Measuring (dia)lectal coherence across time: Change over the lifespan and in the community[1]

Karen V. Beaman and Konstantin Sering

1. Introduction

Fifty years ago, Weinreich, Labov, and Herzog (1968:188) observed that "idiolects do not provide the basis for self-contained or internally consistent grammars;" rather, regularity and coherence, constrained by social factors, is found in the composite grammar of the speech community. Recently, increasing focus has been placed on the role of the individual in ongoing language change and whether individual lifespan trajectories can "speed up" or "slow down" community change (Wagner and Buchstaller 2018). This conundrum precipitates the question as to how well the grammars of individuals and the grammars of a community may cohere or "behave in parallel" (Guy and Hinskens 2016) and what this portends for theories of linguistic change (see also Guy 1980; Walker and Meyerhoff 2013).

The central premise underlying the concept of sociolectal coherence is that linguistic features in a speech community are consistent and act in unison in indexing different lects – dialects, regiolects, sociolects – as well as styles, registers, stances, and so on (Guy 2013; Guy and Hinskens 2016). Greater lectal coherence (i.e., clusters of linguistic features and their relationships) implies that changes in one variant "trigger" changes in another variant such that multiple related variables co-occur within a unified variety. But what constitutes a unified variety? How coherent do variables have to be to be considered a single variety? The aim of this chapter is to evaluate different measures of sociolectal coherence with the ultimate goal of testing a common assumption that sociolects with greater levels of coherence are more resistant to variation and change in coherence across time, while those with lower levels of coherence are more vulnerable, paralleling Milroy's (1987) findings that closed, tightly-knit social networks (i.e., clusters of individuals and their relationships) are more impervious to change while open and diffuse social networks are more accepting of innovations. This chapter starts with a brief review of some of the pertinent research on sociolectal coherence (Section 2), then describes the corpus and the variables under investigation (Section 3), followed by an analysis and discussion of the results (Section 4). In closing, we discuss the important role of sociolectal coherence in a theory of language change (Section 5).

2. Theoretical Background

Guy and Hinskens (2016) maintain that Weinreich, Labov, and Herzog's (1968) concept of ORDERLY HETEROGENEITY implies that "speech communities are sociolinguistically coherent .... [meaning that] the community should collectively behave in parallel: variants (or rates of use of variants) that index a given style, status, or a social characteristic should co-occur" (Guy and Hinskens 2016:2). The concept of sociolinguistic

[1] The authors wish to thanks the following colleagues for their review and constructive feedback on earlier drafts of this research: Lars Bülow, Jenny Cheshire, Gregory Guy, Fabian Tomaschek, and James Walker. Any deficiencies remaining are, of course, our own.

coherence implies that speech communities can be distinguished by their sociolects, which Guy (2013:64) characterizes as a “cluster of variables” that identify a specific social group, e.g., dialect, ethnolect, class-based variety, style, register. “If sociolects are indeed socially and cognitively coherent varieties, we should expect some degree of correlation among the different variables [or variants] present in a community” (Guy 2013:64). This supposition has given rise to a contentious debate in the field regarding the role of community coherence versus that of individual agency (e.g., Bucholtz and Hall 2005; Eckert 2008, 2012; Sharma 2018), which Guy (2014) refers to as “bricks and bricolage” (see Sharma, this volume, for a full discussion of this topic). He maintains that individual linguistic variants can be “mortared together” to construct a sound edifice (“bricks”) or they can be “elaborative additions” to portray a particular style, stance, or identity (“bricolage”) which will be “relatively fluid, and not necessarily have any more coherence and permanence than one’s specific ensemble of clothing worn on a given day” (Guy 2014:2).

Studies have employed various methods in analysing whether linguistic features covary within specific language varieties (see Walker et al., this volume, for a discussion of common approaches). One of the earliest such studies was by Horvath and Sankoff (1987), who investigated variation in four vowels in Sydney, Australia using principal components analysis (PCA). More recently, Meyerhoff and Klaere (2017) used constrained correspondence analysis (CCA), incorporating researcher designated constraints (e.g., “village membership”) to guide the data reduction and aggregation algorithms. Guy (2013), Oushiro and Guy (2015), Newlin-Lukowicz (2016), and Tamminga (2019) used multivariate analyses to obtain factor weights (i.e., speakers’ tendency to use an innovative or nonstandard variant derived through random effects or residuals) and performed cross-correlations to determine whether speakers with different social characteristics cohere in their patterns of variable usage. In his investigation of four variables in Brazilian Portuguese, Guy (2013) discovered some evidence for sociolectal coherence, particularly in women’s tendency to use higher status variables; however, men showed no such coherence, raising the question whether nonstandard phonology is more indexical of masculinity, an effect which overpowers linguistic coherence. Oushiro and Guy (2015) found little covariation in their investigation of six features of Brazilian Portuguese and concluded that coherence might be better explained through “structural similarities” and phonic salience (Naro 1981) than through social groupings. Most recently, in their review of six studies from Guy and Hinskens’ (2016) *Lingua* issue dedicated to the topic of coherence, Woo, Gadanidis, and Nagy (submitted) observed that little over half of the linguistic variables investigated demonstrated coherence, while their own findings on heritage Cantonese spoken in Toronto showed even less covariation: only six out of 21 variable pairs were significantly correlated ($p < .05$) and only two under the more rigorous Spearman test.

The dilemma over why some studies show greater levels of coherence than others is highly influenced by the diverse methods employed and the wide assortment of variables (both dependent and independent) investigated. As Cheshire (1999:65) points out, “in variationist analyses we are limited in what we discover by what we set out to look for.” Thus, the goal of the current study is to explore different methods for quantifying and measuring coherence and then to apply the “best” measure to assess how coherence changes across real- and apparent-time and what this implies for language change.

3. Swabian Corpus

This study analyses a corpus of spontaneously spoken Swabian (Beaman 2020), an upper German dialect belonging to the Alemannic family spoken by approximately 800,000 people or one percent of the German population. The corpus comprises two study types: a real-time panel study of 20 speakers, first interviewed in 1982 and reinterviewed 35-years later in 2017, and an apparent-time trend study of 40 speakers in 2017 who were selected as "social twins", matched for age, sex, education and locality with the panel speakers: the twin speakers exhibit the same mean age (44.8 for the panel study and 45.7 for the trend study) (Blondeau 2001:469) and reflect a similar distribution across speaker gender (100% match), locality (93% match), age group (73%), and education (based on *Abitur* achievement 'German college preparatory exam' (73%), producing an 84% overall match (see Table 1 for a breakdown on the sociodemographics of the sample).

PANEL STUDY

Stuttgart

Age Group	No Abitur		Abitur		Total
	M	W	M	W	
Older	0	1	4	1	6
Younger	0	1	0	0	1
TOTAL	**0**	**2**	**4**	**1**	**7**

Schwäbisch Gmünd

Age Group	No Abitur		Abitur		Total
	M	W	M	W	
Older	1	2	0	0	3
Younger	0	1	6	3	10
TOTAL	**1**	**3**	**6**	**3**	**13**

TWIN STUDY

Stuttgart

Age Group	No Abitur		Abitur		Total
	M	W	M	W	
Older	**0**	**1**	**0**	**0**	**1**
Younger	**3**	**3**	**5**	**2**	**13**
TOTAL	**3**	**4**	**5**	**2**	**14**

Schwäbisch Gmünd

Age Group	No Abitur		Abitur		Total
	M	W	M	W	
Older	1	2	1	1	5
Younger	5	4	7	5	21
TOTAL	**6**	**6**	**8**	**6**	**26**

Table 1. Swabian Corpus.

