Stephan Elspaß and Konstantin Niehaus

The standardization of a modern pluriareal language

Concepts and corpus designs for German and beyond

1 Overview

In this paper, we address the topic "Foundations of Language Standardization" from a theoretical and a methodological point of view, aiming at a more detailed account of standardization processes of a modern pluriareal language like German in the last two hundred years. Part 2 sets off with a short reflection on dominating notions and ideologies about language standardization, which will be contrasted with alternative views. Part 3 introduces a practical corpus design that follows from these considerations. In part 4, we will present three case studies from the standardization history of German which are based on our corpus design. We will compare results from studies on both formal and informal registers from historical (mainly nineteenth century) and present-day German with particular regard to the degree of 'monocentricity' or 'pluriareality' displayed in this corpora. Part 5 has a short conclusion.

2   Conflicting concepts on the standardization and the variation of modern standard languages

Traditionally, textbooks on the histories of ('big') modern standard languages such as English or German present rather monolithic standard varieties as results, closing stages, and sometimes even as objectives of standardization processes. The most powerful linguistic ideologies behind such conceptualisations of standard language and standardization can be identified as 'homogeneism' and 'standard language ideology'. The "dogma of homogeneism" consists of "a view of society in which differences are seen as dangerous and centrifugal and in which the 'best' society is suggested to be one without intergroup differences" (Blommaert & Verschueren 1998:194–5). Although Blommaert & Verschueren have multilingual societies in mind, the basic concept also applies to pluricentric and pluriareal settings. Closely related to homogeneism is the 'standard language ideology', i. e. "a bias toward an abstract, idealized homogeneous language, which is imposed and maintained by dominant bloc institutions and which names as its model the written language" (Lippi-Green 2012:67, mainly based on Milroy & Milroy 1985).

   In the following, we will present two fairly new alternative concepts (or ideologies) to these two ideologies. The first one, 'language history from below', is an approach which in our view can be beneficial for standardization studies of *all* modern standard languages. The pluriareal concept will be less helpful for small languages like Icelandic, but, as we think, essential for the investigation of standardization processes of big(ger) languages. Such alternative approaches could, or rather should have consequences for the design of corpora, as we will demonstrate in Chapter 3, where German will serve as a case for illustration.

2.1 Standardization in a view 'from below'

In textbooks on language histories, standardization processes are often depicted in a teleological way, as long marches toward uniform standard varieties. Moreover, for the modern period such standardization (hi)stories are mostly accounts of the development of *printed* language and thus of conceptually written texts which were written

– or rather: composed – by professional writers. Such histories may be termed "language histories from above". Opposed to such a dominant 'bird's eyes view', there has been a strong plea for 'language histories from below' in the study of standardization (cf. Elspaß 2005, Elspaß, Langer, Scharloth & Vandenbussche 2007). The idea is, firstly, to account for the language use of the vast majority of speakers of modern language communities and its role in standardization processes. Secondly, a view 'from below' advocates a radical shift from conceptually written texts to texts representing the "informal/oral type of linguistic conception", in short "language of immediacy" (Oesterreicher 1997:193–4), as the starting point of standardization studies. "Language of immediacy" differs from "language of distance" with respect to various communicative parameters. Prototypically, such as in an intimate face-to-face conversation, it is characterized by a familiarity between communication partners in private settings and in spatio-temporal proximity, by a free development of topics, spontaneity and a maximum of cooperation between communication partners, etc. (as opposed to unfamiliarity of communication partners in public settings and in spatio-temporal distance, by fixed topics, reflection and a minimum of cooperation between communication partners, etc.). In the context of standardization studies, a shift to a view 'from below' therefore requires a principal focus on historical texts which are as close to 'language of immediacy' and their reflections in grammar and lexis as possible. Private letters, in particular, fulfil this requirement (cf. Elspaß 2012, Elspaß in print). Elspaß (2005) and Nobels (2013) are the first two monograph-size studies applying this approach to periods in the standardization phases of two different Germanic languages, namely nineteenth century German and seventeenth century Dutch respectively.[1]

The second aspect, the inclusion of conceptually oral language (and its inherent variation), is at the heart of the 'principle of viability' which Vilmos Ágel has formulated as a fundamental requirement for the analysis of linguistic phenomena.

> Jede linguistische Beschreibung (bzw. Erklärung) muss mit der Beschreibung (bzw. Erklärung) der Geschichte des zu beschreibenden (bzw. zu erklärenden) Phänomens konform sein.

---

[1] The approach has also proved fruitful for studies on older language periods. Such studies include Schulte (2009) on the younger Fuþark, Ernst & Elspaß (2011) on Old High German glosses and Graser & Tlusty (2012) on Early New High German street songs.

'Every linguistic description (or explanation) has to fit into the de-
scription (or explanation) of the history of the phenomenon under
description (or explanation).'

(Ágel 2001:319, 2003:2)

2.2 Standardization of a modern pluriareal language: the case of German

For 'big' languages such as English or German, an alternative view on
standardization would also account for the co-existence of different
standard language varieties (such as 'the Englishes') and for the idea
of an internal variability of these varieties.

