Zeitschrift für
SPRACHVARIATION
& SOZIOLINGUISTIK
JOURNAL OF LANGUAGE
VARIATION AND SOCIOLINGUISTICS

https://doi.org/10.5282/jlvs/10

Changing socioindexicalities in Swabian: Real-time change in subject pronoun use

Karen V. Beaman[1]

[1] Quantitative Linguistics, University of Tübingen, Tübingen, Germany
E-mail: karenbeamanvslx@gmail.com

Received 22.08.2024
Accepted for publication 05.02.2025
Published 22.05.2025

Abstract

This paper investigates the variable presence or absence of subject personal pronouns in a real-time, spontaneously spoken corpus of Swabian German. Pronominal finite verb constructions were manually extracted and coded for an array of linguistic (e.g., person-number, clause type), cognitive (e.g., coreferentiality, priming, lexical frequency), and social constraints (e.g., age, gender, education, and style). The results reveal considerable stability in the linguistic and cognitive constraints over time, but changing socioindexicalities in the use of subject pronouns: older, more highly educated women in urban environments use more overt pronominal subjects than younger, less educated, men who use more null subjects. The findings underscore the role of the linguistic architecture in understanding morphosyntactic variation and change, specifically the stability of linguistic and cognitive constraints and the instability and individualism of social influences.

Keywords: variationist sociolinguistics, real-time change, sociogrammatical variation, null subjects, German dialects, Swabian

1 Introduction

One linguistic feature that has been used to classify languages is the presence or absence of pronominal subjects, distinguishing so-called "null subject languages" and "non-null subject languages." Spanish, for example, is a canonical null subject languages in which pronominal subjects are generally not expressed (Dryer 2013), as exemplified in (1).

(1) Spanish:
Ø canto ~ yo canto 'I sing'
Ø cantas ~ tú cantas 'you sing'

Typologists and formalists have attributed the "licensing of null subjects" in languages like Spanish, Portuguese, Italian, Persian, and Greek, to name a few, to their inflectional paradigms, which are morphologically salient and referentially unambiguous (Jaeggli & Safir 1989). In null subject languages with rich inflections, subject pronouns are not required for constructing a grammatical sentence, and thus their overt use has been ascribed to stylistic or emphatic effects (cf. Chomsky's *avoid pronoun principle* (1981: 65)). Some sociolinguists maintain that the use of overt pronouns in a null subject language is categorical in contrastive contexts (e.g., Bayley & Pease-Alvarez 1997); however, other

studies have found that this is not always the case (e.g., Otheguy & Zentella 2012).

In contrast to the null subject languages are the non-null subject languages, such as English (Travis & Torres Cacoullos 2012) and German (Volodina & Weiß 2016), in which pronoun omission is not allowed in the standard varieties, as shown in (2) for German. Yet, subject pronoun omission has been attested in both English (Travis & Torres Cacoullos 2012) and Swabian German (Beaman 2022; Bohnacker 2013), as seen in (3).

(2) Standard German:
*ich singe ~ *Ø singe* 'I sing'
*du singst ~ *Ø singst* 'you sing'

(3) Swabian German:
i sing ~ Ø sing 'I sing'
du singsch ~ Ø singsch 'you sing'

This has given rise to the notion of "partial null subject languages" (Holmberg et al. 2009) or "mixed languages" (Dryer 2013), which maintains that different syntactic or discursive environments can influence the expression or omission of a subject pronoun. In a preliminary study of subject pronoun use in Swabian, Beaman (2022) found 15% (*n*=30,108) omission across all singular forms, with a notable 45% (*n*=1802) omission with second-person singular forms. While the formalist/functional argument maintains that this difference in Swabian is largely due to the distinct verbal inflection in the second-person singular (see Tab. 1), further analysis shows that there are a variety of internal and external influences involved. Most prior studies of subject pronoun use in German have taken a descriptivist or functional/formalist theoretical approach (e.g., Abraham 1993; Axel & Weiß 2011; Trutkowski 2010; Volodina & Weiß 2016). To date, no real-time quantitative sociolinguistic variationist studies have been conducted on subject pronoun use in varieties of German – a desideratum this study aims to address.

Table 1: Swabian and Standard German verbal paradigms (present tense). The first line of each row represents typical Central Swabian regiolect (adapted from Frey 1975), the second line is standard German (Duden 2015), and the third is the English gloss.

Person-Number	Present (*Präsens*)	Preterite (*Perfekt*)	Imperfect (*Konjunktiv II*)
SINGULAR:			
First-person	*i sing* *ich singe* 'I sing'	*i hann sunge* *ich habe gesungen* 'I sang/had sung'	*i dääde singe* *ich würde singen* 'I would sing'
Second-person	*du singsch(d)* *du singst* 'you sing'	*du hasch sunge* *du hast gesungen* 'you sang/had sung'	*du däädsch singe* *du würdest singen* 'you would sing'
Third-person	*er/sie/es sing(d)* *er/sie/es singt* 'he/she/it sings'	*er/sie/es hodd sunge* *er/sie/es hat gesungen* 'he/she/it sang/had sung'	*er/sie/es dääde singe* *er/sie/es würde singen* 'he/she/it would sing'
PLURAL:			
First-person	*mer singed/et* *wir singen* 'we sing'	*mer henn sunge* *wir haben gesungen* 'we sang/had sung'	*mer däädet singe* *wir würden singen* 'we would sing'
Second-person	*ihr singed/et* *ihr singt* 'you sing'	*ihr henn sunge* *ihr habt gesungen* 'you sang/had sung'	*ihr däädet singe* *ihr würdet singen* 'you would sing'
Third-person	*sie singed/et* *sie singen* 'they sing'	*sie henn sunge* *sie haben gesungen* 'they sang/had sung'	*sie däädet singe* *sie würden singen* 'they would sing'

Empirical evidence from a corpus of spontaneously spoken Swabian reveals that null subjects can be found in all person-number combinations, as the following examples show[1].

(4) first-person singular:
Ø kenn mi et so aus
ich kenne mich nicht so aus
'I don't know much about it'
[S014-82-I-2-Markus-00:13:07]

(5) second-person singular:
wenn Ø bloß Schwäbisch kãsch
wenn du bloß Schwäbisch kannst
'if you-SG can only speak Swabian'
[S031-82-I-2-Jurgen-00:17:39]

(6) third-person singular:
Ø heert si hald komisch ã
es hört sich halt komisch an
'it actually sounds funny'
[S026-82-I-2-Berdine-00:04:43]

(7) first-person plural:
dann Ø quer durch mit dm Rädle durchgfahre sin
dann wir quer durch [den Teig] mit dem [Pizza-Schneider] durchgefahren sind
'then we went straight through [the dough] with the [pizza cutter]'
[S015-82-I-2-Ricarda-00:00:22]

(8) second-person plural:
Ø lernt echte Menschen kennen
ihr lernt echte Menschen kennen
'you-PL meet real people'
[S036-17-I-1-Helmut-01:16:27]

(9) third-person plural:
Ø gebet sich aso au Mieh
sie geben sich also auch Mühe
'they also put in a lot of effort'
[S034-82-I-1-Bertha-00:27:36]

These examples raise the question: if Swabian German is truly a non-null subject language, why do native speakers sometimes omit pronouns in spontaneous speech and in online media (e.g., Dittmann 2006; Schwitalla 2006)? Are referential subjects dropped "out of the blue" (Trutkowski 2010) in Swabian or do they follow similar structural and discursive constraints found in other language varieties? Are there social and stylistic indexicalities that constrain the use of pronominal subjects in Swabian (e.g., gender, education, age, formality)? And importantly, have pronoun rates and constraints changed or remained stable over time?

To answer these questions, this study investigates whether subject pronoun expression (SPE) in Swabian:

- follows similar structural and linguistic constraints as found in other linguistic varieties → structural uniformity hypothesis
- conforms to cognitive-discursive constraints which appear to be universal across languages → cognitive universal hypothesis
- is more common among women and in more formal contexts in which standard language norms are generally observed → social sensitivity hypothesis

2 Background

According to the standard German literature, null subjects are not allowed in any modern standard varieties of German (Dryer 2013; Rosenkvist 2009). They have, however, been attested in many old Germanic languages, e.g., Old English, Old High German, Old Icelandic, and Old Swedish (Rosenkvist 2009) and in various Germanic vernaculars, e.g., Frisian (Rosenkvist 2009), Övdalian (Rosenkvist 2010), Bavarian (Axel & Weiß 2011), Swabian (Beaman 2022; Bohnacker 2013; Haag-Merz 1996), Swiss German (Cooper & Engdahl 1989), and Yiddish (Gutman 2004). No German varieties appear to allow null subjects across the board; rather, they appear to be restricted to certain persons and numbers, a situation known as "partial pro-drop." Various explanations have been put forward for this, including the weakness of the verbal agreement paradigm, the non-uniformity of verbal inflections, the lack of topic identification structures, and the general incompatibility of null subjects with V2 sentence structure. Historical linguists commonly acknowledge that subject pronouns were often dropped in Old High German and Proto-Germanic. They have been considered "loan syntax" originating from "narrow or slavish" translations (Axel 2005), "syntactic innovations", and "remnants of archaic systems" (Rosenkvist 2009).

Axel (2005: 16) maintains that Old High German (OHG) was a partial null subject language with fully licensed null subject pronouns. She reports a wide variety of null subjects (a) in stressed and unstressed environments, (b) with non-specific references, such as *man* 'one' and *sie* 'they,' (c) in meteorological expressions, and (d) in even in environments where the subject pronoun is considered obligatory. She found null subjects in all person-number combinations, but most commonly, with third-person singular and plural forms (Axel 2007: 314). Abraham (1993: 117) documented the evolutionary stages of pronominal expression, showing how German has evolved from a partial to a non-null subject language – a change that occurred primarily during the Middle High German (MHG) period, from 1050 to 1350. Examples (10) to (12) are from Abraham (1993) and examples (13) and (14) are from the current Swabian corpus.

(10)Gothic:
amen auk Ø qiba izwis
wahrlich sage ich euch
'I tell you the truth'
[Matthew V; Braune & Ebbinghaus (1961: 138)]

(11)Old High German:
duo morgan Ø uuarth
als es Morgen wurde
'when it became morning'
[Mons. Xx111,21; Behaghel (1928)]

(12)Middle High German:
dô tete sî als ir Ø waere gâch
da tat sie, als wäre es ihr Eilig
'then she did it, as if it were urgent'
[Iw. 3612; Abraham (1993: 126)]

(13)Modern Standard German:
Ø ist ähnlich wie hier
es ist ähnlich wie hier
'It is similar to here'
[S016-17-I-1-Manni-00:58:57]

(14)Modern Swabian:
da warsch Ø e klôins Mädle
dann warst du ein kleines Mädchen
'then you were a small girl'
[S106-17-I-1-Mia-00:27:56]

Table 2: Null subjects in two Old Germanic varieties.