3.1 Swabian Communities

Labov claims, "one cannot understand the development of language change apart from the social life of the community in which it occurs" (Labov 1963:275) and, over the last 50 years, this phenomenon has been substantiated by countless sociolinguistic studies. This study investigates two typical Swabian communities, the large urban metropolis of Stuttgart and its neighbouring suburbs and the mid-sized town of Schwäbisch Gmünd with its many surrounding small, rural and semi-rural villages, providing the opportunity to examine variation and change in both an open, diffuse, urban environment and in a closed, denser, semi-rural setting.

Stuttgart is considered to be the ideological norm centre for Swabian (Ruoff 1997:145). It is a large urban hub with over one million inhabitants and is home to many well-known global firms, such as Daimler, Porsche, Bosch, and Siemens. Stuttgart has one of the most diverse populations in Germany, with almost twice as many "foreigners" (individuals with at least one foreign-born parent) as in Germany overall (Auer 2020).[2] The city also benefits from substantial internal migration, with many inhabitants from the north and east who have moved to the region for employment.

[2] *Statistisches Amt, Landeshauptstadt Stuttgart,* https://statistik.stuttgart.de/statistiken/tabellen/7392/jb7392.php

In contrast, Schwäbisch Gmünd, located 100 kilometers east of Stuttgart, has 60,000 inhabitants and is a typical mid-sized German town surrounded by small rural villages with 77% of the land dedicated to woodland and agriculture. Situated in a valley of the Rems between the edge of the *Welzheimer Wald* (part of the Swabian-Franconian Forest) in the north and the foothills of the eastern Swabian Alb in the south, the town derives its name from the word *Gemünde* 'confluence of many streams'.[3]

3.2 Wir können alles außer Hochdeutsch 'We can do everything, except standard German'[4]

A strong Swabian identity, manifested in deeply ingrained linguistic attitudes in the minds of Swabian speakers, wields a formidable influence on the dialect-standard language situation. The following citations from some of the speakers[5] in the study show that the Swabian language plays a complex role in the minds of its speakers, such as cultural identification, pride, shame, and longing for home and the homeland.

(1) *Willard (2017)*

i hätt niemâls e Frau -eh- geheiratet	*'I could never have married a woman*
die ned Schwäbisch schwätzt	*who didn't speak Swabian*
die kulturellen Unterschiede sind	*the cultural differences are*
zu elementar	*too elementary*
du kannsch mit Dialekt einfach	*with the dialect you can easily*
ganz andere Welten erschließen	*unlock completely new worlds'*

(2) *Fabian (2017)*

also gewisserweise isch mã da scho	*'so in a certain way you're really*
e bissle Stolz darauf ã	*kinda proud [to speak Swabian]*
also des isch scho ganz cool	*so it's already pretty cool*
i glaub die Schwabe sind	*I think Swabians are*
au relativ gut angesehen in Deutschland	*also seen relatively good in Germany*
und so des isch ja e fleißiges Volk ja.	*and they are hard-working people yeah'*

(3) *Helmut (2017)*

meine Kinder schämen sich sogar	'my children are actually ashamed
heutzutage Schwäbisch …	of Swabian these days …
auf der einen Seite hast du	on one side you have
dieses Gefühl in der Öffentlichkeit	this feeling in public
du darfst nicht Mundart sprechen	you shouldn't speak dialect
weil du gleich dann	because then you'll immediately
nicht ernst genommen wirst	not be taken seriously
auf der anderen Seite halt man merkt	on the other side like you notice
dass da einfach eine Sehnsucht danach ist	that there is simply a longing for it'

Pursuing the concept of Swabian identity both quantitatively and qualitatively, Svenstrup (2019) carried out a traditional matched-guise language attitude study with 235 adolescents across five locations in northern Swabia. His findings show that more academically-oriented individuals prefer standard German and this effect is stronger for Stuttgart over the other localities, suggesting that language attitudes are "rooted in conceptions and values to do with rurality/urbanity and education" (*Ibid*:161) (cf. Gal 1978). Svenstrup's

---

[3] Drawn from various official government websites: http://www.schwaebisch-gmuend.de/, http://www.iggingen.de/. Viewed on 22-jan-2020.

[4] This quote refers to a campaign that the Baden-Württemberg government ran in 1999 (https://www.bw-jetzt.de/) with the goal of elevating the opinion of Swabian throughout Germany.

[5] All names have been anonymized to protect the privacy of the speakers.

study also signals a difference between the conscious and unconscious attitudes of the speakers, in which the conscious attitudes demonstrate a preference for Swabian, while the unconscious feelings reveal a preference for standard German (*Ibid*:162-163). Standard German is seen as a marker of success, reinforced in the schools where teachers act as "gatekeepers" of the norm (*Ibid*:164).

3.3 Data Collection

The data were collected following Labovian-style sociolinguistic interviews (Labov 1984), conducted by native Swabian speakers with the primary investigator (Beaman) in the role of a friend-of-friend (Milroy and Milroy 1985). The interviewers asked questions about the speaker's childhood, favorite games played, hobbies and interests, as well as questions about the Swabian culture and language and the speaker's participation in local activities and festivals. Interviews lasted approximately an hour, providing a total of 43 hours of recorded speech.

All recordings were transcribed orthographically in ELAN by native German speakers based on the transcribers' perception of a binary choice[6] between the dialect variant and the standard German variant, following the orthographic conventions established for the project. All transcripts were cross-validated at least four times: first, by a different student transcriber to the one who did the original transcription and then by the principal investigator, who made three passes through every transcript to ensure standards were followed, to neutralize any transcriber bias, and to ensure all tokens were correctly annotated. To support the manipulation of large quantities of data, an automated ELAN-to-R (E2R) annotation process was created, based on a bespoke Swabian-German Lexicon (SGL) (see Beaman 2020), which takes ELAN transcripts as input and generates extract files as output which then serve as input to R for data aggregation and statistical analysis (R Core Team 2014).

3.4 Linguistic Variables

In this analysis of coherence, ten phonological and ten morphosyntactic variables were selected as the dependent variables based on their frequency and commonality in depicting the Swabian dialect. A short description of each variable, along with its abbreviation, follows.

1. AIS1 – Shifting of (ai) diphthong of MHG /iː/ origin; e.g., standard German *Teig* [taɪk] 'dough' is realized as *Dêig* [dəɪg] in Swabian.
2. AIS2 – Shifting of (ai) diphthong of MHG /ei/ origin; e.g., standard German *klein* [klaɪn] 'small' is realized as *glôi* [glɔɪ] in Swabian.
3. ANN – Nasalisation of /a/ before /n/; e.g., standard German *man kann* [man kan] 'one can' is realized as *mã kã* [mã kã] in Swabian.
4. FRV1 – Unrounding and lengthening of the standard German front rounded vowel /ø/; e.g.,

[6] We follow traditional sociolinguistic practice in choosing a binary categorisation. "While collapsing detailed transcriptions into categories ... may be undesirable in the sense that one loses the phonetic resolution of the original transcriptions, it should be remembered that the term 'variant' is only meaningful if we choose to impose categories onto what is, after all, a phonetic continuum. As long as this is carried out in a careful, principled, and reproducible way, the approach serves the sociolinguist's purposes well" (Watt 2000:97).

möglich [møːklɪç] ‘possible’ is realized as *meeglich* [mɛːglɪç] in Swabian.