A first step away from the notion of a ‚uniform' standard language
was to conceptualize German as a pluricentric language. Up until the
1980s, many authors saw the standard variety in West Germany as the
only legitimate German standard language. According to the wide-
spread notion of pluricentricity, going back mainly to Clyne (1992),
(pluricentric) German consists of the three 'national varieties' in Ger-
many, Austria and Switzerland and other varieties of standard Ger-
man in smaller countries and regions (Liechtenstein, Luxemburg, East
Belgium and South Tyrol, cf. Ammon, Bickel, Ebner et al. 2004). This
concept, however, has been criticized for different reasons. Firstly, it
is an entirely political concept, based on the notion of *Überdachung* of
the language area by a political state. As for the recent history of Ger-
man, this would have had the somewhat odd consequence that on 3
October 1990, the German language has lost an entire national variety,
namely GDR German, literally overnight. Secondly, from a linguistic
perspective, one may ask whether 'national varieties of German' are
really varieties? One can argue that less than two per cent of varia-
tion in standard German lexis and pronunciation and even less varia-
tion in grammar does hardly make a 'variety'. A third problem from
the linguistic side can be raised in view of empirical evidence. While
some national variants do exist, in a lot of cases the diatopic extension
of Standard German variants is not limited to national borders. They
are not absolute variants, as terms such as 'Germanism', 'Helvetism'
or 'Austriazism' may suggest. Rather, in many cases they are relative
variants, i.e. they are employed in a part of the respective country or
by a part of the members of the speech community only. Moreover,
diatopic standard variants do cross borders, i.e. their distribution is

not restricted to one country alone. 'Pure' national variants may even be the exceptional case in the German speaking countries. Figure 1 (adopted from Scherr & Niehaus 2013:78) may serve to illustrate a more realistic picture of the areal variation in the German-speaking countries.

Because of such empirical evidence, some researchers would prefer to talk of Standard German as a pluriareal language (e.g. Wolf 1994, Scheuringer 1997). This is the approach that we adopt, hence step 2 on the way to our corpus design.

The competing concepts of 'pluricentric German' vs. 'pluriareal German' have sparked a controversy (mainly between German and Austrian scholars) about the historical adequacy of the two models (cf. von Polenz 1999, Reiffenstein 2001). In our view, the debate would profit from empirical evidence from present-day as well as historical data, as we will try to demonstrate in section 4 of this paper.

| area A |
| variant x: 70% |
| variant y: 30% |

area B	area C
variant x: 30%	variant x: 20%
variant y: 70%	variant y: 20%
	variant z: 60%

Fig. 1. Relative areal variation in a pluriareal language community (example)

3   A corpus design for a modern pluriareal language

What follows from such considerations and alternative concepts for the design of corpora to allow for painting a more realistic (or rather adequate) picture of standardization processes? The requirement for such a corpus is that it contains historical as well as present-day data which can help us to explain the areal variation and also the on-going change of a modern language in a long-term perspective. As for German, we introduce a corpus design which both accounts for the pluriareality of German and considers historical data 'from above' as well as 'from below'.

In a 'view from above', we look at written (standard) language

in regional newspapers, thus printed texts. What is new about our project on modern German is that we work with regionally balanced corpora. Surprisingly, whereas corpora on earlier periods of German up until the eighteenth century are all regionally balanced,[2] there are no such corpora for the nineteenth and the twenty-first century. So for present-day German, we use the corpus of our *Variantengrammatik* project which is a joint project of a German, an Austrian and a Swiss team, which are now based at the universities of Zurich, Salzburg and Graz.[3] For the nineteenth and the beginning of the twentieth century, the Salzburg team compiled a small corpus of newspaper texts which we refer to as the *Kaiserreich* ('Empire') corpus. The two corpora are, of course, very different in size: The corpus of the *Variantengrammatik* project comprises more than 600 million words, the *Kaiserreich* corpus so far only 100,000 tokens.

In a view 'from below', we focus on spoken language data or written data which are as close to speech as possible. For present-day German, we used the so-called Pfeffer corpus with interview data mostly from the 1960s (Pfeffer & Lohnes 1984, with a total of 670,000 words) and we also looked at maps from our *Atlas of Colloquial German* (AdA) (cf. Elspaß & Möller 2003ff.). The historical data 'from below' consist of a 880,000 words corpus of nineteenth century emigrant letters (mainly based on the corpus of Elspaß 2005). Again, both corpora are (more or less) regionally balanced.

4   Case studies

Based on the principle that variation is inherent to a modern standard language and with regard to our corpora, our case studies focus on the following research questions:

- How much variation did printed German allow in the nineteenth century?

- Is (and was) the variation in non-standard German similar to or rather different from printed standard German? Which tenden-

---

[2]   Such as the corpora of the Middle High German (1050–1350) and the Early New High German (1350–1650) grammar or the German Manchester Corpus (1650–1800), cf. Paul (2007), Reichmann & Wegera (1993) and Scheible et al. (2011).

[3]   The project is funded by the major research grant organizations of the three countries: the *Schweizerischer Nationalfonds* (SNF) [100015L-134895], the *Deutsche Forschungsgemeinschaft* (DFG) [EL 500/3-1] and the *Austrian Science Fund* (FWF) [I 716-G18].

cies in standard and non-standard German are discernible in the last 150 years?

• Does printed (standard) German change from a stage of 'mono-centricity' to 'pluricentricity', as von Polenz (1989:15) claimed?

Von Polenz' hypothesis needs some explanation here. He assumed that socio-pragmatic criteria played a much bigger role for New High German than for any other period in the history of German. Thus, in his periodization model for New High German, the time between 1770 to 1870 is labelled 'Establishment of an educated bourgeois variety' ("Etablierung der bürgerlichen Bildungssprache"), whereas the 'Establishment of the modern German standard language' took place between 1870 and 1950[4] and was characterized by 'monocentric tendencies' ("Etablierung der modernen Standardsprache, monozentrische Tendenzen"). In contrast, the time after 1950, which he simply called 'Contemporary German' ("Gegenwartsdeutsch"), could be portrayed as the time of 'a polycentric drift' ("polyzentrische Weiterentwicklungen"). Whereas the first phase saw rapidly increasing literacy rates, the second phase can be identified as the crucial period of time of the emergence of a national literature and the standard language ideology (cf. von Polenz 1989:16–27).

