JL | JournaLIPP

Language Variation

Research, Models, and Perspectives

Edited by

Michael Breyl
Yossef Pinhas
Elizabeth Stadtmiller

7

2019

ISSN: 2193-8105

JL | JournaLIPP

Language Variation

Research, Models, and Perspectives

Herausgegeben von

Michael Breyl, Yossef Pinhas und Elizabeth Stadtmiller

JournaLIPP 7 | München 2019

https://lipp.ub.lmu.de/

ISSN: 2193-8105

Table of Contents / Inhaltsverzeichnis

Introduction

25th LIPP Symposium

Language Variation: Research, Models, and Perspectives

Variation is a key concept in sociolinguistics. Not only can members of a speech community influence their use of language intentionally, but a single speaker's unconscious idiosyncrasies may also show significant variation when compared to other individuals in the community. Further, groups within communities, as well as whole communities, commonly share a large set of variants, that set them apart from other groups of speakers sharing the same language. The languages of the world are therefore not each a uniform entity but instead exhibit multifaceted patterns of internal variation. In general, four types of variation are defined: diatopic (based upon geographical location), diastratic (describing the language of a specific sub-set of a society), diachronic (comparing different stages of languages throughout history) and diaphasic/functional (register, based upon specific settings and pragmatics).

The 25th LIPP symposium provided a forum to discuss these patterns of variation across a wide range of languages using various disciplines and methodologies, including synchronic and diachronic approaches to variation as well as quantitative and qualitative research. We saw a plethora of original research on language-variation phenomena at various linguistic levels (phonology, morphosyntax, vocabulary etc.) and in relation to regional, social, stylistic, and medial factors, as well as combinations of these. In addition, papers were presented which explored the perception of, and attitudes towards, variation as well as the construction of identity through linguistic variants. Finally, the symposium investigated the implications of research findings for existing theoretical models and concepts and for the development of new frameworks that enable us to describe and categorize language variation.

The 25th LIPP symposium *Language Variation: Research, Models, and Perspectives* took place June 20–22 2018 and was organized by Michael Breyl, Carolin Harthan, Christoph Hauf, Yossef Pinhas and Elizabeth Stadtmiller of the *Graduate School Language & Literature Munich - Class of Language*.

Contributions

The first contribution is by **Juan Manuel Hernández-Campoy**. In this keynote paper, the author examines five stylistic paradigms as they pertain to sociolinguistics and, in a broader sense, social philosophy. By connecting and rooting each model in its own "long background in social philosophy," Hernández-Campoy offers insight into style change as a "complex, multidimensional phenomenon" that permeates stylistic, linguistic and social components of sociolinguistic variation to respond, to as well as project meaning and identity into, social discourse.

Delving from models into research on specific variation phenomena, the second keynote contribution by **Lars Bülow** considers the case of diatopic dialectal variation in the plural forms of the German verb *sein* (Engl. 'to be'*)* in the region of Salzburg, Austria. By closely examining specific phenomena in *real* and *apparent* time, Bülow shows a tendency toward advergence into a supra-regional, near-standard variety of Bavarian dialects in Austria. [This paper is in German.]

Continuing in the trend of single-variety, **Philipp Striedl**'s paper deals with the diastratic *Zahalit* language of Israeli soldiers. Using data from written and oral sources, the author illustrates both the linguistic characteristics and the social function of the *Zahalit* variation of Modern Hebrew. Striedl goes on to discuss the prevalence of the military variety and, by extension, the relevance of the army, to modern Israeli society.

The next contribution is by **Martin Eberl**, who studies social identities as a basis for linguistic variation. More specifically, the author concerns himself with Twitter data from two subsets – supporters of the 2016 presidential candidates Hillary Clinton and Donald Trump versus Russian online operatives (a.k.a. trolls). Eberl's analysis of lexical variation, hashtags and emojis finds high similarity between the opposing political groups within the US but distinctive differences between this group and the inflammatory Russian tweets.

Alexandra Chudar's article considers diatopic variation, this time in the English-speaking world, examining diminutive forms in Southern Hemisphere Englishes. Comparing diminutives ending on *-ie*, the author describes similarities and differences, the latter arising from language contact and extralinguistic environmental factors, both geographical and social.

Once again considering varieties of English, **Laetitia Van Driessche** and **Hubert Cuyckens** combine diachronic and synchronic perspectives in considering clausal verb complementation. Using various statistical analyses to dissect their data on finite complementation involving *that-clauses* and non-finite complementation comprised of gerundial and infinitival clauses with the words *expect* and *suggest*, the authors' corpora-based study reveals an increase in non-finite use along with a second learner effect in *expect*, but not in *suggest*. For both verbs, *variety* is an influential factor.

The final paper of this journal is a third study on English, contributed by **Monika Pukli**. Closely considering derhotacization in Scottish English and taking into account individual as well as diastratic and diachronic differences, Pukli shows that derhotacization is not necessarily linked to socio-economic factors or speakers' age. The author makes a case that there is a high tolerance for variation within phonological systems without necessarily giving rise to new standards or effecting "any final change."

Thank you

First and foremost, we would like to thank everyone who took part in the 25th LIPP symposium, most especially the contributors of talks and posters. Special thanks go to the Conference's keynote speakers, **Stephan Elspaß** and **Lars Bülow** (University of Salzburg), **Juan Manuel Hernández-Campoy** (University of Murcia), **Daniel Schreier** (University of Zurich) and **Elisabeth Stark** (University of Zurich), whose talks and research helped shape our conference.

In addition, we would like to thank **Andreas Dufter**, **Stephanie Hackert** and **Anthony Rowley**, all of the University of Munich, for their patronage and support throughout.

No journal can be published without a board of reviewers with a wealth of scientific expertise who are willing to take the time to read, comment, re-read and advise on a wide range of papers. We feel deeply grateful that so many highly-qualified individuals were able to support our journal and would like to thank our reviewers for their effort and anonymous but immeasurable contributions.

Finally, none of this would have been possible without our co-organizers **Carolin Harthan** and **Christoph Hauf**, as well the support of the former academic coordinators of the Graduate School Language & Literature, **Daniel Holl** and **Katharina Jakob**, the current academic coordinator **Teresa Barberio** and the student assistants at the time of the symposium, **Christiane Bayer** and **Amina Wittmann**. Thank you!