5. FRV2 – Unrounding of the diphthong /ɔɪ/; standard German *Freund* [fʁɔɪnt] ‘friend’ is realized as *Fraind* [fʁaɪnd] in Swabian.
6. FRV3 – Unrounding of the front vowel /y/; e.g., standard German *Küche* [kyːçə] ‘kitchen’ and *Gmünd* [gmyːnt] are realized as *Kiiche* [kɪːçə] and *Gmiind* [gmɪːnd] in Swabian.
7. FRV4 – Diphthongisation of /u/ to /uə/; e.g., standard German *muss* [mus] ‘must’ and *gut* [gut] ‘good’ are realized as [muəs] and [guəd] in Swabian.
8. LEO – Lowering of long /e:/; e.g., standard German *lesen* [leːzn] ‘read’ and *Lehrer* [leːʁɐ] ‘teacher’ are realized as *lääs* [læːs] and *Läährer* [læːʁɐ] in Swabian and other German varieties.
9. SFV – Stop-fricative variation; e.g., standard German *richtig* [ʁɪçtɪç] ‘correct’ is realized as *richtig* [ʁɪçtɪk] in Swabian and southern German varieties.
10. STPA – Palatalisation of /-st/ to /-ʃt/ in syllable-coda position; e.g., standard German *machst* [mɑxst] ‘you do/make’ are realized as *machscht* [maxʃt] in Alemannic.
11. DAS – Definite neuter article *des*; e.g., standard German *das* [das] ‘the’ is realized as *des* [dɛs] throughout southern Germany.
12. EDP – Present tense plural verb inflexion *-ed*; e.g., standard German *sie finden* [ziː fɪndn] ‘they find’ is realized as *sie finded* [siː fɪndəd] in Swabian.
13. IRV1 – Irregular verb *gehen*; e.g., standard German *ich gehe* [ɪç geːə] ‘I’m going’ is realized as *i gang* [ɪ gaŋ] in Swabian.
14. IRV2 – Irregular verb *stehen*; e.g., standard German *ich stehe* [ɪç ʃteːə] ‘I’m standing’ and *ich verstehe* [ɪç fɛɐˈʃteːə] ‘I understand’ are realized as *i stand* [ɪ ʃtaŋ] and *i verstand* [ɪ fɛɐˈʃtaŋ] in Swabian.
15. IRV3 – Irregular verb *hen*; e.g., standard German *ich habe* [ɪç haːbə] ‘I have’ und *gehabt* [gəhaːpt] ‘had’ is realized as *ghet* [ghɛːt] or *khet* [khɛːt] in Swabian.
16. NEG – Negative marker *net/nette*; e.g., standard German *nicht* [nɪçt] ‘not’ is realized as *et/net* ɛt/nɛt] or *edde/nedde* [ɛtə/nɛtə] in various localities in Swabian.
17. PVB – The periphrastic subjunctive with the auxiliary *tun*; e.g., standard German *es würde beeinflussen* ‘it would influence’ is realized as *es dääd beeinflusse* in Swabian.
18. SAF1 – Diminutive suffix *-le*; e.g., standard German suffix *-chen* (or the older suffix *-lein*), *bisschen* [bɪsçən] ‘a little’ and *Mädchen* [mɛːtçən] ‘little girl’ is realized as *bissle* [bɪslə] and *Mädle* [mɛːdlə] in Swabian.
19. SAF3 – Use of prefix *nââ* [nɔ] instead of *hin* [hɪn], which translates to ‘in’, ‘to’, ‘at’, ‘away’, ‘about’, ‘down’, ‘there’; e.g., standard German *hinkriegt* [hɪnkʁiːgt] is realized as *nââkriegt* [nɔkʁiːg] in Swabian.
20. SAF5 – Dropping of the part participle prefix *ge-;* e.g., standard German *haben gebaut* [haːbn gəˈbaʊ̯t] ‘have built’ or ‘built’ is realized as *hen baut* [hɛn baʊ̯t] in Swabian.

For these 20 variables, a total of 126,250 tokens were extracted covering both the panel and twin study, 86,110 phonological and 40,140 morphosyntactic; 25,030 were from 1982 (20 panel interviews) and 101,220 from 2017 (60 panel and twin study interviews), averaging 1,251 tokens per speaker in 1982 and 1,687 tokens per speaker in 2017 (see Table 2 for detail token counts per variable).

Phonological Variables & Tokens	Panel 1982	Panel 2017	Twin 2017	TOTAL
AIS1 - MHG /ī/ Diphthong [əɪ ~ ai]	2,758	3,849	6,816	13,423
AIS2 – MHG /ei/ Diphthong [ɔɪ ~ ai]	2,371	3,300	5,657	11,328
ANN – Nasal 'a' before 'n' [ã ~ an]	2,402	2,721	4,880	10,003
FRV1 – Unrounded Front Vowel [eː ~ øː]	695	750	1,286	2,731
FRV2 – Unrounded Diphthong [aɪ ~ ɔʏ]	712	892	1,540	3,144
FRV3 – Unrounded Front Vowel [iə ~ ʏː]	1,307	1,988	3,288	6,583
FRV4 – MHG /uo/ Diphthong [uə ~ uː]	1,752	2,266	3,651	7,669
LEO – Lower Long Vowel [ɛː ~ eː]	1,245	2,304	4,124	7,673
SFV – Stop-Fricative Variation [ɪç ~ ɪk]	602	963	1,581	3,146
STPV – Palatal Coda -st [ʃ ~ s] with verbs	376	522	1,004	1,902
STP6 – Palatal Coda -st [ʃ ~ s] with six verbs	568	640	1,324	2,532
STPI – Palatal Coda -st [ʃ ~ s] with 'ist'	1,715	2,486	4,614	8,815
STPO – Palatal Coda -st [ʃ ~ s] with non-verbs	1,409	2,012	3,740	7,161
TOTAL Phonological Tokens	17,912	24,693	43,505	86,110

Morphosyntactic Variables & Tokens	Panel 1982	Panel 2017	Twin 2017	TOTAL
DAS – Definite Neuter Article: des ~ das	2,106	3,414	6,191	11,711
EDP – Plural Verb Inflection: -ed ~ -en	942	2,319	3,647	6,908
IRV1 – Irregular Verb: gange ~ gehen	263	371	718	1,352
IRV2 – Irregular Verb: stande ~ stehen	195	200	456	851
IRV3 – Irregular Verb: hen ~ haben	1,081	1,865	3,462	6,408
NEG – Negative Marker: ned ~ nich(t)	1,388	1,942	3,332	6,662
PVB – Periphrastic Subjunctive: dääd ~ würde	166	206	414	786
SAF1B – Swabian Affix: bissle ~ bisschen	252	307	600	1,159
SAF3 – Swabian Affix: nââ- ~ hin-	133	140	302	575
SAF5 – Swabian Affix: Ø ~ ge-	592	1,238	1,898	3,728
TOTAL Morphosyntactic Tokens	7,118	12,002	21,020	40,140

Dialect Density Index (DDI)	Panel 1982	Panel 2017	Twin 2017
Schwäbisch Gmünd	61.2%	47.4%	44.6%
Stuttgart	47.9%	26.6%	30.8%
OVERALL	56.8%	38.4%	39.9%

Table 2. Token Counts for 20 Linguistic Features under Investigation.