To illustrate our approach, we present three case studies, one involving a lexical variable, and two focussing on grammatical variables in standard German. For grammatical variation, we will investigate the serialization patterns in three-verb clusters (4.1) and discontinuous pronominal adverbs (4.2). For lexical variation, we will look at the distribution of variants for 'Saturday' (4.3).

4.1 Grammatical variation I: serialization in three-verb clusters

The serialization of verb clusters in German has received much attention in the research literature.[5] In our case study, we will focus on

---

[4] This is noteworthy, because traditional German textbooks on the history of German would have it that the German standard language had been established by the end of the 18th century.

[5] To name but a few: Bech (1955/57), Van de Velde (1981) and Askedal (1986) discussed the most important theoretical aspects in detail. The benchmark in historical research on verb clusters is still the study by Härd (1981), which was taken up by Ágel (2001) to exemplify some 'historical principles in today's grammar'.

a particular instance of word order variation in verb clusters. Variation may occur in different respects. We will restrict ourselves to the variation of the finite auxiliary verbs *haben* and *werden* in three-verb clusters, all in subordinate clauses with so-called "Ersatzinfinitiv" ('substitutive infinitive') constructions.

A few notes are necessary on some basic theoretical aspects and the 'rules' as codified by today's grammars. In three-verb clusters containing a past participle, the finite verb in subordinate clauses would be put at the end of the clause:

(1)   … *damit er gesehen³ werden² kann¹*
        … so that he seen[3–inf] become[2–inf] can[1–fin]
        'so that he can be seen'

The dependencies within the verbal cluster therefore usually run from right to left. This serialization pattern has proved to be relatively constant in the history of New High German (cf. Härd 1981:13, 167; Ágel 2001:322), and it is also given as a codified present-day 'basic rule' by the most authoritative present-day grammar, the Duden grammar (cf. Duden grammar 2009:474). However, this rule does not apply for cases in which the second dependency is filled by a modal verb. In such clusters, modal verbs do not have past participle forms, but an infinitive. This is considered to be a substitution of the past participle, hence the name "Ersatzinfinitiv" ('substitutive infinitive'). This grammatical feature also alters the word order; here, the finite verb occurs in the initial position:

(2)   … *dass er das Spiel hat¹ sehen³ können²*
        … that he the match has[1–fin] watch[3–inf] can[2–inf]
        'that he has been able to watch the match'

The Duden grammar (2009:474) pronounces this a 'special rule' in present-day standard German, which is compulsory in cases were the finite verb is *haben*. If the finite verb is *werden*, both positions (initial and final) are possible (cf. Duden grammar 2009:474–475). In contrast to this variation of the finite verbs, the order of the infinite parts of the cluster is restricted to 3–2.

It is noteworthy, that a third possible pattern is not mentioned in the Duden grammar, namely 3–1–2. Here, the finite verb is put in between the infinite parts (so-called "Zwischenstellung"). According to Härd (1981:117), this variant gradually fell into disuse as from the seventeenth century and was virtually extinct in the nineteenth century. Although it is not codified in grammars of standard German,

the 3–1–2 order is considered to be a feature of Austrian standard German (cf. Härd 1981:178; Zeman 1988:78 and 2002:51–53; Askedal 1989:25; Bader & Schmid 2009:219).

So what findings do the corpus studies provide? Which serialization patterns occur in standard language usage? The major caveat which concerns the validity of our results is that all corpora used for this study except for the corpus *Variantengrammatik* are rather small, compared to today's standards in corpus-linguistics, and that, in general, three-verb clusters are a rather rare phenomenon, let alone variation in these clusters. This means that we have to deal with small numbers of variants, particularly in our historical corpora.

The *Variantengrammatik* corpus, however, with ca. 640 million words compiled from some 60 regional newspapers in the German-speaking countries, clearly shows pluriareal variation (also cf. Scherr & Niehaus 2013), cf. figure 2.
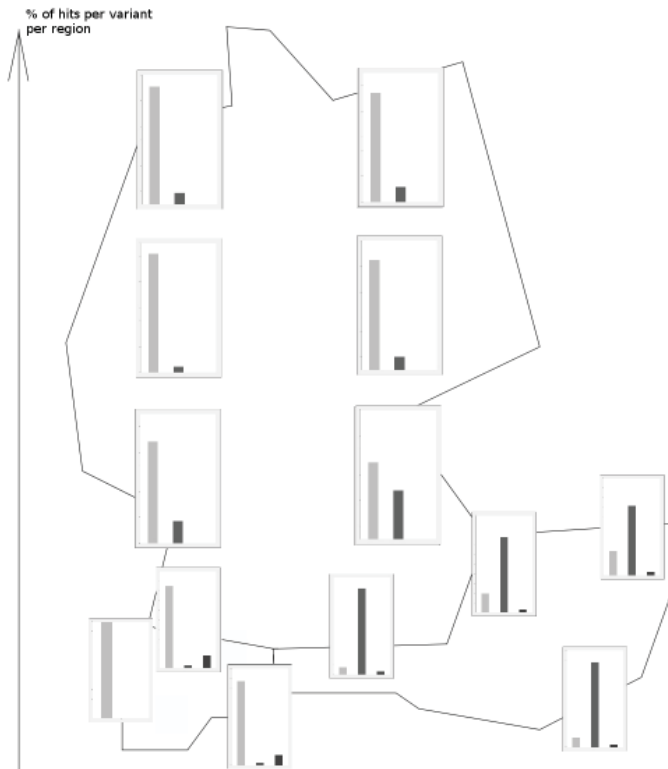

Fig. 2. Serialization in verb clusters in the corpus *Variantengrammatik* (main German-speaking countries) (light grey: 1-3-2; dark grey: 3-1-2; black: 3-2-1)

These proportions (and also a look at true positives) show that in this corpus all three variants are well documented. The 3–1–2 order is very noticeable in the south-eastern regions of the German-speaking language area. In Austria, it is even the dominant (but not absolute) variant. As leading grammar books such as the Duden grammar consider themselves *grammars of usage*, it is surprising that the 3–1–2 variant has not been marked as a standard variant yet, even more so, as it is also highly common in the adjoining areas of Liechtenstein, South Tyrol and the south-east of Germany (i.e. large parts of Bavaria).