The editors: Michael Breyl, Yossef Pinhas & Elizabeth Stadtmiller

Einleitung

25. LIPP-Symposium

Language Variation: Research, Models, and Perspectives

Variation ist ein Schlüsselbegriff der Soziolinguistik. Nicht nur können Mitglieder derselben Sprachgemeinschaft ihre Sprachverwendung mit Intention unterschiedlich verwenden, sondern der Sprachgebrauch einzelner Sprecher*innen kann auch auf unbewusster Ebene im Vergleich mit anderen Individuen der Gemeinschaft erhebliche Variation aufweisen. Weiterhin teilen *speech communities* sowie mitunter Teilgruppen innerhalb solcher Gemeinschaften identitätsbildende sprachliche Variation in großem Umfang – im Kontrast mit anderen Gruppen bzw. Gemeinschaften und obwohl sie derselben Sprache zugehören. Sämtliche Sprachen stellen somit keine homogenen Entitäten dar, sondern zeichnen sich durch vielfältige Muster interner Variation aus. Im allgemein werden dabei vier Arten von Variation definiert: diatopisch (auf Basis des geographischen Standorts), diastratisch (beschreibt die Sprache einer spezifischen Teilgruppe einer Gesellschaft), diachron (vergleicht verschiedene Stadien von Sprachen in der Geschichte derselben) und diaphasisch/funktional (Register, basierend auf spezifischen Settings und Pragmatik).

Das 25. LIPP-Symposium hat ein Forum geboten, um diese Muster in diversen Sprachen und unter Einbezug verschiedener Bereiche und Methodologien zu diskutieren. Unter anderem wurden synchrone und diachrone Ansätze sowie quantitative und qualitative Forschung vorgestellt. Es wurden eine Vielzahl an Studien zu Sprachvariationsphänomenen auf verschiedenen linguistischen Ebenen (Phonologie, Morphosyntax, Wortschatz usw.) beleuchtet. Dabei nahmen die Vortragenden Bezug auf regionale, soziale, stilistische oder mediale Faktoren bzw. eine Kombination derselben. Zusätzlich wurden Paper präsentiert, die sich mit der Perzeption von und Einstellungen gegenüber Variation sowie der Konstruktion von Identität durch linguistische Varianten beschäftigen. Das Symposium hat die Implikationen von Forschungsergebnissen für bereits existierende theoretische Modelle und Konzepte sowie für die Entwicklung neuer Frameworks untersucht, welche die Beschreibung und Kategorisierung von Sprachvariation vereinfachen können.

Das 25. LIPP-Symposium *Language Variation: Research, Models, and Perspectives* hat vom 20. – 22. Juni 2018 stattgefunden und wurde von Michael Breyl, Carolin Harthan, Christoph Hauf, Yossef Pinhas und Elizabeth Stadtmiller von der *Graduate School Language & Literature Munich - Class of Language* organisiert.

Beiträge

Der erste Beitrag wurde von **Juan Manuel Hernández-Campoy** vorbereitet. In diesem Keynote-Artikel untersucht der Autor fünf stilistische Paradigmen in Bezug auf Soziolinguistik und, im breiteren Sinne, Sozialphilosophie. Indem er jedes Modell mit seinem eigenen „Hintergrund in der Sozialphilosophie“ verbindet und verankert, gibt Hernández-Campoy Einsicht in individuellen Stil als ein in sich komplexes, multidimensionales Phänomen, welches nicht nur stilistische, sondern auch linguistische und soziale Komponente der soziolinguistischen Variation beeinflusst, und auf diese Weise sozialem Diskurs Bedeutung und Identität zuschreibt. [Dieser Beitrag ist auf Englisch verfasst.]

Mit dem zweiten Keynote-Paper von **Lars Bülow** geht es direkt von Modellen weiter zur Forschung an spezifischen Variationsphänomenen. In diesem Beitrag betrachtet der Autor den Fall der diatopischen dialektalen Variation in den Pluralformen des Verbs *sein* in der Region um Salzburg. Mit einer genauen Untersuchung spezifischer Phänomene in *real-* und *apparent-time*, zeigt Bülow eine Tendenz in Richtung Advergenz zu einer supra-regionalen, standardnahen bairischen Varietät in Österreich.

Ebenso thematisiert das Paper von **Philipp Striedl** eine diastratische Einzelvarietät: die *Zahalit*-Sprache der israelischen Soldaten. Unter Verwendung von geschriebenen wie gesprochenen Sprachquellen zeigt der Autor linguistische Charakteristika auf und beschreibt die soziale Funktion der *Zahalit*-Variation des Modernen Hebräischen. Darüber hinaus diskutiert Striedl die Prävalenz der militärischen Varietät und, damit einhergehend, die Relevanz der Armee in der modernen israelischen Gesellschaft. [Dieser Beitrag ist auf Englisch verfasst.]

Der nächste Beitrag stammt von **Martin Eberl** und untersucht soziale Identitäten als Basis für linguistische Variation. Spezifisch geht es dem Autor um Twitter-Daten zweier Subsets: Anhänger der US-amerikanischen Präsidentschaftskandidaten von 2016, Hillary Clinton und Donald Trump, im Gegensatz zu russischen online Agent*innen (Trolls). Eberls Analyse der lexikalischen Variation und der Nutzung von Hashtags und Emojis zeigt eine große Ähnlichkeit zwischen den sich gegenüberstehenden politischen Gruppen innerhalb der USA, aber weist auch auf deutliche Unterschiede zwischen diesen und den russischen Troll-Tweets hin. [Dieser Beitrag ist auf Englisch verfasst.]

Der Artikel von **Alexandra Chudar** untersucht die diatopische Variation, diesmal im englischsprachigen Bereich, mit ihrer Untersuchung der Diminutivformen der *Southern Hemisphere Englishes*. Indem sie Diminutive, die auf -ie enden, miteinander vergleicht, beschreibt die Autorin Ähnlichkeiten und Unterschiede, wobei Letztere aus Sprachkontakt sowie extralinguistischen Faktoren geographischer und sozialer Natur stammen. [Dieser Beitrag ist auf Englisch verfasst.]