3.5 Dialect Density

Modelled on the Dialect Density Measure (DDM) by Wolfram and others (Van Hofwegen and Wolfram 2010; Oetting and McDonald 2002), the 20 linguistic variables were aggregated into a Dialect Density Index (DDI), which is used to assess the level of dialect spoken in each community and across time.[7] DDI is a token-based composite metric, calculated as the total dialect variants divided by the total variants (dialect and standard). Beaman (2020) found consistent patterns of dialect levelling and attrition in both communities and in real- and apparent-time: specifically, the later recordings (2017), as well as the younger age groups, showed lower levels of dialect density than the earlier recordings and the older generations, particularly in Stuttgart (see Table 2), signaling the younger, urban generation to be leading the change toward a new supraregionalized southwestern variety of German (Beaman 2020).

4. Analysis and Results

The analysis of this study of coherence is organized into three sections: patterns of variability isolating more or less coherent (dia)lects (Section 4.1), measures for quantifying relative levels of coherence across different (dia)lects (Section 4.2), and coherent variable clusterings identified through PCA's variable weightings (Section 4.3). The discussion is focussed on eight speaker group based two study types (panel and twin), two recording years (1982 and 2017 for the panel study) or age groups (older and younger for the twin study), and two communities (Stuttgart and Schwäbisch Gmünd).

4.1 Patterns of variability

Principal components analysis (PCA) is one of the oldest and most commonly used exploratory methods for reducing the dimensionality of complex datasets into smaller more informative sets of variables to improve interpretability. PCA transforms a large set of variants into a smaller set of uncorrelated variants, preserving as much variability as possible in its first components. Figure 2 presents the PCA results for all 20 linguistic variables for the first six principal components (PCs). Six PCs were selected as they explain 83.5% of the variance and because the unexplained parts are of similar size in all speaker groups (PC1/PC2 accounts for 68.0%, PC3/PC4 for 9.5%, and PC5/PC6 for 6.3%). The left plots show the real-time analysis (i.e., 20 panel speakers each interviewed twice, once in 1982 and again in 2017), and the right plots the apparent-time analysis (i.e., 40 twin study speakers, younger and older age groups). Grey denotes Schwäbisch Gmünd speakers and black Stuttgart speakers; dotted lines with open circles and triangles signify 1982 panel speakers and 2017 older speakers (twin study), while solid lines with solid circles and triangles represent 2017 panel speakers and 2017 younger speakers (twin study).

[7] One reviewer pointed out that this approach may be overly general because the tokens are not evenly distributed across all variables (naturally some variables are more frequent than others) and not all variables are equally dialectal (some are Swabian-specific, others are Alemannic, and others more broadly regional). However, DDI serves the intended purpose which is to provide an overall metric on the level of dialect (as opposed to standard language) spoken by the speakers in each of the communities for comparison purposes.