Person-Number	Old English (Berndt 1956: 65-68)	Old High German (Axel 2007: 315)
SINGULAR:		
First-person	3% (*n*=1629)	19% (*n*=571)
Second-person	9% (*n*=798)	39% (*n*=228)
Third-person	75% (*n*=3073)	56% (*n*=962)
PLURAL:		
First-person	1% (*n*=321)	32% (*n*=97)
Second-person	9% (*n*=1211)	16% (*n*=331)
Third-person	69% (*n*=1709)	60% (*n*=446)

Many historical linguists have quantitatively investigated null subjects in written texts of Old Germanic (Axel 2007; Berndt 1956; Rosenkvist 2009; Walkden 2013) and found varying frequencies of pronominal subject omission across all person-number combinations. Table 2 summarizes the frequency of null subjects from texts in Old English (Berndt 1956) and Old High German (Axel 2007). Both studies show that null subjects were most common in third-person forms, with first- and second-person forms considerably lower.

Contrary to formalist theories, Rosenkvist (2009: 159) argues that the presence or absence of a pronominal subject is not dependent on the richness of the inflectional system; in fact, he maintains that some of the highest frequencies of null subjects can be found with non-ambiguous forms. He conducted a meta-study comparing null subjects from written texts of four old Germanic varieties (i.e., Old High German, Old English, Old Icelandic, Old Swedish) with spoken texts of several modern German dialects (i.e., Bavarian, Swabian, Swiss German, Frisian, Övdalian, and Yiddish).[2] For Swabian, he claims that (a) person and number features of the null subject can always be reconstructed from the verbal agreement, (b) null subjects are not sensitive to the type of clause, specifically, main or subordinate clauses, (c) third-person null subjects are not possible, (d) overt subjects (which could be null) are understood to be emphatic or contrastive, and (e) null subjects are less frequent than overt subjects (Rosenkvist 2009: 173). In his analysis, he used examples from Haag-Merz's (1996)

dissertation, which she drew from an introspective analysis of her own native Swabian dialect. However, Haag-Merz (1996: 155) stresses herself that her own introspections are not always in line with other speakers' or researchers' intuitions. As the current paper will show, an empirical quantitative variationist analysis of spontaneously spoken Swabian calls each of Rosenkvist's (2009) and Haag-Merz's (1996) claims into question.

Bohnacker (2013) conducted the only known quantitative analysis of spoken Swabian with a corpus of six native Swabian speakers from the rural highlands near the city of Ulm. She found that speakers dropped second-person singular subject pronouns, clause internally (i.e., following the finite verb in main clauses and after the complementizer in subordinate clauses), 61% of the time (Bohnacker 2013: 258). While first- and third-person singular neuter subjects were also dropped clause internally, they occurred at much lower frequencies, 1% and 2.5%, respectively, and only in certain phonological and lexical environments. She maintained that the remaining constellations (i.e., third-person singular masculine and feminine, and all plural forms) can never be null in Swabian – a finding contrary to the results of current study.

3 Data and Methods

This section describes the Swabian corpus and speakers, along with a detailed account of the dependent and independent variables.

3.1 Corpus

Swabian German, or Schwäbisch, is an upper German dialect in southwestern Germany belonging to the Alemannic family and spoken by approximately 800,000 people (see Figure 1).

The data for this study are drawn from a corpus of sociolinguistic interviews (752,830 words) collected at two time periods, 1982 and 2017, and in two communities, Stuttgart and Schwäbisch Gmünd (see Beaman 2024 for details on data collection and corpus creation). Stuttgart, with its surrounding suburbs, is a large, international, multiethnic city, with over one million inhabitants, half of whom have at least one foreign-born parent (Auer 2020). In contrast, Schwäbisch Gmünd is a typical mid-sized, semi-rural town with about 60,000 inhabitants, surrounded by many rural villages with 77% of its land dedicated to woodland agriculture.

Figure 1: Division of German Dialects (Lameli 2019).

3.2 Speakers

The corpus comprises 127 native speakers of Swabian, socially stratified for age, gender and education (see Tab. 3). Speakers are divided into two age groups: younger speakers range in age from 18 to 28 (*mean*=22; *sd*=2.8); older speakers range in age from 30 to 88 (*mean*=58; *sd*=14.3. This binary split reflects the age at which most individuals have left home, entered the workplace, and started their own families. While education level is not shown in the table, 72 speakers (38 in Stuttgart and 34 in Schwäbisch Gmünd) have higher education (i.e., completion of the *Abitur*, 'German college preparatory exam') and 55 speakers (26 in Stuttgart and 29 in Schwäbisch Gmünd) have lower education. While a binary split for age and education may appear simplistic, the approach offers simplicity of interpretation, enhanced statistical validity (by avoiding cells with low token counts), and compatibility with other research (Erker et al. in review).

Table 3: Swabian corpus – speakers. Younger speakers are less than 30 years old, while older speakers are over 30.

	1982				2017				Total
	Older		Younger		Older		Younger		
Community	Women	Men	Women	Men	Women	Men	Women	Men	
Schwäbisch Gmünd	4	3	4	7	16	18	6	5	63
Stuttgart	4	1	7	9	14	11	9	9	64
Total by Gender	8	4	11	16	30	29	15	14	127
Total by Age	12		27		59		29		127

3.3 Envelope of variation

The dependent variable is the variable presence or absence of a subject pronoun with a finite verb. To ensure comparability with other studies of subject pronoun variation, I follow the coding conventions set by previous sociolinguistic variationist work (e.g., Erker et al. in review; Otheguy & Zentella 2012). Studies of SPE in Spanish typically exclude inanimates and discourse markers from the envelope of variation since they are invariant; however, in Swabian, these expressions are highly variable and thus are included. Examples (15) and (16) illustrate this variability.

(15)Inanimates:
Ø isch bissle abgedrosche (ref. Aussteige)
'it is a bit over used (ref. 'withdrawing from society')
[S036-82-I-1-Helmut-00:21:49]

(16)Discourse markers:
Ø weiß net 'I don't know'
[S010-82-I-1-Angela-00:23:55]

i weiß net 'I don't know'
[S010-82-I-1-Angela-00:30:12]

ja sag Ø mal so 'yes I say so'
[S036-17-I-1-Helmut-01:17:01]

also sag i mal so 'thus I say so'
[S036-17-I-1-Helmut-01:08:19]

glaub i net 'I don't think so'
[S011-82-I-2-Herbert-00:13:59]

glaub Ø net so 'I don't think so'
[S011-82-I-2-Herbert-00:17:46]

It is important to note that Swabian contains many proclitics, enclitics, and clitic clusters, which are counted as expressed pronouns; only if the pronoun is completely absent in the speech signal is it counted as absent, as shown in example (17).

(17)Clitics:
ob s geht 'whether it works'
[S074-17-I-1-Didrika-00:20:28]

ob Ø dir auch so geht 'whether you feel the same way' [S007-17-I-1-Egbert-00:33:23]

Several environments have been excluded from the envelope of variation due to their invariant nature, i.e., they show either extremely high (over 95%) or abnormally low (below 20%) levels of overt pronominal use:

- indefinite subjects (e.g., *ein* 'one', *kein* 'none', *jeder* 'everyone', *jemand* 'someone', *niemand* 'no one', *n*=416, SPE=98.3%)
- subject relative clauses (e.g., *der Mann, der da steht* 'the man who is standing there', *n*=2,910, SPE=96.3%)
- imperatives (e.g., *geh* 'go', *komm* 'come', *n*=194, SPE=19.4%)
- existentials (e.g., *es gibt* 'there is', *n*=1,225, SPE=94.8%)
- questions (e.g., *wann* 'when', *wer* 'who', *was* ´what', *wie* 'how', *wo* 'where', *n*=547, SPE=97.9%)
- frozen expressions (e.g., *was weiß ich* 'what do I know', *n*=121, SPE=100%)

On average, 300 referential pronominal subjects were manually coded for each speaker by native German-speaking research assistants at the University of

Table 4: Swabian corpus – tokens (finite verbs). Younger speakers are less than 30 years old, while older speakers are over 30.

	1982				2017				Total
	Older		Younger		Older		Younger		
Community	Women	Men	Women	Men	Women	Men	Women	Men	
Schwäbisch Gmünd	1,031	778	1,001	1,657	4,275	4,476	1,728	1,063	16,009
Stuttgart	947	247	1,810	2,243	3,659	2,587	2,266	2,336	16,095
Total by Gender	1,978	1,025	2,811	3,900	7,934	7,063	3,994	3,399	32,104

Tübingen. To provide compatibility across interview topics, annotation started roughly 20 minutes into the interview, beginning with the interview question "what is common sense" (for a list of the interview questions, see Beaman 2024). This created a corpus of 32,104 tokens of subject pronominal clauses, of which 27,166 contain pronouns (see Tab. 4), yielding an overall rate of SPE of 84.6%.

3.4 Conditioning factors

The conditioning factors or constraints (i.e., the independent variables) examined in this study have been selected from a wide range of research that has investigated subject pronoun use across a variety of languages and dimensions. I organize the constraints according to the tripartite typology proposed by Tamminga et al. (2016), which categorizes conditioning factors into three types:

- i-conditioning factors comprise internal structural and language-specific linguistic constraints, such as the person-number of the verb, clause type, tense-mood, etc.;
- p-conditioning factors involve the psychophysiological and cognitive constraints in the discourse, such as reference continuity, priming, and lexical frequency; and
- s-conditioning factors encompass the social and stylistic factors in the social setting, such as aspects of the speakers (e.g., age, gender, education) and the setting (e.g., casual, careful).

Table 5 describes the conditioning factors investigated in this study along with a list of the factor values for each.

Table 5: SPE constraints and factor values. The reference levels used for multivariate analysis are denoted by * before the factor value.