Eine weitere Untersuchung von Varietäten des Englischen bieten **Laetitia Van Driessche** und **Hubert Cuyckens**. Die Autor*innen vereinen in ihrer Untersuchung der *clausal verb complemention* die diachronen und synchronen Perspektiven. Mithilfe statistischer Analysen betrachten Van Driessche und Cuyckens finite *complementation* mit *that* (dt. ‚dass') und nicht-finiter *complementation* von Gerund- und Infinitivphrasen mit *expect* (dt. ‚erwarten') und *suggest* (dt. ‚vorschlagen'). Ihre korporabasierte Studie zeigt eine Zunahme in nicht-finiter Verwendung mit Effekt von Englisch als Zweitsprache bei *expect*, aber nicht bei *suggest*. Der Faktor *Varietät* hat auf beide Verben einen starken Einfluss. [Dieser Beitrag ist auf Englisch verfasst.]

Der letzte Beitrag der Ausgabe ist eine dritte Untersuchung des Englischen, von **Monika Pukli**. Mit einer genauen Betrachtung der *derhotacization* (Entwicklung in Richtung Auslassung oder Veränderung des Lautes /r/) im schottischen Englisch und unter Einbezug individueller sowie diastratischer und diachroner Unterschiede, zeigt Pukli, dass *derhotacization* nicht unbedingt mit sozioökonomischen Faktoren oder Alter der Sprecher*innen zusammenhängt. Die Autorin argumentiert dafür, dass es eine hohe Toleranz für Variation innerhalb phonologischer Systeme gibt, ohne dass diese „notwendigerweise neue Standards hervorrufen oder finale Veränderungen verursachen". [Dieser Beitrag ist auf Englisch verfasst.]

Danksagung

Wir möchten uns recht herzlich bei allen Teilnehmenden am 25. LIPP-Symposium bedanken, vor allem denjenigen, die einen Vortrag oder ein Poster beitragen haben. Besonderer Dank gilt den Keynote-Sprecher*innen der Konferenz: **Stephan Elspaß** und **Lars Bülow** (Universität Salzburg), **Juan Manuel Hernández-Campoy** (Universität Murcia), **Daniel Schreier** (Universität Zürich) and **Elisabeth Stark** (Universität Zürich).

Zusätzlich geht unser Dank an **Andreas Dufter**, **Stephanie Hackert** und **Anthony Rowley** von der LMU München für ihre Schirmherrschaft und Unterstützung während des gesamten Prozesses.

Kein Journal kann ohne ein Team von Gutachter*innen veröffentlicht werden, die enorme Fachkompetenz mitbringen, sowie die Zeit, Artikel (mehrmals) zu lesen und zu begutachten und die Herausgeber zu beraten. Wir sind den hochqualifizierten Individuen sehr dankbar, die sich bereit erklärt haben, unser Journal zu unterstützen und möchten uns bei den Reviewern für ihr Bemühen und ihren anonymen aber unermesslichen Beitrag bedanken.

Zu guter Letzt wäre nichts von alldem möglich gewesen, ohne unsere Mitorganisator*innen, **Carolin Harthan** und **Christoph Hauf**, sowie die Unterstützung der früheren Akademischen Koordinator*innen der Graduate School Language & Literature, **Daniel Holl** und **Katharina Jakob**, der aktuellen Koordinatorin **Teresa Barberio** und der studentischen Hilfskräfte zum Zeitpunkt des Symposiums, **Christiane Bayer** und **Amina Wittmann**. Danke!

Herausgeber: Michael Breyl, Yossef Pinhas & Elizabeth Stadtmiller

Stylistic Models in Sociolinguistics and Social Philosophy

Juan M. Hernández-Campoy (University of Murcia)
jmcampoy@um.es

Abstract

The emergence of new theoretical paradigms along the history of scientific revolutions always occurs in tune with the social philosophy of their time. Historical periods or schools of thought cannot be treated as discrete entities, since they are built upon the immediately preceding stage from which they start and which they normally react against. Stylistic models in Sociolinguistics have not been unaware of the social philosophy of their time. While everybody would agree that intra-speaker variation is a phenomenon conditioned by extralinguistic factors, the resources and mechanisms for reflecting its presence in language production and effective social meaning have been associated with different linguistic constructs and theories trying to account for its nature and functioning. The aim of this paper is to explore and illustrate the main different theoretical models developed to account for the nature, motivations and mechanisms for the use and effect of style-shifting in social interaction: Audio-monitoring, Audience Design, Script Design, Register Model and Speaker Design.

Keywords: *sociolinguistics, style-shifting, intra-speaker and inter-speaker variation, reactive/proactive speakers*

1 Scientific Revolutions

Historical periods or schools of thought cannot be treated as discrete entities, with a monolithic nature and abrupt boundaries, since they are built upon the immediately preceding stage, from which they start and which they normally react against (see Kuhn 1962). The crisis of a current paradigm, because of the dissatisfaction with its accounts accepted until then, and its replacement with another is what gives way to scientific revolutions. But scientific revolutions are accumulative in their theory of knowledge. As Lass (1984:8) suggested: "the history of any discipline involves a lot of old wine in new bottles (as well as new wine in old bottles, new wine in new bottles, and some old wine left in the old bottles)."

As we will see below, the emergence of new paradigms with new the philosophical foundations has been conditioning the development of Linguistic Science and any of its branches, such as Sociolinguistics. Likewise, stylistic models within Sociolinguistics have not been unaware of the social philosophy of their time. While everybody would agree that intra-speaker variation is a phenomenon conditioned by extralinguistic factors, the mechanisms and motivations for reflecting its presence in language production and effective social meaning have been associated with different linguistic constructs and theories trying to account for its nature and functioning.

2 Linguistics

Linguistic theory, as Williams (1992:40) asserted, "has not emerged separately from the social philosophy of its time. Rather, it must be seen as a manifestation of the ongoing debate on the nature and the social world". To simplify somewhat, the nineteenth-century philologist's historicist and comparative urge was in overt opposition to the humanism and classicism of the Renaissance and seventeenth-century rationalism; however, the structuralism of the beginning of the twentieth century was an alternative to nineteenth-century historicism and comparativism, as exemplified by Alcaraz (1990). Even within the same period, different theoretical trends have followed one another in linguistics –structuralism, functionalism,

generativism, variationism, cognitivism, etc.– throughout the course of the twentieth-century (see also Figueroa 1994; Markova 1982). Linguistic studies, therefore, cannot be judged disregarding those of other periods, or even the same period, in order to contrast ontological, epistemological or axiological aspects. In this way, similarities and differences can be established between theories –and this is in fact the function of a metatheory– whereby they can be characterized, classified, and explained in terms of the criteria and relative to the conditions under which the scholar can meaningfully choose between alternative theories and/or their parts (Oyelaran 1970:430-431).