Is this variation and this particular distribution a result of recent developments in standard German? As mentioned above, three-verb clusters are relatively rare, and we cannot expect high numbers of instances in the three smaller corpora. The *Kaiserreichkorpus* gives no hints as to regional variation in nineteenth century printed standard German, there is evidence for the 1–3–2 order only. This may support the monocentric tendencies after 1870, which von Polenz mentioned. But note that the corpus is very small, so that the lack of data for variants may simply be due to limitations of the corpus size.

For the view 'from below', the Pfeffer Corpus and the Emigrant letter corpus, render a different picture. In spite of the low numbers, it quickly becomes clear that the areal distribution of the three-verb cluster variants resembles the spread of the individual variants in present-day standard German. Moreover, there has been a continuity of usage in serialization patterns since the nineteenth century. In southern Germany and Austria the 3–1–2 order is the only variant that occurs in the Emigrant letter corpus (5 matches), while in northern Germany also the initial position of the finite verb (1–3–2) shows up and is highly dominant (> 70% within the variable). The distribution in the Pfeffer corpus is very similar, with the only difference that the 1–3–2 order seems to have become stronger in the southern areas. With all due caution, we may conclude then, firstly, that the areal distribution of the three variants in the *Variantengrammatik* corpus is matched by the pattern that we find in the non-standard corpora and, secondly, that there has been a stable pattern in non-standard German over the last 150 years.

From the results of the corpus study, it becomes evident that an explanation for the variation in verb clusters must be grammatically as well as historically and sociolinguistically founded. Grammatically speaking, prosodic-induced focusshifts (esp. focus on the lexical verb) might have an effect on word order in verbal clusters, according to

dialectal and diachronic studies (cf. Dubenion-Smith 2010, 2011; Sapp 2011:204–205). Sociolinguistically speaking, Bader & Schmid believe that the dominance of the 3–1–2 order in Austrian standard German could be caused by a more 'casual' handling of standard-norms in Austria (cf. Bader & Schmid 2009:219), but this seems disputable as perhaps also the opposite seems likely: the 3–1–2 order might be considered a 'national' norm by Austrian speakers (cf. Patocka 1997:281). This does not change the empirical fact that this word order variant is commonly used in standard context in other German-speaking countries as well and stands as a general feature of south-eastern standard German.

4.2 Grammatical variation II: discontinuous pronominal adverbs

Another grammatical construction which variationist studies have paid a lot of attention to in recent years – also for other Germanic languages such as Dutch, Frisian and Afrikaans[6] – is the pronominal adverb and its discontinuous variants. Example (3) presents a pronominal adverb in its continuous variant, examples (4) and (5) illustrate the two most common discontinuous forms[7] of the German pronominal adverb *damit* 'with this'.

(3) *damit habe ich nichts zu tun*
     this-(PRO)-with have I nothing to do
     'I have nothing to do with this'

(4) *da habe ich nichts mit zu tun*
     this-(PRO) have I nothing with to do

(5) *da habe ich nichts damit zu tun*
     this-(PRO) have I nothing this-(PRO)-with to do

In the discontinuous variant (4), the pronominal (usually *da-*, but also *wo-* and *hier-*) and the prepositional element of the adverb are separated in such a way that at least one other part of the sentence is moved in between. In most of such cases, the pro element is moved to the be-

---

[6] Cf. Fleischer (2002a, b) and Negele (2012) for an overview.
[7] Sometimes the first pronominal element is immediately followed by a 'full' pronominal adverb, resulting in a 'contact' position: *da damit habe ich nichts zu tun* 'this (PRO) this-(PRO)-with have I nothing to do'. In another variant, the pronominal variant is dropped: *__habe ich nichts mit zu tun* 'Ø (PRO) 'have I nothing with to do'.

ginning of the sentence.[8] A similar variant is a construction in which
the pro-element seems to be 'doubled' (5). Except for some dialects,
this 'split construction' occurs only when the preposition begins with
a consonant (e.g. *mit*, *zu*, *von*). In pronominal adverbs connecting *da-*
or *wo-* and a preposition with an initial vowel, an *r* is inserted (e.g.
darauf, *worüber*). Most grammars of standard German, such as the
Duden grammar (2009:581), do not consider discontinuous or other
variants of pronominal adverbs as standard. Moreover, according to
some authors, e.g. Eisenberg (2004:198), the 'double ᴘʀᴏ construction'
is restricted to cases in which the preposition begins with a vowel; as
a consequence the vowel in the second *da-* is usually dropped, e.g.
example (6).

(6)   da habe ich nicht d[a]ran gedacht
      this-(PRO) have I not this-(PRO)-*r*-of thought
      'I have not thought of this'

Fig. 3. Areal distribution of discontinuous pronominal adverbs in the corpus *Varian-
tengrammatik* (left column: split construction, right column: double pro construction)

---

8   The prepositional element, however, does not always move to the last position.
    Therefore, the term 'preposition stranding', which is sometimes used in gram-
    mars (cf. Eisenberg 2004:198) or the research literature (Fleischer 2002a, b), can be
    misleading.