Predictor	**Description**	**Factor values**
i-conditioning factors		
PERSON-NUMBER	Person-number of the verb was coded following the standard German verbal paradigm; specific and non-specific third-person singular forms were coded separately (e.g., SG3 *er/sie/es* 'he/she/it' versus SG3[non] *man* 'one')	*1SG, 2SG, 3SG (specific), 3SG[non] (non-specific), 1PL, 2PL, and 3PL
CLAUSE TYPE	Verbs were coded for the type of clause in which they occur: coordinate clauses were identified by one of four conjunctions: *und* 'and', *aber* or *sondern* 'but', *oder ´or'*; subordinate clauses were identified by the presence of a complementizer (e.g., *dass* 'that', *obwohl* 'although', *weil* 'because', *wenn* 'if', *wie* 'how').	*main, coordinate, subordinate

Predictor	Description	Factor values
TENSE-MOOD	Finite verb forms were coded for their tense and mood and consolidated into three groups based on inflectional distinctiveness: the present (*Präsenz*) and future (*Futur*) were combined as 'present'; the perfect (*Perfekt*) and preterite (*Präteritum*) were combined as 'perfect'; and the conditional and subjunctive (*Konjunktiv II*) were combined as 'imperfect'	*present (*Präsenz*, *Futur*), preterite (*Präteritum*, *Perfekt)*, and imperfect (*Konjunktiv II*)
MORPHOLOGICAL REGULARITY	Finite verbs were coded for their morphological regularity, i.e., *stark* 'strong' (i.e., irregular) or *schwach* 'weak' (i.e., regular)	*strong, weak
SUBJECT ANIMACY	Verbs were coded for whether the referential subject was human/animate or non-human/ inanimate	*animate, inanimate
CLAUSE POLARITY	Clauses were coded for their polarity, based on the presence or absence of a negative particle (e.g., *net, nette, nicht*) or negative pronoun (e.g., *keine, keiner*)	*positive, negative
p-conditioning factors		
REFERENCE CONTINUITY	The pronominal subject of each verb was coded for the continuance of referent from the immediately preceding clause with a finite verb; also referred to as switch reference	*same, different
PRIMING	The pronominal subject of each verb was coded for priming with the pronoun in the immediately preceding clause of the same structural form; also called structural persistence	*present, absent
VERB FREQUENCY	Verbs were coded for lexical frequency based on the finite verb form and calculated locally from the Swabian corpus; two methods of verb frequency are use: binned as 'high' (top 1% of verbs) versus 'low' and scalar, which is log transformed to reduce skewness	* high, low; log-scaled
s-conditioning factors		
RECORDING YEAR	The two recording years are treated separately to investigate real-time change	*1982, 2017
COMMUNITY	The two communities are Stuttgart and Schwäbisch Gmünd	*Gmünd, Stuttgart
GENDER	Gender was coded based on the speakers' self-reported information gathered at the end of the interview	*men, women
AGE LEVEL	Age is binned into two groups: younger (<30 years old) and older (≥30 years old); z-scored age is used for multivariate analyses	*low, high
EDUCATION	Educational level was coded as "high" or "low" based on the completion of an *Abitur,* the 'German college preparatory exam'	*low, high

Predictor	Description	Factor values
GENRE	Utterances were coded for one of six groups using Labov's (2002) *Topic Decision Tree*; no tokens were observed for the categories of "response", "tangent", and "group" discourse	*language, residual, narrative, homeland, kids&games, soapbox
STYLE	Utterances were split into two groups, following Labov's (2002) *Topic Decision Tree*; "homeland", "kids&games", and "narrative" were classified as "casual"; "language", "soapbox", and "residual" speech were classified as "careful" speech	*casual, careful
SPEAKER DYAD	Interviews were coded for whether the interviewer and speaker were of the same or different genders	*same, different

4 Results

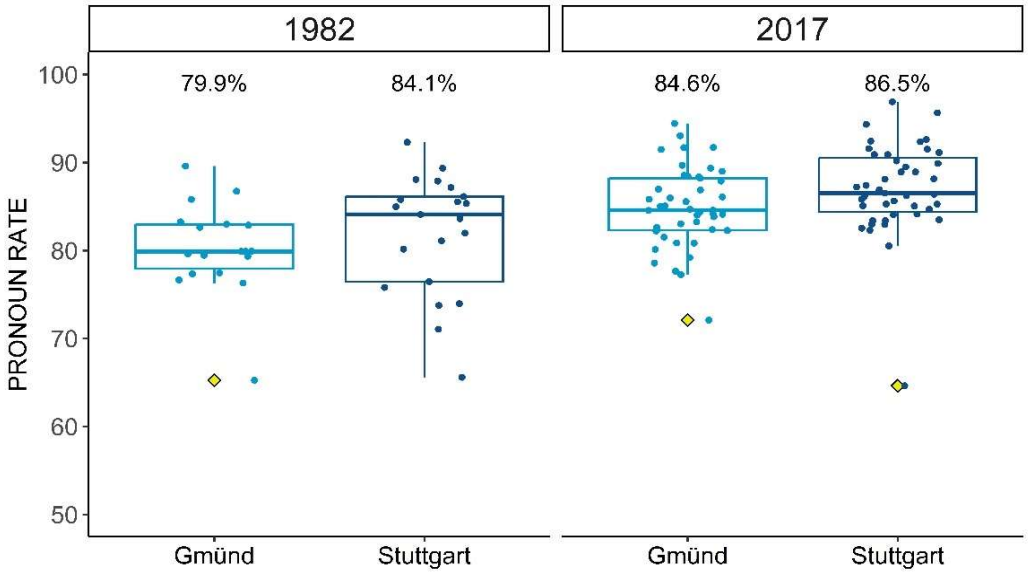
To triangulate the results, I use a variety of statistical approaches, covering univariate analyses for each predictor (see Section 4.1), mixed-effects multiple regression modeling (see Section 4.2.1), and clustering/classification using random forests with the *Boruta* algorithm (see Section 4.2.2).

4.1 Distributional analyses

A crucial component of this Swabian corpus is to what extent real-time change has occurred in pronominal subject use over the 35 years of this study. Figure 2 depicts the pronoun rate for each community by recording year. In 1982, pronoun use was lower in both Schwäbisch Gmünd (79.9%) and in Stuttgart (84.1%), however, this difference between the two communities is only marginally significant (β=0.013, p=0.07). By 2017, both communities show greater pronoun use, 84.6% in Schwäbisch Gmünd and 86.5% in Stuttgart, revealing real-time change toward the standard language (β=0.047, p<0.001 for Schwäbisch Gmünd; β=0.068, p<0.001 for Stuttgart).

Rather than classic dialect attrition, I attribute the change in pronoun use to the changing social situation in Swabia. Rising levels of education and increasing

Figure 2: SPE by community and recording year (n=32,104). Mean rates are presented at the top of each box; median rates are shown by solid lines in the box; individual speakers are represented by dots; the whiskers show the spread of the data, up to 1.5 times the interquartile range; outliers are indicated by yellow diamonds.



mobility over the 35 years of this study, have brought more speakers into greater contact with the standard language – both through increased contact with non-Swabian speakers but also the result of standard language ideologies reinforced in the school system (see Beaman 2024: Chapter 3, for a detailed discussion). Over the last 50 years, educational levels have more than doubled (Frietsch 2003: 38) and immigration has increased by 25%, with a significant influx in 2015, two years before the second data collection. Reports show that one-quarter (24.3%) of the German population has a history of immigration (Destatis, Statistisches Bundesamt 2006: 44). Stuttgart now has more than twice as many individuals with a migration background as in the rest of Germany (Auer 2020).

As a result of these socio-cultural changes, the relationships between the speakers in these communities have also evolved. In 1982, the speakers were more closely connected with "strong ties" (Milroy 1980), particularly in Schwäbisch Gmünd. By 2017, these ties had weakened and connections had become more dispersed, particularly in Stuttgart. Thus, while the interviews in 1982 were casual conversations among family and friends, by 2017, they had become formal interviews with casual or even distant acquaintances (Beaman 2024: Chapter 3). As Section 4.1.3 will show, pronoun omission is indicative of a more casual conversational style, while higher levels of pronoun use signal greater formality. This formality is reflected not only in the change across the years, but also in differences between the semi-rural setting of Schwäbisch Gmünd and the more impersonal environment of the large, urban metropolis of Stuttgart.

i-conditioning factors

To understand the constraints affecting subject pronoun use, I start with the i-conditioning factors (see Tab. 5 for definitions). Table 6 shows the token counts and mean rates of pronoun use for each factor level for each constraint. The significance of each constraint was determined via univariate linear regression models (function *lm*, package *stats*, version 4.3.0) without regard to recording period. All internal linguistic factors analyzed in this study reach significance, except for CLAUSE POLARITY, which will not be discussed further.

PERSON-NUMBER. The PERSON-NUMBER of the verb is one of the most influential constraints on subject pronoun variation across languages (Bohnacker 2013; Erker et al. in review; Otheguy & Zentella 2012). Table 6 shows real-time differences across all person-number combinations, with speakers using more pronouns in 2017 than they did in 1982. Also, as we saw in Figure 2, Stuttgarters use more pronouns than speakers from Schwäbisch Gmünd across all person-number combinations, except for third-person plural (3PL). Most notably, second-person singular forms (2SG) have a much lower level of overtly expressed subjects than other person-numbers, from a low of 36% and 37% in 1982 to 47% and 55% in 2017, Schwäbisch Gmund and Stuttgart, respectively (β=-0.445, p<0.001). This low rate of pronoun use can be attributed to the distinct verbal inflection in the second-person singular, which easily disambiguates it from the other forms, making it more salient (see example (5)). We also see that the second-person singular forms have changed the most over the 35 years of this study, reflecting the S-shaped curve of language change (Kroch 1989; Labov 1994), in which rates in the mid stages of change (between 36% and 65%) are accelerating while those at the beginning (<35%) or end (>65%) stages of change are moving more slowly (Nevalainen et al. 2011).

First-person (1SG) and the third-person singular forms (3SG and 3SG[non]) also show increased frequencies of pronominal use over the 35 years of this study; both are considerably higher than the rates reported for Old High German (see Tab. 2). This change is likely the result of increasing convergence to the standard language brought about rising levels of education and increasing formality, particularly in Stuttgart. As the corpus consists of one-on-one sociolinguistic interviews, the use of plural verbs was limited. Due to their low token counts and to potential ambiguities across forms, plural forms are excluded from further analysis.

Table 6: SPE frequency distributions – linguistic factors (*n*=32,104).

