches on linguistic variation in IRC – implications on qualitative research
Beat Siebenhaar

Introduction

German speaking Switzerland is one of the classical examples for a diglossic situation. Yet, the distribution of the two varieties, standard German and Swiss German, is not a distribution of a high-prestige variety and a low-prestige variety, because the Swiss-German dialects are highly prestigious; rather, it is a distribution of a written and a spoken variety. Therefore, one speaks about a medial diglossia that has some exceptions: In class at school, in certain of TV-programs as well as in discussions with non-dialect speakers, standard German is the required spoken variety. On the other hand, non-standardized dialect writing has established itself for personal communication during the last 20 years. But generally, standard German is used for writing and the Swiss-German dialects are used for speaking. However, these dialects are not a uniform Swiss German; they show a great amount of variation that is documented in the eight volumes of the Sprachatlas der deutschen Schweiz (1962–1997) and many thousands of publications.

This coexistence of the two varieties is mirrored in IRC and it corresponds more to the situation of oral practice than to the written practices. The very first view on Swiss-German IRC reveals that a large share of the communication is not in standard German but in the dialects. This dialect writing is not uniform. A norm for dialect writing does not exist or is not known to the chatters respectively. Accordingly, every chatter has his own dialect writing system, where standard German orthographic principles, phonologic and also phonetic assumptions on the own dialect, difference marking to standard orthography, morphologic regularity, reader perspective and many more influences are mixed in an apparently arbitrary and varying way. This favor for the dialect is a clear marker for or expression of the informal language in IRC. Moreover the non-standardized dialect writing allows for a individualizing on the linguistic level, as one personal trait of every chatter is his recognizable dialect. In this context, standard writing becomes a part of the web personality, too. On top of that, as the overwhelming majority of the chatters reads and writes both varieties they can be functionalized for different communicative tasks.

A closer view on different IRC channels shows that the use of the varieties changes over time, and the chatters make different uses of the varieties. So the dominant variety of a channel fluctuates. Therefore, the language choice for every individual contribution is framed differently depending on the actual and the general linguistic situation of a channel. In this paper I will focus on a quantitative analysis of these framing aspects that have to be respected for a qualitative analysis of IRC in situations where two varieties stand side by side. It is exemplified with Swiss German IRC channels.

---

1 Rash (1998) underlines these exceptions in the oral domain and argues for the term 'functional diglossia'. The developments within the last ten years in the new media, especially the highly frequent use of the dialect in chat communication and in SMS, strengthens this position.

2 Dieth (1938/1986) and Marti (1972) may be known by dialect autors or linguists. As I have shown in IRC those rules are only applied as far as they correspond to general orthographic principles (Siebenhaar 2003).
Against this background I will address the following research question:

• What is the dominant variety in Swiss-German IRC?
• What is the Swiss-German IRC community?
• How do different channels differ?
• How does the individual linguistically fit into the community?
• How does he personalize himself?

I will test the hypotheses:

• Different channels show a different 'normal' linguistic behavior.
• The main varieties have different functions in different channels.
• The individual personalizes himself with linguistic means.
• Code-switches are part of this personalization.

Method

A quantitative approach is selected to establish the dominant variety of a channel and to compare the different channels, as well as for the linguistic preferences of the individual speakers. To focus on situations where the dominant varieties change a windowed or moving average analysis is used. These give the background for some exemplarily qualitative insights.

As basis for all analyses the identification of a variety – Swiss German dialect or standard German – has been made on the word level3. Not every word can be assigned unambiguously to a variety due to the fact that Standard German and dialect words are sometimes identical. A list containing seventy Standard German words that have a straightforward correspondence in the dialect was therefore compiled to be used as a comparison basis for a computer program which will go through the entire corpus. This list consists of auxiliaries, highly frequent verbs, indefinite pronouns, some adjectives, prepositions and a few nouns.

Some problems with this approach shall be mentioned. a) Words without a difference in standard and in one of the dialects cannot be used (er - är/er 'he'). b) For one standard German word there can be many corresponding Swiss German variants, the list has to comprise all of them (musst – muesch/muäsch/musch/muescht/muest '(you) must'), and can be very long. c) A standard word corresponds to a non-related dialectal word form (standard German aus 'out' – western dialects us/uss/uus/uss), this cannot be used. d) A different grammatical construction may bias the results (e.g. different past tense marking: ich war – i bigsii 'I was – I have been'). e) Typing errors and spelling mistakes can blur the results (gut – guet 'well').

With this procedure about 10% of all words used in my data can be assigned to either of the variants. The data were attributed to different channels to give a general view on the data, the were attributed to individual chatters to characterize the chatters language choice, they were attributed to different message types, and they were attributed to time slots in order to trace the dialect-standard-distribution over time.