Fig. 4. Areal distribution of split construction in 19th and 20th c. colloquial German (dark grey area; symbols indicate southernmost documented places of use in the respective century), adapted from Negele (2012:121)
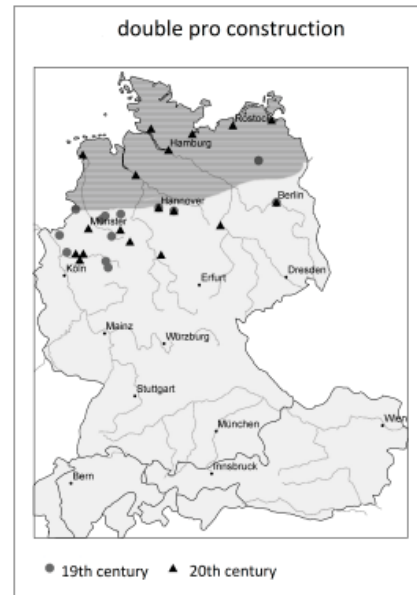
Fig. 5. Areal distribution of double pro constructions in 19th and 20th c. colloquial German (light grey area; symbols indicate northernmost documented places of use in the respective century), adapted from Negele (2012:122)

'Double PRO constructions' in which the preposition begins with a consonant, however, are also widespread. Negele (2012:242–44) argues that such 'double PRO constructions' as in example (5), as well as 'split construction', are standard German, and she gives evidence from standard German texts which hint at a clear areal distribution. 'Double PRO constructions' are employed in the north, 'split construction' in the south of the German speaking countries. In spite of the relatively few instances in the *Variantengrammatik* corpus, Negele's findings can to a certain extent be corroborated by the results of our corpus study. Figure 3 shows a preference for the 'split construction' in the north and west of Germany, whereas the 'double PRO constructions' appear mainly in the southern parts of the German language area. The (small) *Kaiserreichkorpus* gives no hints as to a regional variation in nineteenth century printed standard German. Not a single instance of a discontinuous construction could be found.

Again, it will be safe to assume that discontinuous constructions have not fallen from heaven into present-day German, nor will the

regional distribution be coincidental: In fact, Negele's data from her corpus studies on nineteenth and twentieth century colloquial German, mainly based on the emigrant letter and the Pfeffer corpus (cf. fig. 4 and 5, from Negele 2012:121–2, maps 21 and 22), confirm that the present-day north-south division in the preferred use of discontinuous pronominal constructions in standard German echoes the distribution in nineteenth as well as twentieth century colloquial German.

4.3 Lexical variation: *Sonnabend* and *Samstag* 'Saturday'

Our last case study looks at a prominent instance of lexical variation in standard German, the different names for the last day of the week. The largest dictionary concerned with regional lexical variation in standard German is the *Variantenwörterbuch* ('dictionary of [standard German] variants', Ammon, Bickel & Ebner et al. 2004). For the variable 'Saturday', it states that *Sonnabend* is the common variant in northern Germany (ibid.:653, 724), while *Samstag* is used in central and southern Germany, Austria and Switzerland, with the latter variant expanding more and more into the centre and to the north (ibid.:653).

Again, our data from the *Variantengrammatik* corpus draw a somewhat different picture (cf. Fig. 6). Although the *Variantenwörterbuch*'s statement (and codification) can be confirmed in general, the corpus study renders one notable difference. In central Germany the usage of *Sonnabend* is very much restricted to the middle-east. (There are a few instances of *Sonnabend* in newspapers from Luxemburg, central-west Germany and even southern Germany, but the respective absolute numbers are so low that they would not show up in the columns of our map, cf. fig. 6.)

This areal distribution in figure 6 is repeated in the other three corpora, the Pfeffer Corpus and the two historical corpora for (standard and non-standard) 19th century German. Thus, the overall picture can be interpreted as another case of 'variational continuity'. (This time, however, both variants were and are marked as standard German by the codices.)

The areal (and non-pluricentric) continuity probably reaches back to the Early Modern period. The structure of this variable might be explained by usage conventionalized in times when denomination was a strong factor in the division of regions, as the *Sonnabend*-area

is very similar to the post-1648 regions of Lutheran denomination (cf. Putzger 2006:107). Thus, one could speculate that the two variants originally represented social in-group markers, an effect which then could have been enhanced by restrictions for Lutheran and catholic printers during the counter-reformation, as testified for other variables (cf. Rössler 2005:364–367).
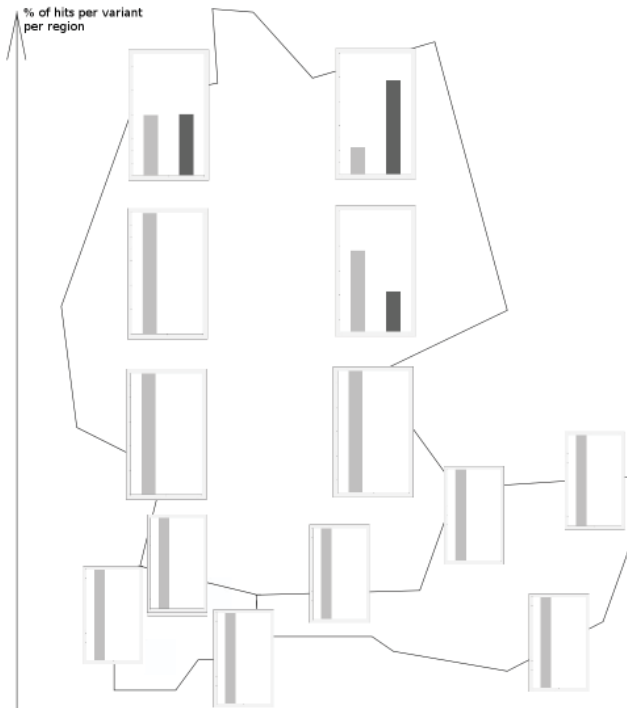

Fig. 6. Lexical variation for 'Saturday' in the corpus *Variantengrammatik* (light: *Samstag;* dark: *Sonnabend*)

5   Conclusion

In this paper, we argued for a consideration of standard as well as non-standard language data from present-day and historical corpora in the study of language standardization. With reference to data from German, we adopted the pluriareal rather than the pluricentric model of 'big' standard languages, as the pluriareal concept seems to depict the factual language variation more adequately and realistically.