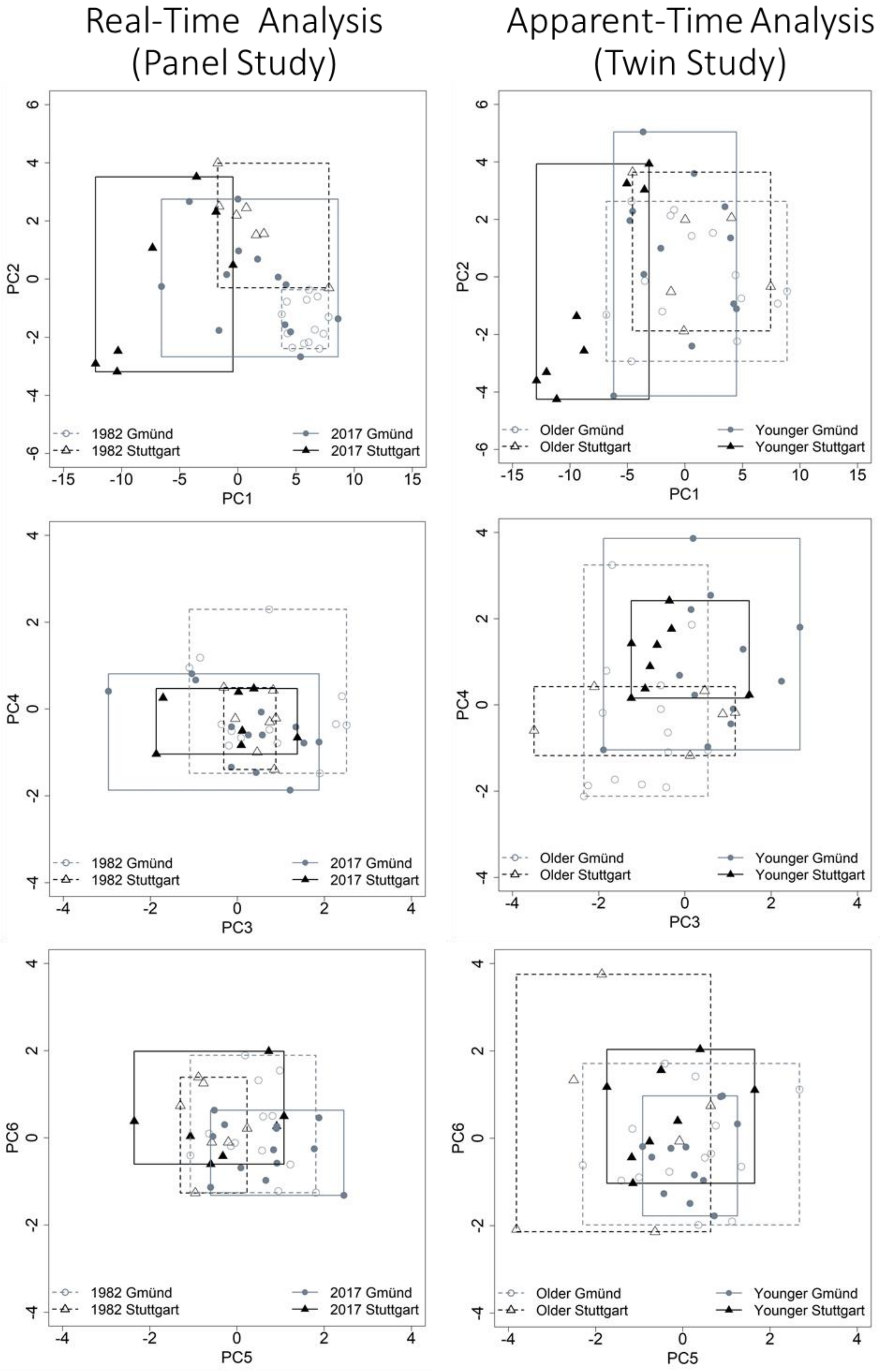

Figure 2. Principal Components Analysis for 20 Swabian Linguistic Features

Based on these 20 variables together, the real-time analysis (Figure 1, left plots) indicates that the 1982 Schwäbisch Gmünd speakers have the tightest coherence: the small grey dashed box designates the least variability. In 1982, the Stuttgart and Schwäbisch Gmünd lects showed no overlap (boxes with dashed lines, top left plot), signalling two distinct varieties. However, by 2017, both lects have expanded (boxes with solid lines, top left plot), exhibiting considerably greater variability and hence weaker coherence. Concomitantly with the decline in coherence is a reduction in dialect density. For the 20 panel speakers in 1982, the average DDI was 56.8%, which dropped to 38.4% in 2017, a decrease of 18.4% over these speakers' lifespans, with the largest decline in Stuttgart, 21.3% (see Table 2). Thus, it appears that the loss of dialect variants is associated with lower levels of coherence, as speakers "juggle" and "mix" two different varieties, Swabian and standard German, in their everyday speech.

The apparent-time analysis (Figure 1, right plots) depicts a congruent, albeit somewhat different, picture. The lectal expansion and overlap seen between the 1982 and 2017 panel speakers is also visible in the older and the younger twin study speakers who are more similar to each other, in both Stuttgart and Schwäbisch Gmünd (top right plot). The average DDI for the older twin speakers in 2017 was 45.1%, comparable to the 2017 older panel study speakers (44.4%), providing support for the general premise that, across their lifetime, individual speakers typically follow the community trend. In the twin study, DDI drops to 20.5% for the younger speakers. The PCA visualization offers a clear illustration of the extensive dialect levelling that has occurred in Swabia over the last 35 years, the consequent fusion of the two local varieties, and the emergence of a new supraregional southwestern variety of German (see Beaman 2020 for further discussion). It is important to note that 2017 panel speakers are the same ages as the older 2017 twin speakers (median ages of 57 and 60, respectively). Thus, a comparison of the solid lined boxes from the real-time analysis (left plots) with the dotted lined boxes in the apparent-time analysis (right plots) reveals similar distributions, establishing the compatibility of both study types for analysing language change.

Considering PC1/PC2 on their own, however, can be misleading. For example, with the 1982 panel speakers in Schwäbisch Gmünd (small grey dotted box, top left plot), a greater amount of the variability can be found on PC3/PC4 (larger grey dotted box, middle left plot). Looking the variable loadings on PC3/PC4, two variables stand out as primarily responsible for this difference: stop-fricative variation (SFV) and the definite neuter article (DAS), two non-traditional Swabian features which have moved into the conservative Schwäbisch Gmünd dialect from middle and northern varieties of German. Two additional variables stand out on PC5/PC6: nasalisation (ANN) and the diminutive suffix *-le* (SAF1B). We discuss the variable loadings further in Section 4.3. With these interesting and indicative findings, the question remains, how do we quantify the amount of variability, and thus the level of coherence, within each lect. And how much do the lects differ from each other in terms of coherence? The following section seeks to answer this question.

4.2 Measuring coherence

To quantitatively measure the level of coherence in a lect, four different metrics have been considered (see Table 3). The first and most straightforward measure is the RECTANGLE VOLUME, i.e., the size and capacity of the rectangle which contains all of the points in a given lect (see Figure 1). Volume is calculated as the rectangle's length times its width times its height times the amount of variance explained by each dimension (in our case, the first six principal components). The volume metric shows substantial increased variability between 1982 and 2017 for the panel study participants, with further increases for the younger and older twin study participants. This increase in variability is likely due to two factors which result from the 35-year time gap: first, the individuals with greater variability are older in each study and hence have vaster life experiences often gained through increased exposure with the standard language (Baayen, Beaman, and Ramscar 2021); and second, the world itself has become increasingly more disparate and complex, bringing uncertainty to the speakers as to which variant to use in a given context. Both of these aspects enrich diversity and variability in language use. Both studies cover spans of 35-37 years, yet the real-time decline in coherence across the years is significantly greater than the apparent-time differences across the generations, reflecting the immense societal changes that have occurred in Germany over the last four decades. It appears that the panel speakers are playing "catch-up" in order to "keep up" with the massive changes and rampant dialect levelling transpiring in Swabia. These findings indicate that real-time change can sometimes occur more rapidly than apparent-time change, particularly in these two communities in the current situation of persistent, pervasive, and precipitous dialect levelling.

Community & Metric	Real-Time (Panel Study)			Apparent-Time (Twin Study)		
	1982	2017	decrease in coherence	Younger	Older	decrease in coherence
Rectangle volume (PC1-PC6)						
Schwäbisch Gmünd	2640	16675	531.5%	34162	64594	89.1%
Stuttgart	980	9010	819.7%	13418	33874	152.4%
Confidence ellipsoid area (PC1-PC2)						
Schwäbisch Gmünd	17.5	113.1	546.3%	194.9	161.3	-17.2%
Stuttgart	29.7	136.9	360.9%	69.6	154.3	121.7%
Convex hull area (PC1-PC2)						
Schwäbisch Gmünd	5.8	47.3	715.5%	59.1	55.1	-6.8%
Stuttgart	7.0	30.4	334.3%	17.9	31.9	78.2%
Functional Diversity (PC1-PC6)						
Schwäbisch Gmünd	0.005	0.099	2055.1%	0.100	0.112	12.6%
Stuttgart	0.021	0.052	145.5%	0.015	0.051	237.7%

Table 3. Lectal Variability Measurements in Real- and Apparent-Time

One advantage of the volume measure is its breadth, covering six (or more) principal components, and the ease with which it can be calculated. However, a major setback is its extreme sensitivity to outliers and its variation under rotations of the latent space. Thus, a second metric considered for assessing lectal variability is the area of the CONFIDENCE ELLIPSOID for PC1 and PC2. The area of the ellipsoid shows a similar decrease in coherence in real- and apparent-time, although significantly smaller than with the volume metric, particularly for Stuttgart in real-time and Schwäbisch Gmünd in apparent-time. This is easily explained because the ellipsoid is only two-dimensional, whereas the rectangle supports multiple dimensions; and, as we have seen, much of the variation lies on the third and fourth PCs which are not accounted for in the two-dimensional confidence ellipsoid. Thus, a main drawback to the ellipsoid is its limited dimensionality, sensitivity to outliers, and lack of stability under rotations of latent space.

A third approach to calculating the variability in a lect is the area of the CONVEX HULL, i.e., the smallest shape that encloses all of the points in a given set, regardless of the shape or distribution of the points. A convex hull can be visualized as a string wrapped around the points along the border of the set (calculated with the *chull* function in the R Package *grDevices*, version 3.6.0). A similar pattern as with the ellipsoid emerges: real-time indicates a greater decrease in coherence than apparent-time. While the area of the convex hull represents an improvement over the volume of a rectangle because it avoids unused “white space”, it poses similar limitations to the ellipsoid with the complexity of incorporating more than two dimensions. All three of these metrics are highly sensitive to low token counts, making smaller groups appear more coherent than they really are.