3 Sociolinguistics

The origins of Sociolinguistics also reflect rebellion as well as epistemological hybridisation. I have always underlined four main motivations for the development of Sociolinguistics (see Hernández-Campoy/Almeida 2005; Hernández-Campoy 2014): i) the dissatisfaction among many linguists in the 1960s with and reaction against previous paradigms: Chomsky, Saussure and Bloomfield; ii) the redefinition and reformulation of *Traditional Dialectology*; iii) the growing interest among linguists in *Sociology* and its scope: poverty and disadvantage as political issues in Western World; and iv) the advent of the Quantitative Revolution. That is: new wine in old bottles, old wine in new bottles, new wine in new bottles, and the new bottles (see also Tucker 1997).

Sociolinguistics, therefore, did not emerge separately from the social philosophy of its time either. It was mostly after the Second World War that the crisis of historicist conceptions began and that a solid neopositivist trend (or even school of thought) arose in science in general, which became the 'quantitative revolution'. There is a confrontation between qualitative and quantitative approaches, also affecting theories, methods and techniques, and, above all, two radically different conceptions of scientific research and the growth of knowledge (truth/reality vs. beliefs/intuitions). In order to express accurately and plainly the results of their analysis, the language of mathematics and logic, conceived as the authentic syntax of science, must be used, for which validity and verifiability are the fundamental criteria and in which coincidence is conceived in terms of probability (Hernández-Campoy/Almeida 2005:10-11). This will lead to the rejection of intuitions and introspective knowledge, which now began to be seen as inferior due to its subjective nature. The new researcher has to be a fieldworker mainly, rather than an armchair one, in order to avoid losing contact with reality.

Within Sociolinguistics, the philosophical foundations of Variation Theory specifically are also broadly anchored to Determinism and Neopositivism, understanding that the only valid knowledge comes from scientific knowledge through empirical evidence –rather than intuitional and introspective (see Hernández-Campoy 2016). In this way, society and social systems are organic models of social structure (Mackenzie 1890/2006; Olssen 2010) that operate according to their own quasi-absolute laws, like those that operate in the physical world because of cause-effect absolute laws of nature and that regard individual behaviour as easily predictable. Additionally, assuming that everything is caused by something in a predictable way, the universe is viewed as a deterministic place where the laws of nature would allow us to easily describe, explain, and predict its state. This causal determinism is a reductionist (and even essentialist) idea that explains the world in terms of a few narrowly defined factors. From a sociological perspective, its application to societal systems means that human behaviour is entirely governed by causal laws, where physiology, environment, population pressures and even genetics determine the organisation of societies; that is, "the attributes and behaviour of socially defined groups can be determined and explained by reference to cultural and/or biological characteristics believed to be inherent to the group" (Bucholtz 2003:400). This position amounts to a probabilistic model of macroscopic analysis where, although people are

not molecules, they can be regarded as predictable in their aggregate behaviour on the basis of mathematical probability (Jones 1990:189). As a result, categorical patterns are derived from probabilistic tendencies.

Similarly, in Sociolinguistics, speech behaviour is thought to reflect social structure simply because it is social structure that determines speech behaviour. It is like a Catch-22 circular logic rule: "… you can't get a job as a banker unless you can talk the way a banker is supposed to talk, and you won't talk like a banker is supposed to talk unless you've grown up in a part of the speech community that is made up of bankers and people like them" (Meyerhoff 2006:147). The study of the relationship between language and society through the correlation of extralinguistic factors (socio-demographic and/or context variables) with intralinguistic elements allowed sociolinguistics to decipher linguistic variation and social meaning, and thus to account for variability in language quantitatively. With their rigorous adoption of scientific methods, assuming Determinism and the mechanistic nature of human behaviour, linguists' explicit neo-positivist desire is to develop a quantified social dialectology where extralinguistic (mostly social) factors are capable of explaining the establishment of laws, relationships, and processes. In this way, the sociolinguistic behaviour was conceived as predictable in terms of diatopic/geographical variation, diastratic/social variation and diaphasic/stylistic variation. The speaker's social provenance, or age, sex, ethnicity, etc. determine the characterisation of their speech to such an extent that it is accurately predictable, providing us with the speaker's sociolinguistic profile/portrait according to their geographic and socio-demographic features, and even their stylistic characterisation through the verbal practices used in their linguistic production (see Hernández-Campoy/Almeida 2005). The so widely-known patterns of sociolinguistic behaviour have been developed in the industrialised Western world from the point of view of social class, age, gender, race, ethnicity, social networks, etc., like functions or equations:

- The higher the speaker's social class, the higher the frequency of overt prestige (standard) features in their speech, and vice versa;
- Female speakers exhibit higher frequencies of overt prestige features than male ones;
- Younger speakers have higher use of non-standard features than middle-aged;
- Older speakers use more conservative features than younger ones;
- Speakers with weaker social ties and high mobility make more use of prestige standard features and less use of local features than those with strong social ties and less mobility.

4 Style in Sociolinguistics

Given the singularly central position of style in the correlation of *linguistic, social* and *diaphasic* elements, intra-speaker variation is undoubtedly seen as consubstantial to sociolinguistic studies now and becoming a major focus of research within the field. While everybody would agree that stylistic variation is a phenomenon conditioned by extralinguistic factors, its presence in language production and effective social meaning has been associated with different linguistic constructs and theories trying to account for its nature and functioning.

Historically, the debate on *responsive-initiative* motivations in stylistic variation constitutes a central issue of the traditional pendulum-oscillating dilemma in social theory about the relationship between *structure* and *agency*, i.e. between sociolinguistic limitations and creativity, and also between speaker intention and listener understanding (Schilling 2013: 342-343; Bell 2014: 305-306): "[a]pproaches which treat speakers as untrammeled agents do not take enough account of the role of structure in interaction and life, just as approaches which

treat speakers as sociodemographic correlates did not take adequate account of individual agency" (Bell 2014: 305-306). Structure refers to the social norms that shape as well as constraint the way we live and sociolinguistically behave. Conversely, agency is our ability to customise that way we live and sociolinguistically behave according to our individual requirements and intentions –taking our own actions, following our own practices, and making our own way and with our own choices (Bell 2014: 305). It is in recent Sociolinguistics that the oscillation of the pendulum is swinging towards agentivity and creativity, and thus moving away from structural constraints and norms (see also Johnstone 2000, 2001).