As to our research questions, we can now answer them as follows:

- How much variation did printed standard German allow in the nineteenth century? –
  The *Kaiserreich* corpus does not indicate any relevant areal variation. The almost complete lack of variation might, however, be due to the small size of the nineteenth century standard German corpus.

- Does printed standard German change from 'monocentricity' to 'pluricentricity' (cf. von Polenz 1989:15)? –
  For the period from 1870 to the First World War, von Polenz' notion of 'monocentricity', which corresponds to a low degree of areal variation, can be supported – at least on the basis of the small corpus that we have so far. Present-day standard German, by contrast, allows for more areal variation. Corresponding findings reported in the research literature (cf. Schmidlin 2011:65; Klein 2013:26) can be supported by the results of our corpus studies. The rise in variation could be attributed to the increasing informality in newspaper style which has become noticeable particularly after the mid-twentieth century (cf. Betz 2006:183–185), as informality is closely associated with the degree of areal variation in language. However, as we can find patterns in the regional distribution of linguistic variants which in many cases are not only determined by political borders, a change 'from monocentricity to pluri*areality*' would describe the linguistic situation more adequately.

- Is (and was) the variation in non-standard German similar to or rather different from printed standard German? Which tendencies in standard and non-standard German are discernible in the last 150 years? –
  As for the last 150 years, variation in non-standard German and printed standard German can differ substantially in a historical perspective, as our case studies indicate. In two of our three case studies (4.1 and 4.2), the variational patterns seem to be more similar to each other in the present, not least due to a tendency towards informal styles in standard German (as we just argued). A major finding of the case studies on the serialisation of verb-clusters and the discontinuous pronominal adverbs is that pluriareal patterns which have emerged in today's standard German can be traced back to present-day as well as earlier non-standard usage. Further corpus studies are needed, of course, to establish whether these are rather singular cases or whether they represent a more general pattern in the recent standardization history.

We hope that our case studies could illustrate one important methodological point and its consequences for research. Studies on standardization and of standard as well as non-standard variation require more and bigger corpora, particularly from the 'age of standardization', i.e. the nineteenth century, onwards. In addition, for big languages such as German we also need more regionally balanced text corpora. As the two case studies on grammatical phenomena clearly demonstrated, one major problem can be a 'descriptive gap' in standardization studies as well as in grammaticography, when variants are overlooked or ignored because of a lack of adequate corpora. This, in turn, limits the validity of such studies, particularly with respect to an adequate description and explanation of linguistic phenomena in terms of the 'principle of viability' (cf. section 2.1). The bigger present-day as well as historic corpora are, and the more they reflect the areal division of a speech community, the better they can help to generate formulations of empirically based 'diachronic principles of grammar'. The comparatively large *Variantengrammatik* corpus gives us a slight hint at what treasures may be hidden in big historical corpora – even if they turn out to be small in terms of absolute numbers.

References

AdA = Elspaß, Stephan & Robert Möller. 2003–2013. *Atlas zur deutschen Alltagssprache*. http://www.atlas-alltagssprache.de/.

Ágel, Vilmos. 2001. Gegenwartsgrammatik und Sprachgeschichte. Methodologische Überlegungen am Beispiel der Serialisierung im Verbalkomplex. *Zeitschrift für Germanistische Linguistik* 29:293–318.

Ágel, Vilmos. 2003. Prinzipien der Grammatik. In: Lobenstein-Reichmann & Oskar Reichmann (eds.), *Neue historische Grammatiken. Zum Stand der Grammatikschreibung historischer Sprachstufen des Deutschen und anderer Sprachen*, pp. 1–46. Tübingen: Niemeyer.

Ammon, Ulrich, Hans Bickel, Jakob Ebner, et al. 2004. *Variantenwörterbuch des Deutschen. Die deutsche Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol*. Berlin/New York: de Gruyter.

Askedal, John Ole. 1986. Zur vergleichenden Stellungsfelderanalyse von Verbalsätzen und nicht-verbalen Satzgliedern im Deutschen. *Deutsch als Fremdsprache* 23:269–273 and 342–348.

Bader, Markus & Tanja Schmid. 2009. Verb clusters in colloquial German. *Journal of comparative Germanic Linguistics* 12:175–228.

Bech, Gunnar. 1955/57[1983]. *Studien über das deutsche Verbum Infinitum*. 2nd vol. Tübingen: Niemeyer.

Betz, Ruth. 2006. *Gesprochensprachliche Elemente in deutschen Zeitungen.* Radolfzell: Verlag für Gesprächsforschung. http://www.verlag-gespraechs-forschung.de/2006/betz.htm [accessed 12.07.2013].

Blommaert, Jan & Jef Verschueren. 1998. The role of language in European nationalist ideologies. In: Schieffelin et al. (eds.), pp. 189–210.

Clyne, Michael. 1995. *The German language in a changing Europe*. Cambridge: Cambridge University Press.

Dubenion-Smith, Shannon A. 2010. Verbal Complex Phenomena in West Central German: Empirical domain and multi-causal account. *Journal of Germanic Linguistics* 22,2:99–191.

Dubenion-Smith, Shannon A. 2011. Der Verbalkomplex im Schlesischen. *Zeitschrift für Dialektologie und Linguistik* 78:281–320.