PERSON-NUMBER	**Tokens 1982 Gmünd**	**Tokens 1982 Stuttgart**	**Tokens 2017 Gmünd**	**Tokens 2017 Stuttgart**	**Rates 1982 Gmünd**	**Rates 1982 Stuttgart**	**Rates 2017 Gmünd**	**Rates 2017 Stuttgart**
1SG	1,555	1,856	4,341	4,361	86.8	87.9	90.7	91.5
2SG	330	393	927	595	35.8	36.6	47.0	54.6
3SG	1,366	1,783	3,430	2,874	78.0	81.4	84.6	85.8
3SG[non]	648	601	1,235	1,583	86.9	89.4	90.3	90.8
1PL	175	199	668	593	90.0	93.0	95.5	94.9
2PL	10	11	46	41	80.0	90.9	84.0	75.6
3PL	383	404	895	801	86.2	81.7	84.8	81.6
					Intercept = 1SG; Estimate (2SG)=-0.445; p<0.001; Adjusted R^2=0.097			
CLAUSE TYPE								
subordinate	839	1,048	1,888	1,756	92.3	94.9	94.6	96.2
main	2,922	3,346	7,452	7,094	81.5	82.5	86.8	89.7
coordinate	687	824	2,117	1,944	63.6	64.4	73.9	72.5
					Intercept = main; Estimate (subordinate)=0.24; p<0.001; Adjusted R^2=0.042			
TENSE-MOOD								
imperfect	217	250	543	423	86.6	90.4	91.9	92.4
preterite	191	298	690	653	91.1	87.9	91.7	90.4
present	4,029	4,657	10,271	9,750	79.6	80.9	84.3	86.8
					Intercept = present; Estimate (imperfect)=0.071; p<0.001; Adjusted R^2=0.003			
MORPHOLOGICAL REGULARITY								
strong	3,395	4,148	8,907	8,172	81.4	83.8	86.2	88.2
weak	1,072	1,099	2,635	2,676	77.4	74.2	81.4	84.3
					Intercept = strong; Estimate (weak)=-0.05; p<0.001; Adjusted R^2=0.003			
SUBJECT ANIMACY								
inanimate	740	928	2,046	2,001	84.5	87.2	86.6	88.7
animate	3,708	4,290	9,411	8,793	80.0	81.1	85.5	87.4
					Intercept = animate; Estimate (inanimate)=0.031; p>0.001; Adjusted R^2=0.001			
POLARITY								
positive	3,841	4,430	10,035	9,271	80.2	82.4	85.8	88.0
negative	607	788	1,422	1,523	83.9	81.0	85.0	85.4
					Intercept = positive; Estimate (negative)=-0.009; p=n.s.; Adjusted R^2=0			

CLAUSE TYPE. A second powerful effect on pronoun use is CLAUSE TYPE, specifically whether the pronominal subject occurs in a main clause, coordinated clause, or subordinate clause (Nagy 2015; Otheguy & Zentella 2012; Torres Cacoullos & Travis 2019). In this study, I consider only syntactic linking or "syndetic coordination," i.e., the presence of a conjunction, and not prosodic linking or "asyndetic coordination" (Torres Cacoullos & Travis 2019). As Table 6 shows, in Swabian, as in many other languages, the highest frequency of pronouns is found in subordinate clauses (90% range) (β=0.24, p<0.001), followed by main clauses (80% range), and then coordinated clauses (70% range).

TENSE-MOOD. The third i-conditioning factor in this study is the TENSE-MOOD of the verb. This constraint has been widely researched in the Romance languages Spanish and Portuguese (Erker et al. in review), where, due to functional differences in the verbal paradigm, the lowest rate of pronoun use is in the preterite, followed by the present and then imperfect. Swabian also shows a significant effect of verbal inflection; however, the lowest rate of pronoun use is with the present tense, followed by the preterite (β=0.662, p<0.001) with the imperfect showing the highest rates of pronouns use (β=0.071, p<0.001). This difference is likely due to the differing inflectional paradigms and ambiguities that exist across these forms (see Tab. 1).

MORPHOLOGICAL REGULARITY. The fourth *i*-constraint I evaluate is the MORPHOLOGICAL REGULARITY of the verb. German verbs are typically classified into three categories: weak (*schwache*) verbs follow a regular pattern of conjugation; strong (*starke*) verbs have irregular conjugations; and, mixed (*gemischte*) verbs are also irregular but with mixed conjugations. While there are only around 200 strong verbs in German, they are by far more frequent, making up 77% of the verbs in the Swabian corpus; thus, strong collinearity is expected with verb frequency. Table 6 shows the distribution of strong and weak verbs in the corpus, revealing that the irregular strong verbs show higher rates of pronoun use across both communities and time periods (Schwäbisch Gmünd 1982: 81.4% versus 77.4%; Stuttgart 1982: 83.8% versus 74.2%; Schwäbisch Gmünd 2017: 86.2% versus 81.4%; Stuttgart 2017: 88.2% versus 84.3%) (β=−0.05, p<0.001). It is noteworthy to mention that these results are the reverse of what Erker and Guy (2012) found in Spanish, where regular verb forms show greater use of overt pronouns. This difference is likely explained by the differing inflectional paradigms in the two languages.

SUBJECT ANIMACY. The fifth internal linguistic factor I analyze in this study is the ANIMACY OF THE SUBJECT. Studies of subject pronoun variation in Spanish have typically eliminated inanimate subjects because such references occur very rarely with a pronoun (Otheguy & Zentella 2012). However, in Swabian, pronominal use with animate and inanimate subjects is variable. Animate subjects account for 82% (n=26,389) of the tokens in the Swabian corpus, while inanimates comprise 18% (n=5,715) (see Tab. 6). The overall mean rate of pronoun use for animate subjects is 84%, while inanimate subjects show a mean rate of 87% (β=0.031, p<0.001), revealing that expressed subjects are more common with inanimate subjects.

p-conditioning factors

I consider three p-conditioning factors in this study: REFERENTIAL CONTINUITY, PRIMING, and VERB FREQUENCY. These cognitive-discursive factors, in particular priming and the temporal proximity of a variant, have been shown to have the most powerful effect on subject pronoun use (Tamminga et al. 2016), surpassing the internal linguistic and external social factors (see also Erker et al. in review). Table 7 shows the token counts and mean rates of pronoun use for each factor level (see Tab. 5 for definitions).

REFERENCE CONTINUITY. Numerous studies show that SPE is particularly sensitive to maintaining the continuity of reference across utterances. In almost all studies of subject pronoun variation, verbs that exhibit a change in referent from the prior clause show higher rates of pronoun use than verbs that maintain the same referent as the preceding clause (Erker & Guy 2012; Otheguy & Zentella 2012). Swabian also adheres to this pattern (β=0.139, p<0.001). Of the 31,104 tokens in the corpus, 61% (n=15,376) show a change in reference, while 39% (n=9,977) maintain the referent. Overall, the same referent as the prior clause yields a mean rate of pronoun use of 75%, while a switch in reference shows a mean rate of 88%. Table 7 provides the breakdown of pronoun rates by community and recording year.

Table 7: SPE frequency distributions – cognitive factors (*n*=25,353).

	Tokens				Rates			
	1982		2017		1982		2017	
	Gmünd	Stuttgart	Gmünd	Stuttgart	Gmünd	Stuttgart	Gmünd	Stuttgart
REFERENT CONTINUITY								
different	2,307	2,554	5,529	4,986	85.3	85.4	89.6	90.9
same	1,425	1,668	3,583	3,301	68.4	72.7	75.6	77.6
					Intercept=same; Estimate (different)=0.139; p<0.001; Adjusted R^2=0.033			
PRIMING								
present	3,087	3,696	8,363	8,082	82.4	82.4	86.2	87.9
absent	597	678	1,234	1,004	66.2	74.8	73.3	75.3
					Intercept=absent; Estimate (present)=0.128; p<0.001; Adjusted R^2=0.014			
VERB FREQUENCY								
high	2,581	3,165	6,835	6,642	81.4	84.4	87.2	89.4
low	1,886	2,082	4,707	4,206	79.2	77.7	82.1	83.7
					Intercept=low; Estimate (high)=0.052; p<0.001; Adjusted R^2=0.005			

PRIMING. Tamminga et al. (2016: 311) describe structural priming as "a preference for using a recently processed syntactic structure to form a novel utterance in cases with multiple syntactic options available." This implies that speakers tend to use the same form as the most previously used form (Bock 1986; Travis & Torres Cacoullos 2012), a tendency that has been found in most previous studies of SPE, as well as in the current corpus of Swabian (β=0.128, p<0.001). Of the total tokens in the Swabian corpus with a priming context, 87% (n=23,228) are primed by the presence of a pronoun in the previous utterance, and 13% (n=3513) have pronoun absence in the priming position (tokens with no priming context were eliminated, n=5363). Across the entire corpus, a present pronoun in the prior clause shows an overall rate of pronoun use of 86%, while utterances with prior pronoun absence show a mean rate of 73%. Table 7 shows the differences in rates by community and recording year.

VERB FREQUENCY. Previous quantitative studies of the effect of lexical frequency on linguistic variables have produced conflicting findings. Bybee (2017) maintains that change spreads first in high-frequency words through processes of analogy, while Hay et al. (2015) claim the opposite, arguing that change moves slower through high-frequency words due to their greater entrenchment in the mental lexicon. Tomaschek et al. (2018) argue that high-frequency words "get more practice," which makes them more resistant to change. Erker and Guy (2012) find that a non-monotonic relationship between pronoun use and verb frequency, such that more frequent forms are more differentiated in their use of pronouns, while less frequent forms are more congruent. Table 7 shows the results of the univariate analysis based on a binary split in verb frequency (i.e., the top 1% of verbs are considered "high" frequency verbs). Based on this binary distinction, across both time periods and communities, high-frequency verbs exhibit higher frequencies of pronoun use in Swabian (β=0.052, p<0.001).

s-conditioning factors

I investigate six s-conditioning factors in this study (see Tab. 5 for definitions). As discussed in Section 4.1, Figure 2 shows that there are significant differences in pronoun use for RECORDING YEAR and COMMUNITY: SPE has increased over the 35-year timespan of this study and is

higher in the urban center of Stuttgart. This section addresses the remaining social factors and their influence on pronominal use. Table 8 presents the frequency distributions for the six s-conditioning factors, by community and recording period.

SPEAKER GENDER. Few sociolinguistic studies in varieties of German have shown an effect of speaker GENDER on language use (Auer 2020; Beaman 2024); however, GENDER appears to be an evolving predictor of subject pronoun use. As Table 8 shows, in 1982, men and

Table 8: SPE Frequency distributions – social factors (*n*=32,104).

	Tokens				**Rates**			
	1982		**2017**		**1982**		**2017**	
	Gmünd	Stuttgart	Gmünd	Stuttgart	Gmünd	Stuttgart	Gmünd	Stuttgart
GENDER								
women	2,032	2,757	6,003	5,925	81.3	81.2	85.9	89.3
men	2,435	2,490	5,539	4,923	79.7	82.4	84.2	84.7
	Intercept=1982 men; Estimate (2017 women)=0.029; p<0.01; Adjusted R^2=0.005							
AGE LEVEL								
older	1,809	1,194	8,751	6,246	82.8	84.3	85.1	87.7
younger	2,658	4,053	2,791	4,602	78.9	81.0	85.2	86.5
	Intercept=1982 older; Estimate (2017 younger)=0.031; p<0.01; Adjusted R^2=0.004							
EDUCATION								
high	2,180	3,209	6,354	6,244	80.2	81.9	86.2	87.9
low	2,287	2,038	5,188	4,604	80.6	81.6	83.8	86.2
	Intercept=1982 high; Estimate (2017 low)=-0.02; p<0.05; Adjusted R^2=0.005							
GENRE								
language	1,490	919	2,995	2,813	84.5	86.8	87.0	89.3
narrative	82	116	286	362	73.2	81.9	85.0	89.2
residual	229	332	433	63	79.0	86.4	86.6	88.9
soapbox	742	887	3,324	3,596	77.8	79.6	84.4	86.8
homeland	1,082	1,559	3,387	3,313	78.5	82.7	85.2	86.2
kids&games	842	1,434	1,117	701	79.2	77.8	81.3	84.0
	Intercept=language; Estimate (kids&games)=-0.072; p<0.001; Adjusted R^2=0.003							
STYLE								
careful	2,461	2,138	6,752	6,472	82.0	83.8	85.7	87.9
casual	2,006	3,109	4,790	4,376	78.6	80.4	84.3	86.1
	Intercept=careful; Estimate (casual)=-0.026; p<0.001; Adjusted R^2=0.001							
SPEAKER DYAD								
same	2,191	2,120	6,255	5,704	79.6	81.7	84.5	88.9
different	2,276	3,127	5,287	5,144	81.2	81.8	85.9	85.3
	Intercept=different; Estimate (same)=0.008; p<0.05; Adjusted R^2=0							