Because of these shortcomings, we considered a fourth metric, FUNCTIONAL DIVERSITY (FD), a concept borrowed from ecology for measuring ecosystem processes and resilience to environmental changes (Laliberté and Legendre 2010; Villéger, Mason, and Mouillot 2008). FD provides a method for understanding and interpreting the patterns of co-occurrence and the role of different “traits” (cf. linguistic features) on the functioning of an ecosystem (Tilman 2001). One component of functional diversity, FUNCTIONAL RICHNESS, uses principal coordinates analysis (PCoA) to estimate the volume of data in the minimum convex hull for a community (Laliberté and Legendre 2010:310). Functional richness is defined as the amount of “niche space” filled by a feature or set of features in a community: the more functionally different the variables are (in our case, the more variability in the variables), the higher the index (Legras, Loiseau, and Gaertner 2018). The results are shown in the fourth section of Table 3 (calculated using *FRic* option with the *dbFD* function in *FD* R package, version 1.0-12). The *FD* metric shows a similar trend as the others, with the greatest decrease in coherence appearing in Schwäbisch Gmünd over the lifespan and the second-largest decrease emerging in Stuttgart between the younger and older generations. The advantages of FD over other methods of calculating variability explored here are its ability to handle multiple dimensions, asymmetric binary variables, variable weights, and empty cells (due to sparse data), all critical parameters with sociolinguistic data.

In sum, the younger speakers in both study samples show greater coherence, i.e., less diversity, than the older speakers, signaling greater linguistic stability. Recall that the real- and apparent-time analyses both cover time spans of 35-37 years, meaning that the 1982 panel speakers and the younger 2017 twin speakers are roughly the same age, albeit separated by 35 “real” years. Younger speakers may exhibit greater lectal coherence simply because they are more conditioned to using a single variety, the standard language, in school, and quite naturally, due to their age, have had more limited life experiences and contact with speakers of other language varieties. Older speakers have developed more extensive and diverse repertoires over their lifespans, reflecting a mixture of both Swabian and standard German, which produces a broader feature pool to draw from (see Baayen, Beaman, and Ramscar (2021) for an empirical analysis of dialect and standard vocabulary growth across the lifespan of the 20 Swabian panel speakers).

4.3 Variable clusterings and weightings

The third coherence question this chapter addresses is in what ways do the linguistic features characteristic of a sociolect covary and which ones carry the most coherence weight? Since PCA predicts the proportion of variance explained by each variable, the variable loadings can be visualized in latent space, as seen in Figure 2 for PC1 and PC2 (plotted with the function *autoplot* in R package *ggplot2*, version 3.2.1). The arrows show the loadings and clustering for the 20 linguistic features. The angles of the arrows indicate the degree of correlation between the linguistic feature and the principal component, and their length signifies how much of the variation is explained by that feature: specifically, the more horizontal or vertical the arrow, the stronger the correlation with PC1 or PC2, respectively, while the longer the arrow, the greater the explained variance.

Figure 2 exposes two meaningful clusters which have been enlarged for readability in Figure 3. The first cluster is primarily defined by the first principal component: four front rounded vowels (FRV1, FRV2, FRV3, FRV4), two (ai) diphthongs (AIS1, AIS2), three irregular verb stems (IRV1, IRV2, IRV3), and three Swabian-specific morphosyntactic constructions (EDP, SAF3, PVB). These are the traditional, conservative features unique to the Swabian dialect; thus, this cluster is called Traditional Swabian (TS). The Swabian dialect is a conservative variety with a long tradition of evoking sundry and opposing images such as "inventive", "hard-working" and "thrifty" but also "backward", "uneducated" and "miserly" (Beaman 2020). Hence, it is not surprising that speakers would react differently to traditional Swabian-specific features than with features entering the dialect from other varieties of German.

The second cluster lies between PC1 and PC2 and consists of regional features that are prevalent throughout southern Germany, including Bavaria to the east, Baden to the west, and Switzerland to the south: palatalisation of coda -*st* (STPA), diminutive affix -*le* (SAF1), dropping of the past participle *ge*- (SAF5), negative markers (NEG), and nasalisation of *an* (ANN). Since these features are common in many southern German dialects, this cluster is called Southern Regional (SR). There are two exceptions, the periphrastic subjunctive (PVB) and the lowering of [eː] (LEO), southern regional features which pattern more closely with the Swabian-specific variables. Although LEO lies on the cusp and could theoretically be grouped with either cluster, based on its correlation coefficient (see Table 4), it fits better with the Traditional Swabian variables.

This leaves two features that do not cluster with the others: the definite neuter article (DAS) and stop-fricative variation (SFV), both of which are strongly aligned along the second principal component, one positive and one negative. These features are spreading into Swabian from the north (Spiekermann 2008:186), thus, this group is called Northern Influence (NI). Both of these variables are relatively stable and not undergoing rampant change and levelling as are the other variables in this study (Beaman 2020).