Duden grammar (2009) = *Duden. Die Grammatik. Unentbehrlich für richtiges Deutsch. Herausgegeben von der Dudenredaktion*. 8th, rev. ed. Mannheim/Zürich: Dudenverlag.

Dürscheid, Christa, Stephan Elspaß & Arne Ziegler. 2011. Grammatische Variabilität im Gebrauchsstandard: das Projekt 'Variantengrammatik des Standarddeutschen'. In: Konopka et al. (eds.), pp. 123–140.

Eisenberg, Peter. 2004. *Grundriss der deutschen Grammatik*. 2nd, rev. ed. Vol. 2: Der Satz. Stuttgart: Metzler.

Elspaß, Stephan. 2005. *Sprachgeschichte von unten. Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert*. Tübingen: Niemeyer.

Elspaß, Stephan. 2012. The Use of Private Letters and Diaries in Sociolinguistic Investigation. In: Hernández-Campoy, Juan Manuel & Juan Camilo Conde-Silvestre (eds.), *The Handbook of Historical Sociolinguistics*, pp. 156–169 Oxford: Wiley-Blackwell.

Elspaß, Stephan (in print). Private letters as a source for an alternative history of Late Modern German. In: Auer, Anita, Daniel Schreier & Richard J. Watts (eds.), *Letter Writing and Language Change*. Cambridge: Cambridge University Press.

Elspaß, Stephan, Nils Langer, Joachim Scharloth & Wim Vandenbussche (eds.). 2007. *Germanic language histories 'from below' (1700–2000)*. Berlin/New York: de Gruyter.

Ernst, Oliver & Stephan Elspaß. 2011. Althochdeutsche Glossen als Quellen einer Sprachgeschichte ‚von unten'. *NOWELE* 62/63:249–283.

Fleischer, Jürg. 2002a. *Die Syntax von Pronominaladverbien in den Dialekten des Deutschen. Eine Untersuchung zu Preposition Stranding und verwandten Phänomenen*. Stuttgart: Steiner.

Fleischer, Jürg. 2002b. Preposition Stranding in German Dialects. In: Barbiers, Sjef, Leonie Cornips & Susanne van der Kleij (eds.), *Syntactic Microvariation*. (= Meertens Institute Electronic Publications in Linguistics, 2), pp. 116–151. Amsterdam: Meertens Institute. http://www.meertens.knaw.nl/books/synmic/pdf/fleischer.pdf [accessed 12.07.2013].

Graser, Helmut & B. Ann Tlusty. 2012. Sixteenth Century Street Songs and

Language History 'from below'. In: Langer, Nils, Steffan Davies & Wim Vandenbussche (eds.): *Language and History, Linguistics and Historiography. Interdisciplinary Approaches*, pp. 362–388. Oxford [etc.]: Lang.

Hagemann, Jörg, Wolf Peter Klein & Sven Staffeldt (eds.). 2013. *Pragmatischer Standard*. Tübingen: Stauffenburg.

Härd, John Evert. 1981. *Studien zur Struktur mehrgliedriger deutscher Nebensatzprädikate. Diachronie und Synchronie*. Göteborg: Acta Universitatis Gothoburgensis.

Klein, Wolf Peter. 2013. Warum brauchen wir einen klaren Begriff von Standardsprachlichkeit und wie könnte er gefasst werden? In: Hagemann, Jörg et al. (eds.), pp. 15–33.

Konopka, Marek, Jacqueline Kubczak, Christian Mair, František Štícha & Ulrich H. Waßner (eds.). 2011. *Grammar and Corpora 2009. Third International Conference, Mannheim 22.–24.09.2009*. Tübingen: Narr.

Lippi-Green, Rosina. 2012. *English with an Accent. Language, ideology, and discrimination in the United States*. 2nd edition. London/New York: Routledge.

Milroy, James & Lesley Milroy. 1985. *Authority in Language. Investigating Language Prescription and Standardisation*. London/New York: Routledge and Kegan Paul.

Negele, Michaela. 2012. *Varianten der Pronominaladverbien im Neuhochdeutschen. Grammatische und soziolinguistische Untersuchungen*. Berlin/New York: de Gruyter.

Nobels, Judith. 2013. *(Extra)Ordinary Letters. A view from below on seventeenth-century Dutch*. Utrecht: LOT.

Oesterreicher, Wulf. 1997. Types of orality in text. In: Bakker, Egbert & Ahuvia Kahane (eds.), *Written voices, spoken signs*, pp. 190–214. Cambridge, MA/London: Harvard University Press.

Patocka, Franz. 1997. *Satzgliedstellung in den bairischen Dialekten Österreichs*. Frankfurt a. Main [u.a.]: Peter Lang.

Paul, Hermann. 2007. *Mittelhochdeutsche Grammatik*. Rev. by Thomas Klein et al. Tübingen: Niemeyer.

Pfeffer, J. Alan & Walter F. W. Lohnes. 1984. *Grunddeutsch. Texte zur gesprochenen deutschen Gegenwartssprache*. Tübingen: Niemeyer.

von Polenz, Peter. 1989. Das 19. Jahrhundert als sprachgeschichtliches Periodisierungsproblem. In: Cherubim, Dieter & Klaus J. Mattheier (eds.), *Voraussetzungen und Grundlagen der Gegenwartssprache*, pp. 11–30. Berlin/New York: de Gruyter.

von Polenz, Peter. 1999. Deutsch als plurinationale Sprache im postnationalistischen Zeitalter. In: Gardt, Andrea, UlrikeHaß-Zumkehr & Thorsten Roelcke (eds.): *Sprachgeschichte als Kulturgeschichte*, pp. 115–132. Berlin / New York: de Gruyter.