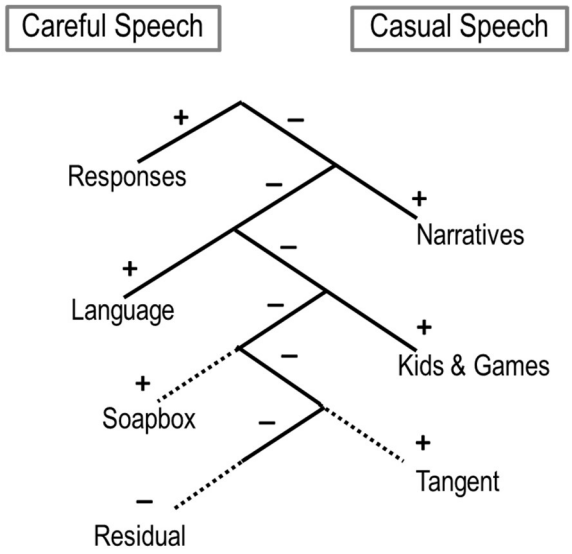
women were similar in their rates of subject pronoun use. However, by 2017, a significant difference has emerged with women in Stuttgart using more subject pronouns than the men in Stuttgart (β=0.029, p<0.001). Interestingly, this effect is not seen in Schwäbisch Gmünd. This change suggests that social stratification is expanding in Germany, particularly in the urban metropolis of Stuttgart, brought about by the changes in the socio-cultural landscape which has been evolving over the last 35 years (e.g., rising education, immigration, geographic mobility).

SPEAKER AGE. As Figure 2 showed, there has been considerable real-time change in expressed pronominal subjects over the 35 years; apparent-time change across the generations shows older speakers using more pronouns than younger speakers. But younger speakers across both time periods and communities use fewer pronouns (β=0.031, p<0.01, Tab. 8), indicating a process of age-grading rather than of community change (Hockett 1950; Wagner 2012). I surmise that the reasons for this lie in the more casual, informal style that younger speakers adopt vis-à-vis older speakers, who have become well-established in their professional lives (cf. *linguistic market* (Bourdieu 1977)). Standard language pressures, characteristic of the workplace, are likely driving the higher rates of pronoun use with the older speakers. The following section on genre and style dissects this formality prediction further.

SPEAKER EDUCATION. Higher levels of EDUCATION generally correlate with lower levels of nonstandard, dialect use, particularly for Swabian (Beaman 2024). Prestige and prescriptivism in the schools largely explain this phenomenon in that institutions of higher education (e.g., university) also promote standard language norms (e.g., Ammon 2001 for Germany; Vergeiner 2021 for Austria). Subject pronoun use in Swabian shows this same tendency, albeit slight, revealing that speakers with higher levels of education use more pronouns (β=-0.02, p<0.05, Tab. 8).

GENRE AND STYLE. Prior research shows that the topic of conversation can play a large role in stylistic variation (Coupland 1980; Labov 2002). When talking about familiar or emotional topics, such as the "homeland" or "danger of death" situations, speakers tend to revert to the vernacular; however, when they contemplate unfamiliar topics, such as the question, "what is common sense?", they tend toward more careful speech patterns. Labov's (2002) *Topic Decision Tree* (see Fig. 3) provides an effective heuristic for categorizing utterances into casual and careful speech styles. Each utterance in the Swabian corpus was coded for one of 10 different topics, and then grouped into two categories, casual and careful speech according to the *Topic Decision Tree*.

Figure 3. Labov's Topic Decision Tree (adapted from Labov 2002).



The results, as expected, show greater pronoun use with topics related to "language" and the lowest pronoun use in topics about "kids and games" (β=−0.72, p<0.001, Tab. 8). This effect was stronger in 1982 and has weakened somewhat over time. Overall, more careful styles promote more pronouns (β=−0.026, p<0.001, Tab. 8). These tendencies and the trend are depicted in Figures 4a (1982) and Figure 4b (2017). One noticeable difference is the placement of "narratives" across the two recording periods. In the 1982 recordings, the narratives were considerably more informal than they were in 2017, revealing a full 10% difference in pronoun use. Again, I attribute this to the more formal nature of the 2017 interviews, which were conducted between casual acquaintances, unlike the 1982 interviews, which were carried out between close family and friends (see Beaman 2024 for an ethnographic description of the two communities).

Figure 4a: SPE and style shifting (1982) (*n*=9,714).

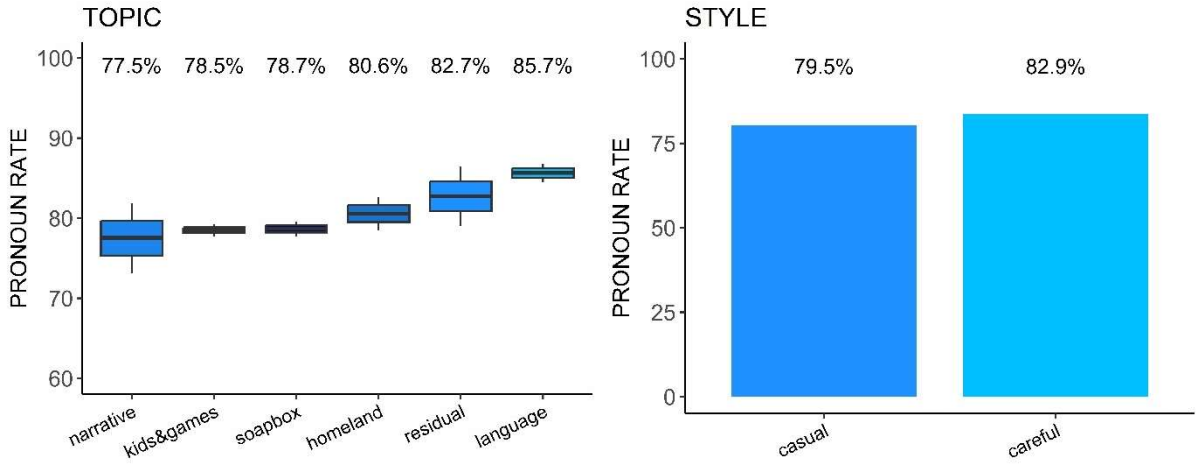
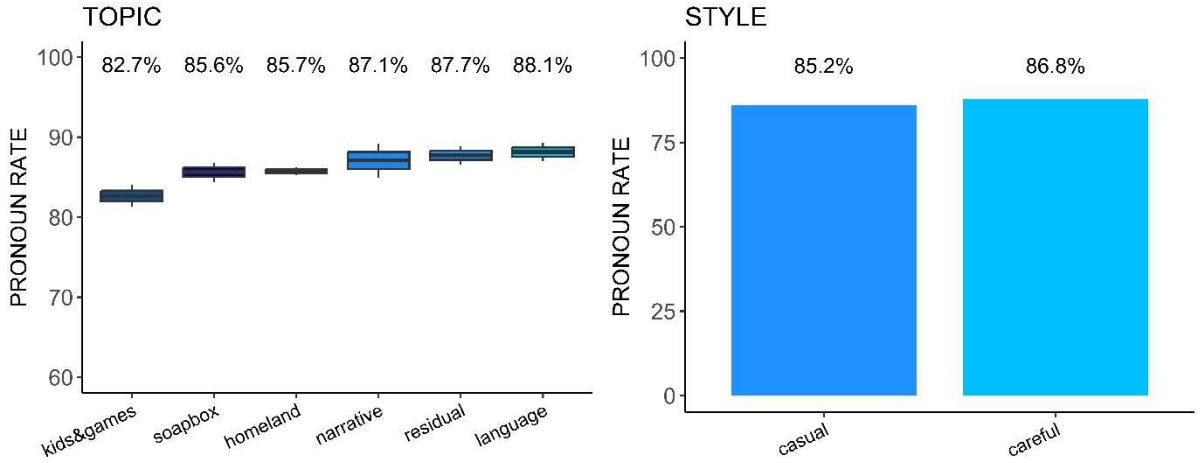


Figure 4b: SPE and style shifting (2017) (*n*=22,390).



SPEAKER DYAD. The last s-conditioning factor I consider in this paper is speaker dyad, the effect of the relationship between the interviewer and interviewee. Prior studies show that the power relationship in the interview can affect speakers' use of linguistic variants (Gregersen et al. 2018; Trudgill 1981). Table 8 shows a slight, but significant, tendency for same-gender interviews to produce lower rates of pronoun use, except for Stuttgart in 2017 (β=0.008, p<0.05, Tab. 8). I attribute this difference to the closer relationships and more casual interview situations in 1982 and in Schwäbisch Gmünd than in 2017 and in Stuttgart.

4.2 Multivariate Modeling

To make sense of this profusion of conditioning factors, I now turn to multivariate modeling. I explore two different methods: generalized mixed-effects regression analysis (Section 4.2.1) and clustering/classification with random forests using the *Boruta* algorithm (Section 4.2.2).