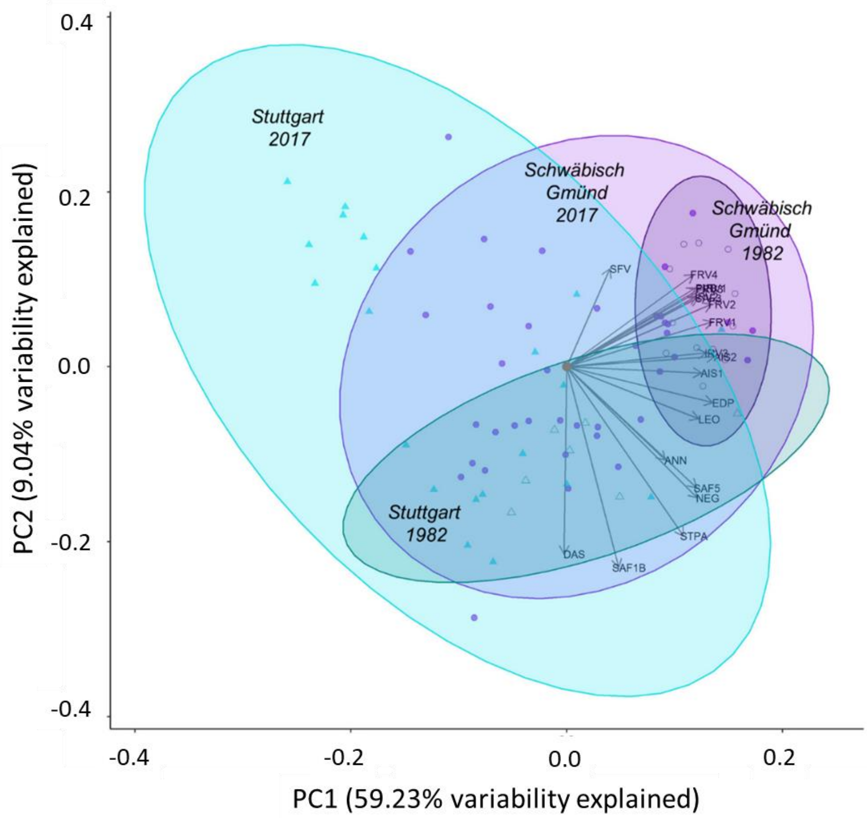

Figure 2. Variable Loadings for PC1 and PC2 – Panel and Twin Study

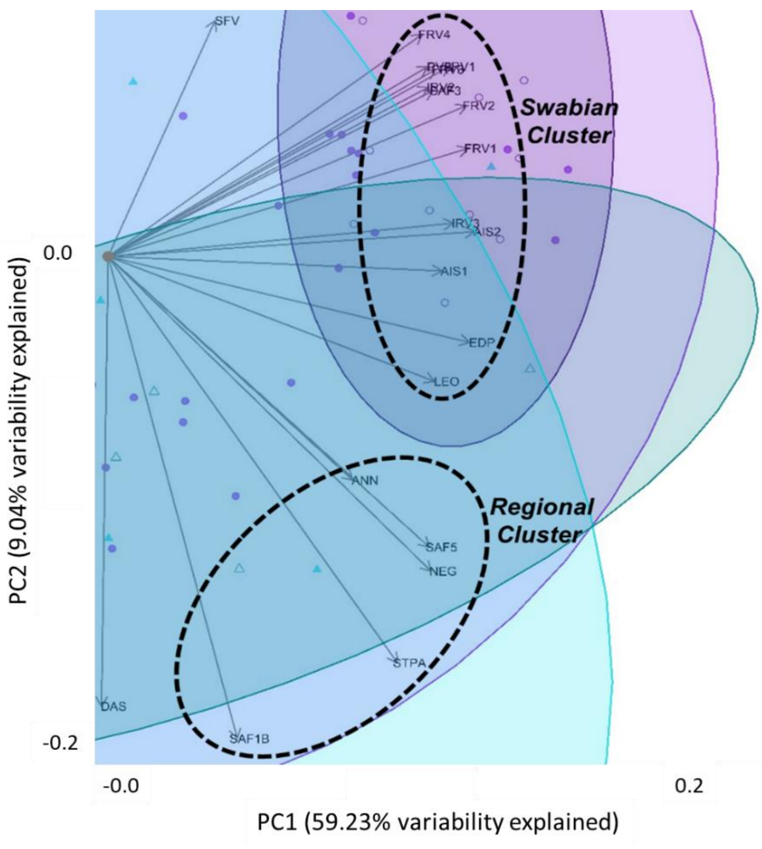

Figure 3. Traditional Swabian and Southern Regional Clusters – Panel and Twin Study

Table 4 reports the correlation coefficients for the first three principal components for the 20 linguistic variables, with respect to a reference variable (indicated as "RefVar"), sorted by cluster and then descending by the proportion of explained variance. The irregular verb *hen ~ haben* (IRV3) was chosen as the reference for the TS cluster, and palatalization of coda -*st* (STPA) was chosen for the SR cluster since both features are at or near the center of their cluster and are considered prototypical dialect variants. Correlations over .8 and under -.8 (shaded in grey) are considered to be the most coherent features, divulging that the strongest coherence is with the TS features, followed by the SR features, and finally, the NI features which exhibit negative correlations. Another exception is nasalisation of *an* (ANN), which shows a strong negative correlation as opposed to the positive correlations of the other TS features in the first two PCA components, supporting Auer's claim that ANN may be lexically constrained (personal communication).

Linguistic Features	Type	Cluster	TS	SR
IRV3 – Irregular Verb: hen ~ haben	SWG	TS	RefVar	0.188
FRV1 – Unrounded Front Vowel [eː ~ øː]	SWG	TS	0.998	0.127
FRV2 – Unrounded Diphthong [ai ~ ɔʏ]	SWG	TS	0.978	-0.023
IRV1 – Irregular Verb: gange ~ gehen	SWG	TS	0.945	-0.143
SAF3 – Swabian Affix: nââ- ~ hin-	SWG	TS	0.943	-0.149
IRV2 – Irregular Verb: stande ~ stehen	SWG	TS	0.942	-0.154
FRV3 – Unrounded Front Vowel [iə ~ ʏː]	SWG	TS	0.936	-0.170
PVB – Periphrastic Subjunctive: dääd ~ würde	REG	TS	0.933	-0.177
AIS2 – MHG /ei/ Diphthong [ɔi ~ ai]	SWG	TS	0.910	0.577
FRV4 – MHG /uo/ Diphthong [uə ~ uː]	SWG	TS	0.862	-0.335
EDP – Plural Verb Inflection: -ed ~ -en	SWG	TS	0.827	0.708
AIS1 – MHG /iː/ Diphthong [əi ~ ai]	SWG	SR	0.287	0.995
SAF5 – Swabian Affix: Ø ~ ge-	REG	SR	0.744	0.796
NEG – Negative Marker: ned ~ nich(t)	REG	SR	0.489	0.949
LEO – Lower Long Vowel [ɛː ~ eː]	REG	SR	0.462	0.958
ANN – Nasal 'a' before 'n' [ã ~ an]	SWG	SR	0.346	0.987
STPA – Palatal Coda -st [ʃt ~ st]	ALM	SR	0.188	RefVar
SAF1A – Swabian Affix: -le ~ -chen	ALM	SR	-0.368	0.844
DAS – Definite Neuter Article: des ~ das	REG	NI	0.983	0.365
SFV – Stop-Fricative Variation [ɪç ~ ɪk]	REG	NI	-0.965	0.078

Table 4. Correlations of Explained Variances by Variable – Panel and Twin Study
TS=Traditional Swabian (reference: IRV3); SR=Southern Regional (reference: STPA);
NI=Northern Influence; correlations greater than .8 or less than -.8 are shaded in grey

5. Conclusion

This investigation has confirmed the importance of the concept of coherence in a comprehensive theory of linguistic change. Methodologically, the findings support the use of principal components as an effective tool for examining and assessing levels of sociolectal coherence and its change across time. Our analysis has identified a reliable, quantifiable metric for calculating the level of coherence in a lect by borrowing a method from ecology for examining the functional richness within an ecosystem. This multidimensional approach measures the amount of (in)coherence (i.e., diversity), while properly handling asymmetric binary variables, differing variable weights, and sparse or empty cells. We compared levels of coherence across eight sociolects and confirmed the overall hypothesis of this study: more coherent lects are more resistant to change, while less coherent lects are more susceptible to variation and change. The results of this investigation also reveal that the etymological origin of the variable – traditional Swabian versus more widely dispersed regional features – has a strong coercive effect on linguistic coherence. Prior research shows that Swabian orientation, as expressed through greater use of the traditional dialect, provides a sense of shared social motivation across the community (Beaman 2020), inciting individuals to move in a concerted, unified, and coherent direction (Tamminga 2019).

The findings confirm that coherence is a matter of relativity and degree (Tomaschek, Hendrix, and Baayen 2018) and changes over time. In 1982, speakers used greater levels of dialect, and concomitantly, demonstrated higher levels of coherence. Over the 35-year timespan of this study, the speakers in this study have converged toward the standard language, and broad dialect mixing has become the norm, rendering individual lects more diverse and thus less coherent. It is evident that the less coherent lects are undergoing the greatest amount of change, making them less stable over time. As Swabian has continued to converge toward the standard language, resulting in lower levels of dialect-standard mixing and greater use of standard German, coherence has begun to increase. This process leads to the speculation that in the ensuing years, as Swabian continues its convergence to the standard language and as more of the changing variables stabilize, coherence will continue to strengthen. But what will happen when new changes not targeted toward the standard language occur? As the Roman scholar Marcus Terentius Varro (116-27 BC) observed, *consuetudo loquendi est in motu*, ‘the vernacular is always in motion’ (translated by Taylor 1975). As some variables stabilize, others begin to change, keeping sociolinguists constantly on the move in search of the cognitive coherence of sociolects.