Putzger. 2006. *Historischer Weltatlas. Kartenausgabe*. Berlin: Cornelsen.

Reichmann, Oskar & Klaus-Peter Wegera (eds.). 1993. *Frühneuhochdeutsche Grammatik*. Von Robert Peter Ebert, Oskar Reichmann, Hans-Joachim Solm und Klaus-Peter Wegera. Tübingen: Niemeyer.

Reiffenstein, Ingo. 2001. Das Problem der nationalen Varietäten. Rezensionsaufsatz zu Ulrich Ammon: Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten. Berlin/New York 1995. *Zeitschrift für deutsche Philologie* 120:78–89.

Rössler, Paul. 2005. *Schreibvariation – Sprachregion – Konfession. Graphematik und Morphologie in österreichischen und bayerischen Drucken vom 16. bis ins 18. Jahrhundert*. Frankfurt a. Main [u.a.]: Peter Lang.

Sapp, Christopher D. 2011. *The Verbal Complex in Subordinate Clauses from Medieval to Modern German*. Amsterdam/Philadelphia: Johns Benjamins.

Scheible, Silke, Richard Jason Whitt, Martin Durrell & Paul Bennett. 2011. Investigating diachronic grammatical variation in Early Modern German. Evidence from the GerManC corpus. In: Konopka et al. (2011), pp. 539–548.

Scherr, Elisabeth & Konstantin Niehaus. 2013. … weil man den Gebrauchsstandard erheben wird wollen. Variabilität und funktionale Äquivalenz in der Standardsyntax am Beispiel der 'Zwischenstellung' in Verbalkomplexen. In: Hagemann, Jörg et al. (eds.), pp. 75–84.

Scheuringer, Hermann. 1997. Sprachvarietäten in Österreich. In: Stickel, Gerhard (ed.), *Varietäten des Deutschen. Regional- und Umgangssprachen*, pp. 332–345. Berlin/New York: de Gruyter.

Schieffelin, Bambi, Kathryn Woolard & Paul Kroskrity (eds.). 1998. *Language Ideologies: Practice and theory*. New York: Oxford University Press.

Schmidlin, Regula. 2011. *Die Vielfalt des Deutschen: Standard und Variation. Gebrauch, Einschätzung und Kodifizierung einer plurizentrischen Sprache*. Berlin/New York: de Gruyter.

Schulte, Michael. 2009. Neue Überlegungen zum Aufkommen des Jüngeren Fuþarks. Ein Beitrag zur Schriftgeschichte „von unten". *Beiträge zur Geschichte der deutschen Sprache und Literatur* 131:229–251.

Van de Velde, Marc. 1981. Zur Reihenfolge von verbalen Elementen. In: Kohrt, Manfred & Jürgen Lenerz (eds.), *Sprache: Formen und Strukturen. Akten des 15. Linguistischen Kolloquiums Münster 1980* (Vol. 1), pp. 223–235. Tübingen: Niemeyer.

Wolf, Norbert Richard. 1994. Österreichisches zum österreichischen Deutsch. Aus Anlaß des Erscheinens von: Wolfgang Pollak: Was halten die Österreicher von ihrem Deutsch? Eine sprachpolitische und soziosemiotische Analyse der sprachlichen Identität der Österreicher. *Zeitschrift für Dialektologie und Linguistik* 61:66–76.

Zeman, Jaromír. 1988. Zu Stellungsvarianten des Verbs im Österreichischen: „daß nur ich es hören habe können". *Brünner Beiträge zur Germanistik und Nordistik* 6:71–82.

Zeman, Jaromír. 2002. *Die deutsche Wortstellung*. Wien: Edition Praesens.

Keywords

language history, variationist linguistics, grammar, lexical variation, grammatical variation (verbal clusters, discontinuous pronominal adverbs)

Lykilorð

málsaga, tilbrigðarannsóknir, málfræði, tilbrigði í orðaforða, málfræðileg tilbrigði

Útdráttur

Lengi vel voru hugtök eins og "einsleitni" og "staðalmálshugmyndafræði" ríkjandi í tengslum við málstöðlun og það leiddi til þess að gögnin sem rannsóknir voru byggðar á einskorðuðust gjarnan við formlegt mál og bókmenntatexta. Á síðustu árum hafa tvær hugmyndir tekið að hreyfa við hefðbundnum viðhorfum til málstöðlunar. Í fyrsta lagi hefur nálgun sem kennd er við málsögu "að neðan" dregið til sín athygli sem nýtt sjónarhorn, ekki bara gagnvart þýsku heldur einnig öðrum germönskum málum (cf. Elspaß, Langer, Scharloth & Vandenbussche 2007). Í rannsóknum innan þessa fræðilega ramma er áhersla lögð á söguleg gögn sem endurspegla talmálsleg einkenni, einkum texta sem skrifaðir voru af fólki úr lægri lögum þjóðfélagsins, í þeim tilgangi að leiða í ljós breytingar og/eða samhengi í málsögunni sem áður höfðu ekki komið fram. Í öðru lagi hefur líkan sem gerir ráð fyrir mörgum svæðum (fremur en einum kjarna) fengið útbreiðslu, einkum með tilliti til "stórra" mála eins og þýsku. Í greininni er gerð grein fyrir skipulagi málheilda sem tekur tillit til beggja þessara hugmynda við athuganir á stöðlunarferlinu og leitast er við að sýna að rannsóknir sem byggjast á stærri og jafnvægari málheildum geta styrkt grundvöll athugana á málstöðlun.

Stephan Elspaß / Konstantin Niehaus
Universität Salzburg
Fachbereich Germanistik
Erzabt-Klotz-Straße 1
A-5020 Salzburg
stephan.elspass@sbg.ac.at / konstantin.niehaus@sbg.ac.at