Regression Analysis

The gold standard in sociolinguistics for analyzing the significance of constraint effects and controlling for intraspeaker variances is generalized mixed-effects regression modeling (function *glmer*, package *lme4,* version 1.1-33) (Baayen et al. 2008). Mixed-effects regression modeling incorporates both fixed and random effects, providing reliable and generalizable estimates across different levels of variability, even in the face of unbalanced datasets (e.g., Peduzzi et al. 1996). Separate models were built for the two recording periods, 1982 and 2017, to allow for a real-time comparison. Tokens for regression analysis were pruned to remove

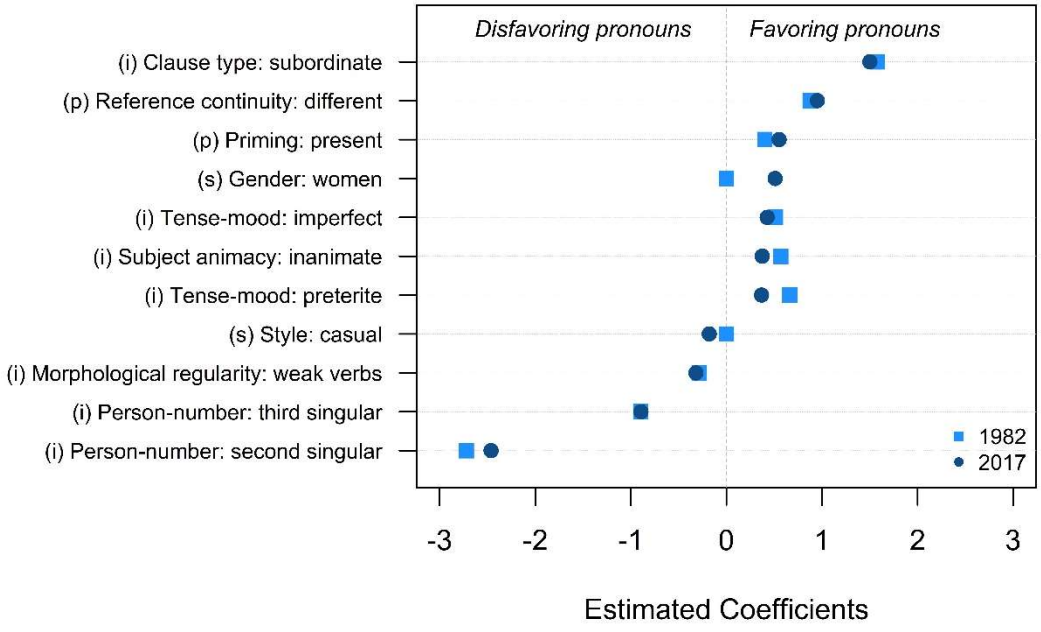
insignificant factors (e.g., third-person singular non-specific *man* 'one', *n*=4067), factors with low token counts (e.g., plural forms, *n*=4226), utterances with no referential or priming context, and "other" clause types and verb forms, reducing the total token count to 16,431 (1982: *n*=5,135, speakers=39; 2017: *n*=11,296, speakers=88). Age (neither binned nor scaled) reached significance in either recording period and was also removed. Table 9 and Figure 5 show the results of the two best-fit regression models (1982 and 2017) for subject pronoun use in Swabian based on AIC (Akaike Information Criterion), a commonly used metric for model selection, balancing goodness of fit and model simplicity (Burnham & Anderson 2004).

Table 9 and Figure 5 list the constraints in descending order, showing the environments most favoring pronoun use to those least favoring, based on the 2017 estimated coefficient. At the top, the most powerful constraints are subordinate clauses with a change in reference that are primed with a previous pronoun. This finding confirms the powerful role that linguistic and cognitive constraints play in sociogrammatical variation

Table 9: SPE estimated coefficients (*n*=16,431). Reference level values are listed in Table 5; constraints presented in descending order by 2017.

Constraint	**1982**	**2017**
(i) Clause type: subordinate	1.577 ***	1.500 ***
(p) Reference continuity: different	0.876 ***	0.951 ***
(p) Priming: present	0.402 ***	0.551 ***
(s) Gender: women	-0.067 n.s	0.510 ***
(i) Tense-mood: imperfect	0.513 *	0.427 **
(i) Subject animacy: inanimate	0.568 ***	0.375 ***
(i) Tense-mood: preterite	0.665 ***	0.368 **
(s) Style: casual	0.018 n.s	-0.180 **
(i) Morphological regularity: weak verbs	-0.285 **	-0.318 ***
(i) Person-number: third singular	-0.897 ***	-0.894 ***
(i) Person-number: second singular	-2.721 ***	-2.462 ***

Figure 5: SPE estimated coefficients. Reference level values are listed in Table 5; constraints presented in descending order by 2017.



and change. At the bottom, the strongest constraints disfavoring pronoun use are second- and third-person singular verb forms. It is not surprising to see pronoun omission favored in these environments given the unambiguous inflections with regular second- and third-person singular verb forms in Swabian. Also not surprising, none of the linguistic or cognitive constraints show more than modest real-time change between 1982 and 2017. This confirms much other sociolinguistic research which consistently shows that "the constraints on a variable are shared across members of a speech community, despite inter-speaker differences in the rate of use of the alternating variants'' (MacKenzie 2019: 4). This study shows that this principle also holds true for SPE in Swabian in real-time. These results support that the p- and i-conditioning factors exert the strongest effects on subject pronoun use, while the s-conditioning factors exhibit the weakest influence (see Erker et al. in review for further discussion).

However, the two s-conditioning factors – gender and style – show a noteworthy trend: neither of these social factors were significant in 1982, yet both have become significant in 2017. The effect of gender shows that, in 2017, women have begun using more pronouns than men (β=0.510, p<0.001), 88% (women) to 84% (men), increasing to 89% for women in Stuttgart. This supports the previously mentioned supposition of changing gender norms in Swabian society. Style has also emerged as a significant predictor in 2017: casual conversational topics reflect fewer pronouns (β=−0.180, p<0.01), particularly by younger speakers (β=−0.021, p<0.05). Recall that the univariate results show that younger speakers use fewer overt pronouns (see Tab. 8), suggesting the SPE is an age-graded variable. This leads to the speculation that younger speakers may be trying to differentiate themselves from the formal, structured school system and world of work by "dropping" pronouns to sound cool, friendly, and nonchalant.

Classification

To assist in triangulating and validating the constraints on subject pronoun use in Swabian, I leverage an alternative multivariate method: random forests with the *Boruta* algorithm. Random forests are an ensemble learning method for classification and regression that models complex, non-linear interactions with high levels of accuracy (Tagliamonte & Baayen 2012). The *Boruta* algorithm (function *Boruta,* package *Boruta,* version 8.0.0) (Kursa & Rudnicki 2010) serves as a wrapper around a random forest calculation. It works by making copies of the original predictors, called shadow predictors*,* shuffling their values to break the relationships, and then iteratively comparing the original features with the randomly generated shadow predictors, dropping those that are deemed less important, until all predictors are classified as either important, unimportant, or undecided. This yields a measure for each predictor of their importance to the model, in effect, an estimate of how much the model is improved by that predictor. *Boruta* is a complementary method to multiple regression modeling which helps with initial feature selection by reducing the dimensionality of the data and thus simplifying the model to focus only on the most important features. It works effectively with scant datasets, non-normal distributions, unbalanced samples, and multicollinearity, all of which are problematic areas in regression modeling. *Boruta* also helps to avoid overfitting by excluding features that do not meaningfully contribute to the model.

Figures 6a and 6b display the *Boruta* results for subject pronoun use in Swabian, separate *Boruta* runs for each recording period, 1982 and 2017. For both periods, person-number and referential continuity are the strongest constraints, followed by clause type and priming, demonstrating the powerful effect of linguistic and cognitive constraints on grammatical variation. The ranking of the constraints is almost identical for the two time periods, showing again that, while rates of use may change over time, constraint systems are more stable across the community (MacKenzie 2019: 4) and time. Figures 6a and 6b confirm the findings from the regression modeling that gender was not found to be a significant constraint on pronoun use in 1982 but has become an important influence in 2017.

Figure 6a: SPE constraint weightings – 1982 (*n*=5,135).

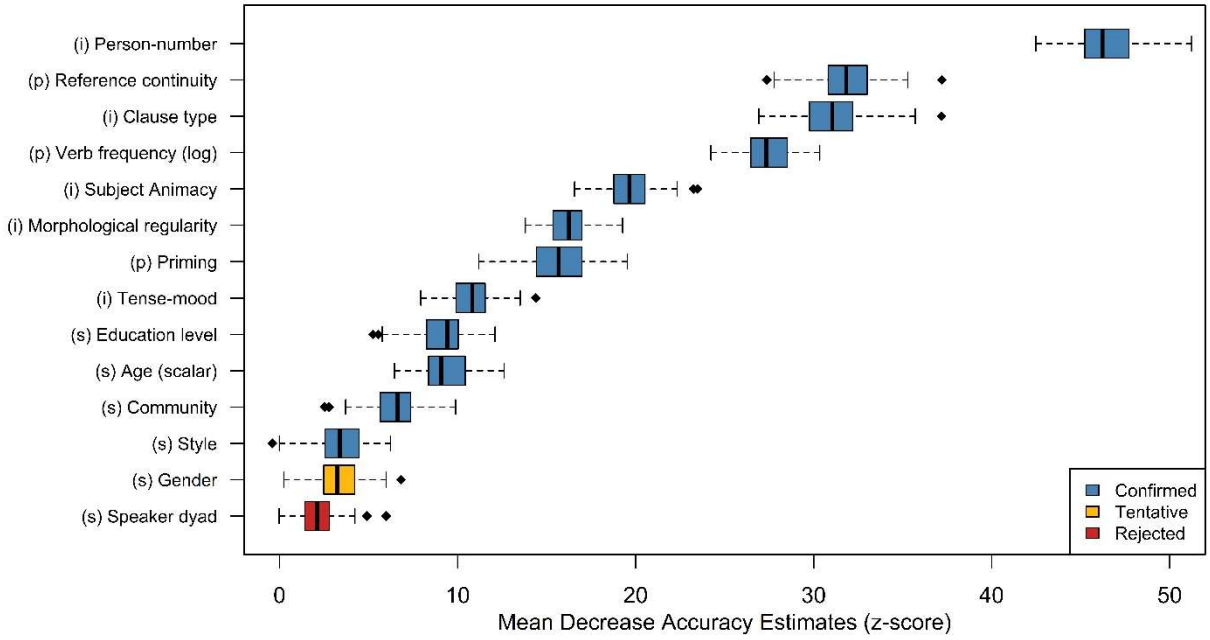
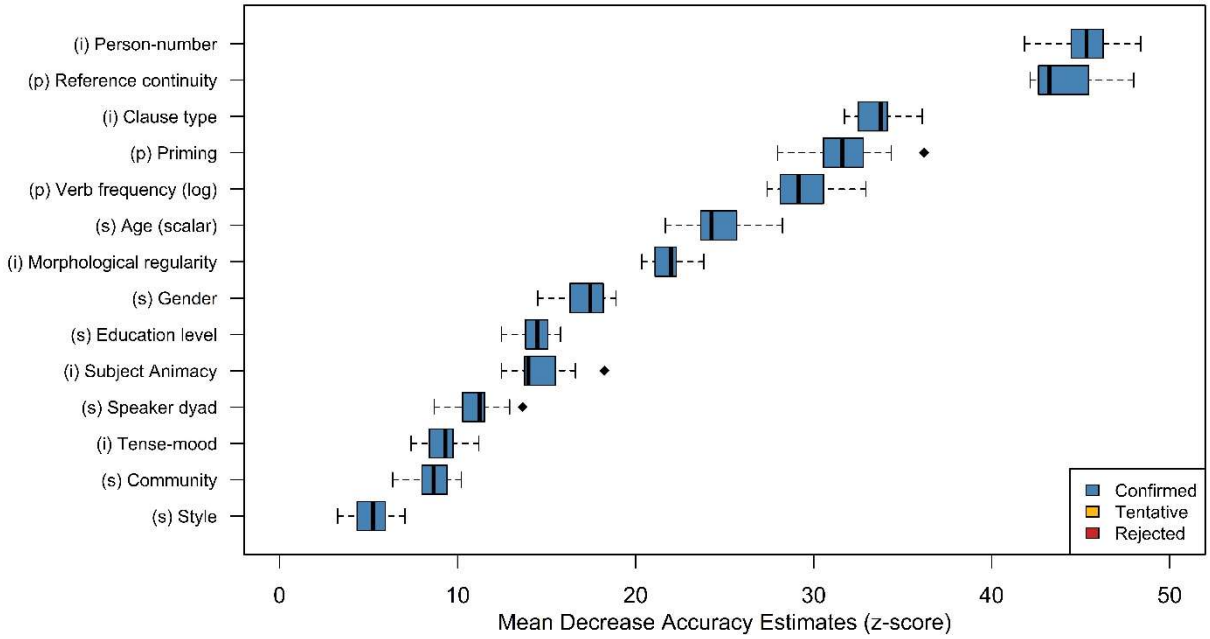


Figure 6b: SPE constraint weightings – 2017 (*n*=11,296).