Notions and concepts such as *ethos* and *pathos* in ancient Greek Rhetoric, *elocutio* and *pronunciatio* in Roman Oratory, *foregrounding*, *predictability* and *expectancy* in 20th-century Stylistics, or more recently, *enregisterment*, *stylisation*, *stance*, *authenticity*, *persona management*, and *crossing* in intra-speaker Sociolinguistics have all contributed to a greater understanding of the nature, functioning, and effectiveness of style-shifting processes in social interaction (see Hernández-Campoy 2016). In this evolution of language from the most baroque rhetorical resources to the most direct linguistic forms for economy reasons – though always searching a particular effect – stylistic and rhetorical devices to enhance speech or writing do not necessarily have to be complex figures of speech (*figurae verborum*) or figures of thought (*figurae sententiarum*), but rather mere linguistic variables or just the alternation between standard and non-standard uses in linguistic varieties (see Figure 1).

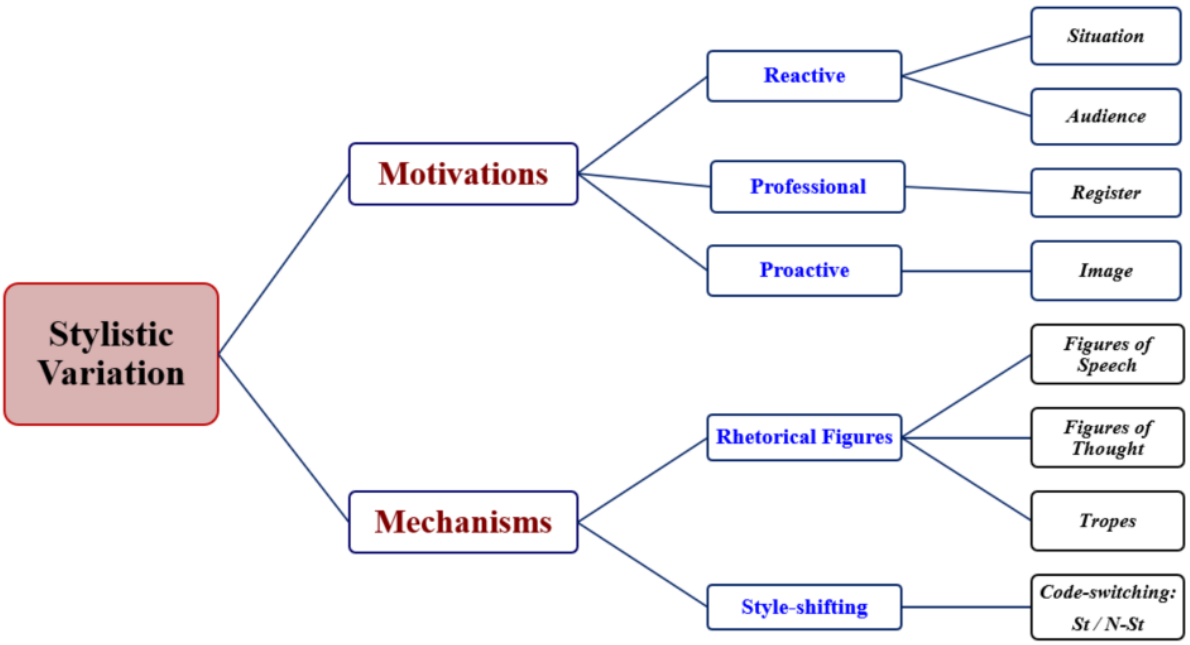


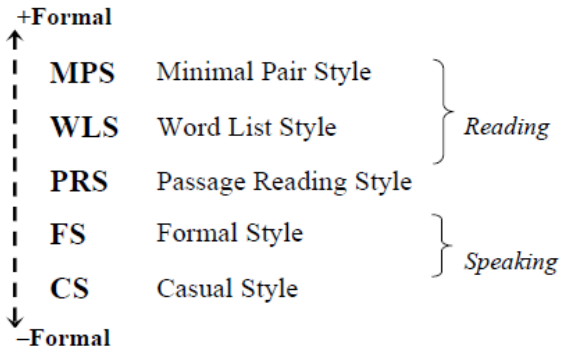
Figure 1: Motivations and mechanisms in stylistic variation

The following sections will explore the main different theoretical models developed to account for the nature, motivations and mechanisms for the use and effect of style-shifting in social interaction: Audio-monitoring, Audience Design, Script Design, Register Model, and Speaker Design.

4.1 Attention to speech model

The Labovian approach to stylistic variation (Labov 1966; 1972) was also inevitably inspired by that neopositivist determinism (see Hernández-Campoy 2016:65-94). Based on mechanistic foundations (considering speakers as androids), style-shifting is conceived as a conscious social

reaction (response) to a situation and appears scaled within a formality continuum –ranging from least to most formal[1].



Labov (1966) was able to quantify stylistic variation and to extract its indexical relationship with the individual's social background and situation. He found that although the different social class groups have different levels of usage of a given variable, their evaluation of the different variants is exactly the same: speakers of all classes change their pronunciation in exactly the same direction –i.e. by increasing the percentage of prestige forms in their speech as stylistic context becomes more formal, and vice versa (see Figure 2).