Data

Since 2002 I have logged 13 IRC channels, run by bluewin.ch. The channels are accessible via a specific IRC application or via a java-applet in any browser. Each year, the respective

3 A word is defined as a string of characters between two delimiting characters (space, punctuation marks, end of line)
channels were recorded for 200 to 700 hours within one month. Two different kinds of channels were traced:

a) regional channels named after a town or an area, with most chatters aged between 15 and 25, and coming from the respective region (#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).

Over all 6.6 Mio messages were recorded with 30.5 Mio words.

**Results**

**Dialect share of the channels**

Figure 1 shows the mean dialect share of the different channels by year. In general, the regional channels have a higher dialect share (69%–95%) than the supraregional channels (49%–92%), which have a greater spread. From year to year, the dialect share can vary, albeit the relative position of the channels is quite stable. The development over the years is not very sound; nevertheless, it seems that in most channels, the dialect share has declined for the last or the last two years.

Where do the differences between the channels come from? Within the regional channels #bern, which covers the region of the capital Berne, has mainly one of the highest dialect shares, #wallis and #graubuenden – alpine regions – are on the following positions. #aargau, which represents a rural and suburban area between the centers Zurich, Basel, and Berne, always has the lowest dialect share. #zurich and #basel with the most urban centers have a mean position. Berne has traditionally had strong dialectal awareness, including 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. Moreover, Zurich, Basel, and Aargau are cantons that border on Germany, which may also attract German chatters, who do not have to abandon Standard German.

![Figure 1: Dialect share of regional (left) and supraregional (right) IRC channels by year](image-url)
Comparing the age groups of the supraregional channels in figure 2, for every year we find a standard-dialect ratio in the form of a U with the younger and older chatters at the tails of the U and the middle age group at the bow. This form is regarded as typical of a stable variation (Labov 2001: 101 f.). The middle-aged group #flirt40plus uses more Standard German, the variety associated with overt prestige for writing, while younger (#flirt20plus) and older chatters (#flirt60plus) use more dialect. Thus, with reference to the apparent-time hypothesis, the distribution does not point to a language change, but to a variation that remains stable over time. Still, this result is surprising since dialect writing in private communication only emerged in the early 1980s. We also can look at the data with a real-time perspective. Yet, the data collected in the last five years do not really allow for a stable real-time interpretation. The decline of the dialect share for the three middle groups and in most of the regional groups in the last two years is still remarkable and can be seen in the general context of an official pressure towards a reinforced use of Standard German since the publication of the PISA-studies 2000 and 2003. However, this is only a short-term development in the data and must be followed in the next years.

**Dialect share of individual chatters**

Figure 3: Dialect share in percent 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-standard ratio of all chatters with more than 50 messages of three regional channels on the left and of three age related channel on the right. All channels show a two tail distribution with a favor for the dialect, with the exception of #flirt40plus where one third of the chatters almost exclusively use the dialect; in all other channels this value is between 2 % und 12 %. Nevertheless, more than half of all chatters in all channels use both varieties.

**Dialect share in time**

Figure 4 focuses on the fact that the dialect share is not a fix value, but it changes over time. The figures track the value for the evening of May 3, 2007.

![Figure 4: Dialect share development in time for the evening of May 3, 2007 in #bern, flirt20plus, and flirt40plus](image)

It is obvious that the development in the three channels is different. In #bern the dialect share is at 100 % for most of the time, it is only in times with a lower activity that the dialect share drops to a very low value. For the whole month that was recorded there is a medium correlation of activity and dialect share (r=0.41). In #flirt20plus this correlation is much lower (r = 0.31), for #flirt40plus it is almost not remarkable (r=0.14). The channel #flirt20plus shows a more volatile dialect share, while the channel #flirt40plus has a similar volatility, albeit on a lower level.

**Impact for a qualitative analysis**

These quantitative results reveal a broad background for qualitative analyses. What has been shown here makes clear that the selection of the time slot to be analyzed, the channel, or the chatters to be traced have a major impact on these results. Selecting a dataset that is too small can bias the results to the one or the other variety, while a rather large dataset can blur all internal differences.

So, a code-switch from dialect to standard German has another meaning in channel #bern or in #flirt60plus than in #flirt40plus. It has another meaning when it is performed by a chatter, who uses the dialect in 95% of his messages, than by a chatter, who only uses the dialect in 10% of his messages. It has another meaning if the other chatters, which are present, use mainly the dialect or if they are bi-variety-chatters. It has another meaning if the partners use standard German or if they use the dialect. In this sence, a quantitative approach can give a detailed background to a qualitative analysis. It helps embedding the qualitative results, and on the other hand the qualitative results allow for an interactional explanation of the quantitative results. This is not done here. A longer qualitative analysis of 2003 is given in

References