Both the regression models in Figure 5 and the *Boruta* models in Figure 6 were run on the same pruned dataset (*n*=16,431). While the results are similar, the *Boruta* models provide some additional insights not easily seen in the regression models. First, the effect of gender has visually changed in 2017, moving up in the constraint hierarchy revealing its growth in importance. Table 10 lists the *Boruta* importance measures for each recording period, showing that two social factors have changed considerably between 1982 and 2017: gender (from 3.3 to 17.2), followed by speaker dyad (from 2.2 to 11.0). The only other constraint that has increased more is priming (from 15.7 to 31.7). Second, because *Boruta* expands the database using shadow predictors, permutating hundreds of conditional inference trees, it can handle a larger set of predictors, in this case, age

Table 10: SPE importance measures (*n*=16,431). Constraints presented in descending order by 2017.

Constraint	1982	2017
(i) Person-number	46.415	45.279
(p) Reference continuity	31.914	44.082
(i) Clause type	31.034	33.622
(p) Priming	15.655	31.717
(p) Verb frequency (log)	27.450	29.549
(s) Age (scalar)	9.385	24.595
(i) Morphological regularity	16.276	21.780
(s) Gender	3.305	17.214
(i) Subject animacy	19.742	14.664
(s) Education level	9.163	14.396
(s) Speaker dyad	2.156	11.020
(i) Tense-mood	10.796	9.143
(s) Community	6.482	8.573
(s) Style	3.491	5.223

(scalar), verb frequency (logarithm), community, and education level, all of which failed to converge in the fully saturated regression models. In sum, regression models and classification provide different perspectives on the sociolinguistic situation: multiple regression analysis assumes linear relationships, testing each individual predictor's contribution to the model; *Boruta* handles non-linear relationships, evaluating each predictor's importance to a randomly generated set of shadow predictors. These different approaches can lead to differences in model outputs, which, when analyzed together, provide broader insight into the sociolinguistic patterns constraining variation.

5 Discussion

The results of this study confirm the three hypotheses laid out in the beginning of the paper. Subject pronoun expression in Swabian: (1) follows similar internal linguistic constraints found in other language varieties, the structural uniformity hypothesis; (2) conforms to common cognitive and psychophysiological constraints which appear to be uniform across languages, the cognitive universal hypothesis; and, (3) is more frequent in situations of higher social prestige, such as by women, with higher levels of education, in more urban settings, and with more formal topics of conversation, the social sensitivity hypothesis.

Five of the six i-conditioning factors examined in this study – person-number, clause type, tense-mood, morphological regularity, and subject animacy – validate the structural uniformity hypothesis, substantiating the overwhelming strength of internal linguistic factors on sociogrammatical variation. While the details of the pronoun rates for these factors have changed over the 35 years of this study, the constraint ranking has remained largely the same: person-number and clause type are the strongest constraints, followed by morphological regularity, subject animacy and tense-mood. This finding supports the general linguistic principle of variability in rates and the stability in constraint hierarchies.

All three p-conditioning factors considered in this analysis – referential continuity, priming, and verb frequency – affect pronominal use, with reference continuity being the strongest, validating the cognitive universal hypothesis for this constraint. While priming is also a significant discursive constraint in Swabian, it appears to be less important than in other languages, such as Portuguese and Chinese (see Erker et al. in review), which I attribute to the outsized effect of person-number in Swabian. The role of verb frequency remains in question. While the univariate and *Boruta* results show that high-frequency verbs favor pronouns, this could not be confirmed in the regression analysis. Further

analysis is needed to uncover the role of lexical frequency in subject pronoun use.

All seven s-conditioning factors analyzed in Swabian – recording year, community, age, gender, education, style, and speaker dyad – have an important role to play in subject pronoun use, albeit to a weaker degree than the i- and p-conditioning factors, revealing that subject pronoun use is also governed by the social sensitivity hypothesis. Pronoun omission is higher in situations of lower sensitivity to social prestige and formalness, by men, by younger speakers, and in the semi-rural community of Schwäbisch Gmünd. In addition, the earlier recordings show evidence of age-grading, whereby younger speakers, both men and women, drop more pronouns and then increase their pronoun use as they age. By the later recordings, however, we see greater differentiation between the older men and women, revealing the evolution of what was a stable, age-graded variable to a gender-stratified variable, led by women and driven by change-from-above.[3]

While the linguistic findings confirm the relative consistency and stability of constraint effects over time, the social findings expose changing indexicalities resulting from societal change in progress (Eckert 2019; Silverstein 2003). Most striking, the results suggest that Swabia is becoming a more gender-stratified society. In 2017, women from both Schwäbisch Gmünd and Stuttgart favor the presence of pronominal subjects, an effect which is even stronger in Stuttgart. This behavior supports considerable established sociolinguistic research that women and large urban centers are the leaders of language change, which in Swabia, means convergence to the standard language (Beaman 2024). It is worth pointing out that, in prior investigations of both phonological and morphosyntactic variables, Swabian shows no gender-based differences (Auer 2020; Beaman 2021, 2024), although, a gender effect has been found with a few socio-grammatical and lexical variables (e.g., *wo*-relatives Beaman (2021) and intensifiers Stratton & Beaman (2024)).

German society has always been situationally stratified. Conversations with strangers and acquaintances promote the use of the formal, third-person plural *Sie* verb forms, while situations among family and friends expect the informal, second-person singular *Du* verb forms. Almost all interviews in the Swabian corpus were conducted using the *Du* pronoun, with the aim of bringing a more informal tone to the relatively formal interview situation. Because the sociolinguistic interview does not entail a change in interlocutor, the actual use of formal and informal politeness forms cannot be tested in this corpus. However, from a formality perspective, the data show that the use of null subjects is greater with more casual conversational topics (e.g., homeland, family, friends), while the overt use of subject pronouns is more common with formal topics (e.g., language, soapbox).

6 Conclusion

This study set out to examine variable subject pronoun expression in a language traditionally considered a "non-null subject language." The findings reveal a wide range of variability, along with real-time change, constrained by various conditioning factors: linguistic-structural, cognitive-discursive, and social-stylistic. The real-time increase in pronominal subjects over the 35-years of this study confirms considerable other research that pervasive dialect leveling and convergence to the standard language is occurring in Swabian (Beaman 2024). Increasing mobility, escalating immigration, and rising education are bringing more Swabians into greater contact with non-Swabians, increasing the formality of communicative situation. Importantly, these findings of this study contribute to our theoretical understanding on how change and increased exposure to the standard language affect linguistic and social conditioning. The results clearly demonstrate how language change reflects societal change, particularly the emerging socioindexicalities in Swabian with respect to age-grading and gender-stratification. Finally, this research confirms the vital importance of taking a quantitative approach to the analysis of spontaneous speech to determine what speakers actually "do" and not what they "say" they do or what linguists "think" they do. However, we must be cautious to not overgeneralize the applicability of these findings to all German society or even to all communities of Swabian. Much additional work remains to be done. It is only through real-time, verifiable, and reproducible analyses that the full nature of linguistic variation and change can be unveiled.

Acknowledgements

I am grateful for feedback on the present article from Gregory Guy, Robert Bayley, Xiaoyun Jin, and Xinye Zhang, along with three anonymous reviewers of this journal. I would also like to thank the editors of this special issue for their helpful feedback and recommendations. Any deficiencies remaining are, of course, my own.

Author Statement

My research on subject pronoun variation in Swabian has been inspired by the work of Gregory Guy, Daniel Erker, and our crosslinguistic study examining subject pronoun use across six languages and 27 communities (Erker et al. in review). While some of the concepts and findings presented here have been motivated through our collaboration, this article presents original work on subject pronoun expression in Swabian, not previously published.

Data availability

This research has been supported by the Tübingen Universität Schwäbische Forschungsstipendium (TUSFS-2016-20). A detailed description of the corpus is in preparation: Beaman, Karen V. & Zhuge Gao. In prep. SWG*SLX: A multi-stage sociolinguistic corpus of spontaneously spoken Swabian.

Endnotes

1 In these and all examples, the first row is transcribed in a bespoke Swabian orthography (see Beaman 2024), the second row is modern standard German, and the third row is the English translation – a "loose", not literal, English translation to aid readability. Citations are in the following format: Snnn-nn-I-n-name-hh:mm:ss, where "Snnn" is the speaker ID, "nn" is the year of the recording (i.e., 82 or 17), "I" is for sociolinguistic interview, "n" is a sequential number when there are multiple recordings of the same speaker, "name" is the speaker's pseudonym (to protect their privacy), and "hh:mm:ss" is the location in the recording.

2 One reviewer commented on the validity of comparing written and spoken texts; however, in considering historical data, written texts are the only available sources. While written and spoken texts are indeed not fully comparable, these texts offer important insight into the evolution of subject pronoun expression in German.

3 I would like to thank Danny Erker for uncovering this finding in the data.

References

Abraham, Werner. 1993. Null Subjects in the History of German: From IP to CP. *Lingua* 89. 117–42.

Ammon, Ulrich. 2001. Dialect as an Educational and Social Challenge. In Eneko Barrutia (ed.). *Euskalkia Eta Hezkuntza*, 201–25. Bilbo: Mendebalde Kultura Elkartea.

Auer, Peter. 2020. Dialect (Non-)Acquisition and Use by Young People of Migrant Background in Germany. *Journal of Multilingual and Multicultural Development* 0(0). 1–15.

Axel, Katrin. 2005. Null Subjects and Verb Placement in Old High German. In Stephan Kepser & Marga Reis (eds.). *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, 27–48. Berlin: De Gruyter Mouton.

Axel, Katrin. 2007. *Studies on Old High German Syntax: Left Sentence Periphery, Verb Placement and Verb-Second*. Amsterdam: Benjamins.

Axel, Katrin & Helmut Weiß. 2011. Pro-Drop in the History of German from Old High German to the Modern Dialects. In Melani Wratil & Peter Gallmann (eds.). *Empty Pronouns*, 21–52. Berlin: De Gruyter.

Baayen, R. Harald, Douglas J. Davidson & Douglas Bates. 2008. Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items. *Journal of Memory and Language* 59(4). 390–412.