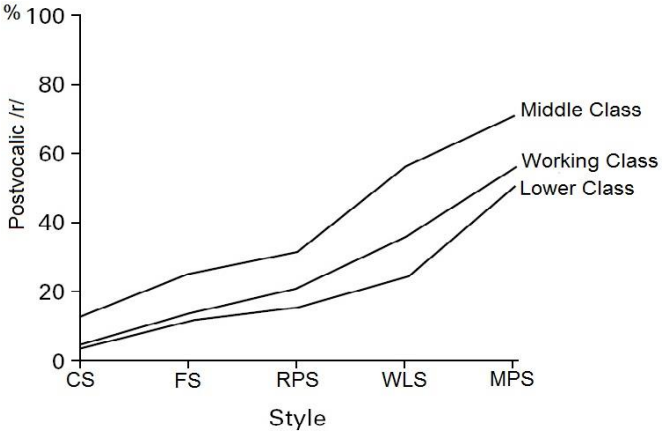


Figure 2: Postvocalic /r/, social class and styles in New York City (adapted from Labov 1966/2006: 141; Figure 7.1)

As a result, the same speaker uses different linguistic varieties in different situations and for different purposes, and the totality of linguistic varieties so used by a particular community of speakers constitutes the *verbal repertoire* (Trudgill 1983:100). Shared patterns of style-shifting are thus one of the defining characteristics of membership in a particular speech community (Rickford/Eckert 2001:10). Intra-speaker (stylistic) variation is largely a function of inter-speaker variation, where some individuals exhibit a much wider range of stylistic variation than

[1] In the contrast between spoken and written language, Koch/Oesterreicher (1994; 2012) understand the relationship between the characteristics of oral and written dimensions as a more-or-less continuum rather than as an either-or dichotomy, beyond the phonic-graphic distinction. Within their orality-literacy model, oral language is seen as the 'language of immediacy' typically associated with private settings, a high degree of familiarity and low emotional distance between the interactants, as among family and friends. In contrast, written language is viewed as a 'language of distance' and typically associated with public/official settings, situations of distance and formality, as in legal texts. But a piece of writing can be medially written but conceptually oral, as in church sermons, and the reverse situation, as in postcards. Consequently, private correspondence belongs to the written level from a medial perspective, although conceptually closer to the immediacy end of the continuum.

others: the different social class groups have different levels of usage of sociolinguistic variables drawing a perfect symmetry, where the most formal style of the lowest social group is similar to the most informal style of the highest social class, as Trudgill (1974) was able to quantify (see Table 1).

Table 1: (ng) Indexes by social class and style in Norwich (Trudgill 1974). Usage of non-standard variants

Social Class	**Style**			
	WLS	**PRS**	**FS**	**CS**
Middle Middle Class	000	000	003	028
Lower Middle Class	000	010	015	042
Upper Working Class	005	015	074	087
Middle Working Class	023	044	088	095
Lower Working Class	029	066	098	100

It is in this intersection between the stylistic and the social dimensions that makes style be a crucial sociolinguistic concept: there is a point along the symmetrical axis where, as Labov (1972:240) illustrated, objectively and quantitatively, it would be difficult to distinguish "a casual salesman from a careful pipefitter".

Style is then a reflection (or the product) of the awareness and attention paid by the speaker to their own speech, depending on external factors (such as topic, addressee, audience and situation) which 'determine' the level of formality and, thus, the linguistic variety to be employed –as well as the degree of self-monitoring in speech production: the more attention a speaker pays, the more formal their style will be, and vice versa.

The Attention to Speech Model (or Audio-monitoring) became 'the universal factor' with the status of a quasi-absolute law operating to cause style differences (Bell 2007b:96), and its basic principles are inherently related to the theoretical foundations of Variationist Sociolinguistics:

i) *The Principle of Graded Style-shifting*: No single speaker is mono-stylistic, though some have a wider verbal repertoire than others;

ii) *The Principle of Range of Variability*: The variation that any individual shows in their speech is never greater than the differences between the social groups that their style-shifting is derived from;

iii) *The Principle of Socio-stylistic Differentiation*: The linguistic features involved in stylistic variation are mostly the same as those marking social variation; i.e. those features typically found at the high end of the social scale are equally high on the stylistic scale, and vice versa;

iv) *The Principle of Sociolinguistic Stratification*: Variation originates in a hierarchy of evaluative judgments, where *indicators* denote social stratification only and *markers* show both social stratification and style-shifting;

v) *The Principle of Stylistic Variation*: Different styles constitute different ways of saying the same thing;

vi) *The Principle of Attention*: Styles can be classified unidimensionally according to the degree of attention paid to speech;

vii) *The Vernacular Principle*: The vernacular is the most natural, spontaneous and requires the least attention to the way of speaking;

viii) *The Principle of Formality* (The Observer's Paradox): Any systematic observation of the vernacular must minimise its effects on the informant's language production in order to guarantee the capture of the genuinely most natural and spontaneous speech.

However, this Labovian view of style-shifting as if speakers were automata, through its notion of attention paid to speech, and the formal-informal distinction on a linear scale of style-shifting, which had been the 'received wisdom' in the dominant variationist strand of Sociolinguistics until the late 1970s (Bell 1984:147; 2007a:91), began to be questioned in the early 1980s because of its mechanistic approach. The amount of attention paid to speech, or audio-monitoring, according to Bell (1984:150), is just a mechanism of response intervening between a situation and a style, but it is not intra-speaker variation in itself. Labov's axiom operated more as a kind of descriptive framework than an explanatory model, where style was not characterised in itself (Gadet 2005:1357). Also, as discovered in media language, and as we will see below, the effect of the present or absent audience on speech is also crucial in linguistic production, given that the same speaker can consciously shift into a different style not as a reaction to a topic or situation but to an addressee (Bell 1977/1979; 1984).

The Labovian axiom, as well as its sociolinguistic theory, conceived language as a mere reflection of social structures and norms for interpersonal communication, and speakers as androids that modify their speech production through style-shifting passively and as a response (or reaction) to an external situation, without considering any possible agency in choice and use of the stylistic resources proactively. The Labovian axiom became unable to explain all cases of stylistic variation in interpersonal communication. As Bell (2007a:91) states, "[w]hat happens when a speaker talks in any social situation involves many linguistic features almost simultaneously, at all levels of language, all contributing to the mosaic of the sociolinguistic presentation of self in everyday life." The androids in our speech communities are also able to think.