References

Auer, Peter. 2020. “Dialect (Non-)Acquisition and Use by Young People of Migrant Background in Germany.” *Journal of Multilingual and Multicultural Development*: 1–15.

Baayen, R. Harald, Karen V. Beaman, and Michael Ramscar. 2021. “Deconfounding the Effects of Competition and Attrition on Dialect across the Lifespan: A Panel Study Investigation.” In *Language Variation and Language Change Across the Lifespan: Theoretical and Empirical Perspectives from Panel Studies*, eds. Karen V. Beaman and Isabelle Buchstaller. New York: Routledge.

Beaman, Karen V. 2020. “Coherence in Real- and Apparent-Time: A Sociolinguistic Study of Swabian. Unpublished PhD Thesis.” Queen Mary University of London.

Blondeau, Hélène. 2001. “Real-Time Changes in the Paradigm of Personal Pronouns in Montreal French.” *Journal of Sociolinguistics* 5(4): 453–74.

Bucholtz, Mary, and Kira Hall. 2005. “Identity and Interaction: A Sociocultural Linguistic Approach.” *Discourse Studies* 7(4–5): 585–614.

Eckert, Penelope. 2008. "Variation and the Indexical Field." *Journal of Sociolinguistics* 12(4): 453–76.

———. 2012. "Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation." *Annual Review of Anthropology* 41(June): 87–100.

Guy, Gregory R. 1980. "Variation in the Group and the Individual: The Case of Final Stop Deletion." In *Locating Language in Time and Space,* ed. William Labov. New York: Academic Press, 1–36.

———. 2013. "The Cognitive Coherence of Sociolects: How Do Speakers Handle Multiple Sociolinguistic Variables?" *Journal of Pragmatics* 52: 63–71.

———. 2014. "Bricks and Bricolage: The Question of Sociolinguistic Coherence." *New Ways of Analyzing Variation 43. Keynote Lecture. Chicago, IL. October 23-26, 2014*.

Guy, Gregory R., and Frans Hinskens. 2016. "Linguistic Coherence: Systems, Repertoires and Speech Communities." *Lingua* 172–173: 1–9.

Van Hofwegen, Janneke, and Walt Wolfram. 2010. "Coming of Age in African American English: A Longitudinal Study." *Journal of Sociolinguistics* 14(4): 427–55.

Horvath, Barbara M., and David Sankoff. 1987. "Delimiting the Sydney Speech Community." *Language in Society* 16(2): 179–204.

Labov, William. 1984. *Field Methods of the Project on Linguistic Change and Variation*.

Laliberté, Etienne, and Pierre Legendre. 2010. "A Distance-Based Framework for Measuring Functional Diversity from Multiple Traits." *Ecology* 91(1): 299–305.

Legras, G., N. Loiseau, and J. C. Gaertner. 2018. "Functional Richness: Overview of Indices and Underlying Concepts." *Acta Oecologica* 87: 34–44. https://doi.org/10.1016/j.actao.2018.02.007.

Meyerhoff, Miriam, and Steffen Klaere. 2017. "A Case for Clustering Speakers and Linguistic Variables: Big Issues with Smaller Samples in Language Variation." In *Language Variation - European Perspectives VI,* eds. Isabelle Buchstaller and Beat Siebenhaar. John Benjamins, 23–46.

Milroy, James, and Lesley Milroy. 1985. *Authority in Language: Investigating Standard English. Kindle Edition*. Kindle. London: Routledge.

Milroy, Lesley. 1987. *Language and Social Networks*. 2nd Ed. Oxford: Basil-Blackwell.

Naro, Anthony J. 1981. "The Social and Structural Dimensions of a Syntactic Change." *Language* 57(1): 63–98.

Oetting, Janna B., and Janet L. McDonald. 2002. "Methods for Characterizing Participants' Nonmainstream Dialect Use in Child Language Research." *Journal of Speech, Language, and Hearing Research* 45(3): 505–18.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ruoff, Arno. 1997. "Sprachvarietäten in Süddeutschland." In *Värietäten Des Deutschen: Regional- Und Umgangssprachen,* ed. G. Stickel. Berlin: De Gruyter, 142–54.

Sharma, Devyani. 2018. "Style Dominance: Attention, Audience, and the 'Real Me.'" *Language in Society* 47(1): 1–31.

Spiekermann, Helmut. 2008. *Sprache in Baden-Württemberg. Merkmale Des Regionalen Standards*. Tübingen: Max Niemeyer Verlag.

Svenstrup, Christoph Hare. 2019. "'weil de Zukunft in Hochdeutsch Liegt...' ('because the Hochdeutsch Is the Future...'): Language Attitudes amongst Adolescents from the Stuttgart Area." University of Copenhagen.

Tamminga, Meredith. 2019. "Interspeaker Covariation in Philadelphia Vowel Changes." *Language Variation and Change* 31(2): 119–33.

Taylor, Daniel J. 1975. *Declinatio: A Study of the Linguistic Theory of Marcus Terentius Varro (Studies in the History of Linguistics, Volume 2)*. Amsterdam: John Benjamins.

Tilman, D. 2001. "Functional Diversity." In *Encyclopedia of Biodiversity*, ed. S. A. Levin. Academic Press, 109–20.

Tomaschek, Fabian, Peter Hendrix, and R. Harald Baayen. 2018. "Strategies for Addressing Collinearity in Multivariate Linguistic Data." *Journal of Phonetics* 71: 249–67. https://doi.org/10.1016/j.wocn.2018.09.004.

Villéger, Sébastien, Norman W.H. Mason, and David Mouillot. 2008. "New Multidimensional Functional Diversity Indices for a Multifaceted Framework in Functional Ecology." *Ecology* 89(8): 2290–2301.

Wagner, Suzanne Evans, and Isabelle Buchstaller. 2018. *Panel Studies of Variation and Change*. eds. Suzanne Evans Wagner and Isabelle Buchstaller. New York: Routledge.

Walker, James A., and Miriam Meyerhoff. 2013. "Studies of the Community and the Individual." In *Oxford Handbook of Sociolinguistics*, eds. Robert Bayley, Richard Cameron, and Ceil Lucas. Oxford: Oxford University Press, 175–94.

Watt, Dominic. 2000. "Phonetic Parallels between the Close-Mid Vowels of Tyneside English: Are They Internally or Externally Motivated?" *Language Variation and Change* 12(1): 69–101.

Weinreich, Uriel, William Labov, and Marvin I. Herzog. 1968. "Empirical Foundations for a Theory of Language Change." In *Directions for Historical Linguistics*, eds. Winfred Lehmann and Yakou Malkiel. Austin: University of Texas, 95–188.

Woo, Joyce, Tim Gadanidis, and Naomi Nagy. "SUBMITTED. Co-Variation in Toronto's Heritage Cantonese: Primarily a Phonological Process."