Bayley, Robert & Lucinda Pease-Alvarez. 1997. Null Pronoun Variation in Mexican-Descent Children's Narrative Discourse. *Language Variation and Change* 9(3). 349–71.

Beaman, Karen V. 2021. Swabian Relatives: Variation in the Use of the Wo-Relativiser. In Karen V. Beaman, Isabelle Buchstaller, Sue Fox & James A. Walker (eds.). *Advancing Socio-Grammatical Variation and Change: In Honour of Jenny Cheshire*, 134–64. New York: Routledge.

Beaman, Karen V. 2022. wenn Ø bloß Schwäbisch kãsch 'If You Can Only Speak Swabian': Null Subjects in Real- and Apparent-Time. In *International Conference on Language Variation in Europe (ICLaVE) 11. University of Vienna, Austria. April 11-14, 2022.*

Beaman, Karen V. 2024. *Language Change in Real- and Apparent-Time: Coherence in the Individual and the Community*. New York: Routledge.

Berndt, Rolf. 1956. *Form und Funktion des Verbums im Nördlichen Spätaltenglischen*. Halle: Max Niemeyer Verlag.

Bock, Kathryn. 1986. Syntactic Persistence in Language Processing. *Cognitive Psychology* 18. 355–387.

Bohnacker, Ute. 2013. Null Subjects in Swabian. *Studia Linguistica* 67(3). 257–289. https://doi.org/10.1111/stul.12014

Bourdieu, Pierre. 1977. *Outline of a Theory of Practice*. Cambridge: Cambridge University.

Burnham, Kenneth P. & David R. Anderson. 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods and Research* 33(2). 261–304.

Bybee, Joan L. 2017. Grammatical and Lexical Factors in Sound Change: A Usage-Based Approach. *Language Variation and Change* 29. 273–300.

Chomsky, Noam. 1981. *Lectures on Government and Binding. Studies in Generative Grammar* 9. Dordrecht: Foris.

Cooper, Kathryn & Elisabet Engdahl. 1989. Null Subjects in Zürich German. *Working Papers in Scandinavian Syntax* 44. 31–44.

Coupland, Nikolas. 1980. Style-Shifting in a Cardiff Work-Setting. *Language in Society* 9(1). 1–12.

Destatis, Statistisches Bundesamt. 2006. *Germany's Population by 2050: Results of the 11th Coordinated Population Projection*. Wiesbaden: Statistisches Bundesamt.

Dittmann, Jürgen. 2006. Konzeptionelle Mündlichkeit in E-Mail Und SMS. In Ulrike Reeg (ed.). *Interkultureller Fremdsprachenunterricht: Grundlagen und Perspektiven*, 79–97. Bari: Editioni di Pagina.

Dryer, Matthew S. 2013. Expression of Pronominal Subjects. In Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Duden. 2015. *Das Aussprachewörterbuch: Betonung Und Aussprache. Band 6*. Berlin: Dudenverlag.

Eckert, Penelope. 2019. The Limits of Meaning: Social Indexicality, Variation, and the Cline of Interiority. *Language* 95(4). 751–776. https://doi.org/10.1353/lan.2019.0072

Erker, Daniel & Gregory R. Guy. 2012. The Role of Lexical Frequency in Syntactic Variability: Variable Subject Personal Pronoun Expression in Spanish. *Language* 88(3). 526–557.

Frey, Eberhard. 1975. *Stuttgarter Schwäbisch: Laut- Und Formenlehre Eines Stuttgarter Idiolekts*. Marburg: Elwert.

Frietsch, Rainer. 2003. *"Intensivierung" von Bildungsabschlüssen zwischen 1970 und 2000. Analysen im Rahmen der jährlichen Berichterstattung zur technologischen Leistungsfähigkeit Deutschlands. Studien zum deutschen Innovationssystem. Nr. 5-2004*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).

Gregersen, Frans, Torben Juel Jensen & Nicolai Pharao. 2018. Comparing Speech Samples: On the Challenge of Comparability in Panel Studies of Language Change in Real Time. In Suzanne Evans Wagner & Isabelle Buchstaller (eds.). *Panel Studies of Variation and Change*, 155–180. New York: Routledge.

Gutman, Eynat. 2004. Third Person Null Subjects in Hebrew, Finnish and Rumanian: An Accessibility-Theoretic Account. *Journal of Linguistics* 40(3). 463–490.

Daniel Erker, Gregory R. Guy, Karen V. Beaman, Robert Bayley, Aria Adli, Rafael Orozco & Xinye Zhang. In Review. *Subject Pronoun Variation: A Cross-Language Sociolinguistic Study*. Cambridge: Cambridge University Press.

Haag-Merz, Christine. 1996. *Pronomen im Schwäbischen: Syntax und Erwerb*. Marburg: Tectum.

Hay, Jennifer, Janet B. Pierrehumbert, Abby Walker & Patrick LaShell. 2015. Tracking Word Frequency Effects through 130 Years of Sound Change. *Cognition* 139. 83–91.

Hockett, Charles F. 1950. Age-Grading and Linguistic Continuity. *Language* 26(4). 449–457.

Holmberg, Anders, Aarti Nayudu & Michelle Sheehan. 2009. Three Partial Null-Subject Languages: A Comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica* 63(1). 59–97.

Jaeggli, Osvaldo & Kenneth J. Safir. 1989. The Null Subject Parameter and Parametric Theory. In Osvaldo Jaeggli & Kenneth J. Safir (eds.). *The Null Subject Parameter*, 1–44. Dordrecht: Springer Netherlands.

Kroch, Anthony S. 1989. Reflexes of Grammar in Patterns of Language Change. *Language Variation and Change* 1(3). 199–244. https://doi.org/10.1017/S0954394500000168

Kursa, Miron B. & Witold R. Rudnicki. 2010. Feature Selection with the Boruta Package. *Journal of Statistical Software* 36(11). 1–13. https://doi.org/10.18637/jss.v036.i11

Labov, William. 1994. *Principles of Linguistic Change, Volume I, Internal Factors*. Oxford: Wiley-Blackwell.

Labov, William. 2002. The Anatomy of Style-Shifting. In Penelope Eckert & John R. Rickford (eds.). *Style and Sociolinguistic Variation*, 85–108. Cambridge: Cambridge University Press.

Lameli, Alfred. 2019. Areale Variation im Deutschen „horizontal“: Die Einteilung der arealen Varietäten des Deutschen. In Joachim Herrgen & Jürgen Erich Schmidt (eds.). *Language and Space: An International Handbook of Linguistic Variation. Volume 4: Deutsch,* 185–205*.* Berlin: De Gruyter Mouton.

MacKenzie, Laurel. 2019. Perturbing the Community Grammar: Individual Differences and Community-Level Constraints on Sociolinguistic Variation. *Glossa: A Journal of General Linguistics* 4(1). 1–23.

Milroy, Lesley. 1980. *Language and Social Networks*. Oxford: Basil-Blackwell.

Nagy, Naomi. 2015. A Sociolinguistic View of Null Subjects and VOT in Toronto Heritage Languages. *Lingua* 164. 309–27. https://doi.org/10.1016/j.lingua.2014.04.012

Nevalainen, Terttu, Helena Raumolin-Brunberg & Heikki Mannila. 2011. The Diffusion of Language Change in Real Time: Progressive and Conservative Individuals and the Time Depth of Change. *Language Variation and Change* 23(1). 1–43. https://doi.org/10.1017/S0954394510000207

Otheguy, Ricardo & Ana Celia Zentella. 2012. *Spanish in New York: Language Contact, Dialectal Leveling, and Structural Continuity*. Oxford: Oxford University Press.

Peduzzi, Peter, John Concato, Elizabeth Kemper, Theodore R. Holford & Alvan R. Feinstein. 1996. A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis. *Journal of Clinical Epidemiology* 49(12). 1373–1379.

Rosenkvist, Henrik. 2009. Referential Null Subjects in Germanic Languages – an Overview. *Working Papers in Scandinavian Syntax* 84. 151–180.

Rosenkvist, Henrik. 2010. Null Referential Subjects in Övdalian. *Nordic Journal of Linguistics* 33(3). 231–267.

Schwitalla, Johannes. 2006. *Gesprochenes Deutsch: Eine Einführung*. Berlin: Schmidt.

Silverstein, Michael. 2003. Indexical Order and the Dialectics of Sociolinguistic Life. *Language and Communication* 23(3–4). 193–229.

Stratton, James M. & Karen V. Beaman. 2024. ‘Fei Schee’: The Social Meaning of Intensifier Use in Swabian. In James M. Stratton & Karen V. Beaman (eds.). *Expanding Variationist Sociolinguistics: Studies in Linguistic Variation and Change in Varieties of German*, 111–131. New York: Routledge.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, Forests, and Trees of York English: *Was/Were* Variation as a Case Study for Statistical Practice. *Language Variation and Change* 24(2). 135–178.

Tamminga, Meredith, Laurel MacKenzie & David Embick. 2016. The Dynamics of Variation in Individuals. *Linguistic Variation* 16(2). 300–336.

Tomaschek, Fabian, Benjamin V. Tucker, R. Harald Baayen & M. Fasiolo. 2018. Practice makes perfect: The consequences of lexical proficiency for articulation. *Linguistics Vanguard*. 4(s2). 1–13.

Torres Cacoullos, Rena & Catherine E. Travis. 2019. Variationist Typology: Shared Probabilistic Constraints across (Non-)Null Subject Languages. *Linguistics* 57(3). 653–692.

Travis, Catherine E. & Rena Torres Cacoullos. 2012. What Do Subject Pronouns Do in Discourse? Cognitive, Mechanical and Constructional Factors in Variation. *Cognitive Linguistics* 23(4). 711–748. https://doi.org/10.1515/cog-2012-0022

Trudgill, Peter. 1981. Linguistic Accommodation: Sociolinguistic Observation on a Sociopsychological Theory. In Carrie S. Masek, Roberta A. Hendrick & Mary F. Miller (eds.). *Papers from the Parasession on Language and Behavior*, 218–237. Chicago: Chicago Linguistic Society.

Trutkowski, Ewa. 2010. Referential Null Subjects in German. In George Walken (ed.). *Proceedings of the Sixth Cambridge Postgraduate Conference in Language Research*, 206–217. Cambridge: Cambridge Institute of Language Research (CILR).

Vergeiner, Philip C. 2021. *Bewertungen – Erwartungen – Gebrauch: Sprachgebrauchsnormen Zur Inneren Mehrsprachigkeit an Der Universität.* Stuttgart: Steiner.

Volodina, Anna & Helmut Weiß. 2016. Diachronic Development of Null Subjects in German. In Sam Featherston & Yannick Versley (eds.). *Quantitative Approaches to Grammar and Grammatical Change*, 187–206. Berlin: De Gruyter.

Wagner, Suzanne Evans. 2012. Age Grading in Sociolinguistic Theory. *Language and Linguistics Compass* 6(6). 371–382.

Walkden, George. 2013. Null Subjects in Old English. *Language Variation and Change* 25(2). 155–178.