4.2 Audience design model

Factors such as "audienceship," "addressivity," "responsiveness" and "speaker agency" became crucial for new theories in the 1980s, putting the audience at the centre of intra-speaker variation. Founded on Speech Accommodation Theory (Giles 1979) and Linguistic Marketplace (Bourdieu/Boltanski 1975 or Sankoff/Laberge 1978), Bell's *audience design* theory (1984; 1991; 2001) conceived stylistic variation as an essentially responsive action to the characteristics of a present or absent audience, and introduced an initiative (proactive) dimension—in addition to the responsive (reactive) dimension. Intra-speaker variation appeared then as a response to interspeaker variation (Bell 1984, 158). All stylistic variation began to be explained through the audience. Largely rooted in Behaviourism and the Social Psychology, this model underlined the influence of external (audience) rather than internal stimuli (attention): the sources conditioning inner states and behaviour are external (in the environment, stimuli, responses, reinforcements, etc.) rather than internal (in the mind), and can be systematically observed. Crucial here is the Social Identity Theory (see Tajfel 1978; 1979), which highlights the importance of language to transmit identity, group solidarity, and language loyalty, as well as the Bakhtinian Multiple Voices theory and Dialogism (see Bell 2007a). The individual's multiplicity of social networks fosters the development of a polyhedral and versatile image, as well as a multifaceted behaviour, accommodating to their audience, as an ability to project different social identities in interpersonal communication for different purposes in also different moments, places, relational and interactional social contexts.

The best example is the pioneering case studied by Allan Bell on a radio presenter who worked for two radio stations in the same New Zealand public broadcasting service and was able to switch between them very quickly: YA Station, the 'National Radio'–playing classical music and attracting a higher-status audience– and ZB Station –a local community radio station playing popular music and attracting a wider range of social groups. Bell found that the speech of the same individual newsreader was different when reading bulletins in one radio station or the other, making considerable style shifts to suit the audience. Under these conditions, Bell characterises style as follows (see Hernández-Campoy 2016:95-130):

i) *Relational activity*: Style is what an individual speaker does with a language in relation to other people;

ii) *Indexicality*: Style derives its meaning from the association of linguistic features with particular social groups;

iii) *Responsiveness* and *Audienceship*: Speakers design their style primarily for and in response to their audience;

iv) *Linguistic repertoire*: Audience design applies to all codes and levels of a language repertoire, whether monolingual or multilingual;

v) *Style Axiom*: Variation on the style dimension within the speech of a single speaker derives from and echoes the variation which exists between speakers on the 'social' dimension;

vi) *Accommodative competence*: Speakers have a fine-grained ability to design their style for a range of different addressees, as well as for other audience members;

vii) *Discourse function*: Style-shifting according to topic or setting derives its meaning and direction of shift from the underlying association of topics or settings with typical audience members;

viii) *Initiative axis*: Besides the 'responsive' dimension of style, there is the 'initiative' dimension, where the style-shift itself initiates a change in the situation rather than resulting from such a change;

ix) *Referee design*: If 'audience design' is responsive (reactive) and addressed to a present second-person, 'referee design' is initiative (proactive) and addressed to an absent but highly influential third-person reference group in order to express identification;

x) *Field and object of study*: Style research requires its own designs and methodology.

Within the framework of Historical Sociolinguistics, Hernández-Campoy/García-Vidal (2018b) also found the same pattern of audience design at the end of the Middle Ages in the case of the dissappearance of the runic symbol thorn <þ> and its replacement with the Latin-based grapheme <th> in English.

4.3 Script design

In Cutillas-Espinosa/Hernández-Campoy (2006; 2007), the speech so eminently standard (92%) of a radio presenter in the traditionally non-standard local community of Murcia Region (Santomera) during his programme MQM was investigated, as well as the speech of his audience when making phone calls (mostly non-standard: ±13,4%). This diverging pattern of verbal behaviour meant our proposal of the Script Design Model as use of a professional voice. Interested in this sociolinguistic behaviour 'on air' of the radio presenter (with no audience design at all, unlike Bell's results), we contacted him for a private interview (recorded with

informed consent). As we can see now in the results of the analysis of his speech produced during the interview, his sociolinguistic behaviour was then radically different, being more local and attached to non-standard frequencies (30% standard as opposed to his 'on air' speech: 92%).

Therefore, the Script Design model stresses the need to consider not only responsive and even initiative-based performance, but also the *script*, as part of structural constraints that condition the individual linguistic behaviour in public occupations. Script takes the form of a professional voice used strictly following a particular linguistic policy which is based on canonical sociolinguistic norms and attitudes to language. This view urges us to consider community-specific structural factors anchored to linguistic norms, correctness and appropriacy restraints in the explanation of, at least, some cases of stylistic variation.

4.4 Register model

Based on Malinowski and Firth's context of situation and Bailey's polylectal grammar, Biber/Finegan (1994) developed a completely different conception of style. They argue that stylistic variation should not be considered a mirror image of interspeaker variation. Rather, they assume that "the patterns of register variation are basic and the patterns of social dialect variation result from differential access among social groups to the communicative situations and activities that promote register variation" (Finegan/Biber 1994:337). From this viewpoint, style is basically context-dependent and social class differentiation is just an echo of the different registers that are most commonly used in one's professional and personal life (see Biber 1994; 1995; Biber/Finegan 1989a; 1994; Finegan/Biber 1994; 2001; or Biber/Conrad 2009, for example).

4.5 Speaker design model

Yet, in addition to the effect of audience and script, there are other factors involved in stylistic variation. There is now a new tendency towards the development of multidimensional, multidisciplinary, and interdisciplinary approaches to style phenomena. This new approach focuses on the proactive facet of style-shifting and the individuality of speakers, where, as Giddens (1991:82-85) remarks, self-identity requires creativity and agency, and where, as Johnstone (2000:417) states, the individual voice is seen as an active agent for the transmission of sociolinguistic meaning. Linguistic variation thus becomes viewed as the verbal instrument for semiotic identificational and interactional meanings in public: a resource for identity projection and positioning in society, where individuals and individual voices are actively responsible for the transmission of sociolinguistic meaning (Giddens 1991:82-85; Johnstone 2000:417). The speaker's sociolinguistic behaviour is now conceived as inevitably based on social meaning, where language is a social practice, and style-shifting is socially motivated through its diverse linguistic resources and mechanisms. Given the strong relationship that exists between language and society, the social meaning is stressed, conceiving language not solely as a means of communicating information (oral and written), but also as a means of establishing and maintaining social relationships (building bridges between speakers). Crucially then, language production becomes a very important instrument for conveying social information about the speaker –chiefly identificational and ideological– and/or situation (Trudgill 1983; Pride 1971): "when people talk they communicate not only information but also images of themselves" (Tannen 1984/2005:3), because "language as a social phenomenon is closely tied up with the social structure and value systems of society" (Trudgill 1983:19). In addition to enabling communication and establishing social relations, language transmits social meaning through sociolinguistic variation and the choices speakers make between them. Geographical, socio-demographic, or stylistic variation conveys some kind of social meaning in terms of identity, attitudes, and/or ideology (see Hernández-Campoy 2016:51-62). The

identificational axiom iniciated by the Audience Design Model is now developed further through the Social Identity Theory: language acts *are* acts of identity, a very important symbol of group consciousness and solidarity, a signal of group identity and linguistic loyalty (Le Page/Tabouret-Keller 1985:14). In this way, language variation is understood as agentive, interactive and socially meaningful, where accents, dialects and their styling are markers of this intended social meaning (Auer 2007; Podesva 2006). This means that every single time that speakers produce an utterance, they are signalling some kind of identity (standard/non-standard), ideology (norm/usage) and attitudes (correction, adequacy, aesthetics), and the use of one variant of a variable or another expresses their social affiliation. Like any other social stereotypes, the different ways of speaking constitute prototype categories within a wider frame that comprises not only ideological components, but also markers from a wide variety of dimensions, such as speech, physical appearance, dress, dance and music (Kristiansen 2008:72-73). Styles thus represent our ability to take up different social positions (Bell 2007b:95), because styling is a powerful device for linguistic performance, rhetorical stance-taking, and identity projection. Consequently, this means that identity is dynamic and that every speech act is performance –with speakers projecting different roles in different circumstances–, because we are always displaying some particular type of image.

In this context, the philosophical thought more recently inspiring third-wave Sociolinguistics is a post-modernist social theory of knowledge known as Social Constructionism (see Hammersley 1992; Craib 1997; or Andrews 2012, among many others). Unlike Determinism, Scientism and Neopositivism, and as a reaction against them, Socio-Constructionism is essentially an anti-realist, relativist and interpretivist approach to thinking. It denies that knowledge is a direct perception of an independent and objective reality, which now appears as the product of experience and discourse. In fact, assuming that reality and elements of knowledge are not objectively given by nature and absolute laws, but subjectively constructed and institutionalised by humans in an ongoing, dynamic process, they explore how individuals and groups participate in this construction, perception and interpretation of social phenomena. As Schwandt (2003) states, this means that 'knowledge' and 'truth' are created rather than discovered by the mind, since reality is socially defined by individuals, or groups of individuals, through the subjective experience of everyday life within society and transmitted through the dialectical interaction of those individuals themselves. Society is actively and creatively produced by individuals just as individuals are the product of society; or, to put it another way, society is created by people but people are also created by society. Correspondingly, as Burr (1995) pointed out, the experience of society as a subjective reality provides human beings with an identity and a place within society itself, since "our identity originates not from inside the person but from the social realm" (Andrews 2012): much of what we are and do is the result of social and cultural influences. Gender, racial, ethnic, and age identities, for example, are thus social constructions and categorisations beyond any biological difference. Also under a strong influence of Relativism, socio-constructionist theory assumes the existence of multiple realities and also multiple interpretations of those realities. Likewise, 'truth' is a socially constructed concept, and therefore socially relative, which is the origin of current 'post-truth' concept and phenomena. Accordingly, there is no single valid methodology in science, but rather a diversity of useful methods (Schofield 2010). This means the realism-relativism polarisation and confrontation.

The most recent sociolinguistic trends are not unaware of these new philosophical and social conceptions. In this setting, styles and stylistic variation represent our ability to take up different social positions through linguistic choice (Bell 2007b:95), because style-shifting is a powerful device for linguistic performance, rhetorical stance-taking, and identity projection. With these philosophical ingredients in sytilistic variation, now the Speaker Design theory has recently emerged as a multidimensional socio-constructionist model that analyses stylistic variation in

terms of multiple co-occurring parameters –language and society conceived as co-constitutive realities– taking into account a wide range of contextual factors that might customise people's speech (see Coupland 1985; 2001a; 2001b; 2007): internal (purpose, key, frame, etc.) and external (audience, topic, setting, age, familiarity, etc.) characteristics as factors influencing speaker agency in the shaping of style or language choice. Building on individual agency, the Speaker Design Model views stylistic variation as a resource in the performance of speakers' personal and interpersonal social identity (active creation, presentation, and even recreation); in other words, stylistic variation is a resource for creating as well as projecting one's persona. Style-shifting is therefore now understood as a proactive (initiative) rather than responsive (reactive) phenomenon. Accordingly, identity is dynamic and all speech is performance – speakers projecting different roles in different circumstances– since we are always displaying some particular type of identity. Speakers do *identity work* using language to create and recreate their multiple identities, regardless of social categories, because they constantly shape, re-shape and create the situation through strategic use of language style. That is persona management, where *performativity*, *agency*, *stylisation*, *enregisterment*, *identity construction* and *authenticity* are crucial concepts (see also Coupland 2003; 2009; Johnstone 2000; 2014; Hernández-Campoy/Cutillas-Espinosa 2012a).

This theory was developed in sociolinguistic styling by Nikolas Coupland (1981; 1985) with his study on the multiple personal identity images projected by a Cardiff travel agent through her speech when addressing her clients and co-workers; or in the case of a disc jockey in a Cardiff radio station (Coupland 1985). Similarly, this phenomenon had also been observed by Trudgill (1980) in his study on the use of American vs. British working class linguistifc features in British pop-rock music bands. Whereas singers in the mainstream pop tradition showed a tendency towards the use of American features, those in the punk-rock movement, particularly Ian Dury, exhibited an exclusive tendency towards British features, in line with the self-image they wanted to project and with the profile of those fans.

Hernández-Campoy/Cutillas-Espinosa (2010; 2012; 2013) also found this speaker design phenomenon in Spain in their study on a former president of Murcia Region (María Antonia Martínez) and her use of vernacularisation (see also Cutillas-Espinosa/Hernández-Campoy/Schilling-Estes 2010). A similar situation of speaker design was found by Hernández-Campoy/García-Vidal (2018b) in a communicative situation of late medieval England making use of the Labovian *Uniformitarian Principle*. In this case, the construction of rhetorical stance and projection of identity performatively in order to achieve a particular aim takes place through the instrumentalisation of an orthographic feature in his language production: <th> versus <þ>.

5 Conclusion

The paradigm of Sociolinguistics is in a process of theoretical reformulation and redefinition as well as methodological updating in consonance with the evolution of epistemology and the development of new fieldwork methods, data collection techniques and even statistical analysis (see Eckert 2012; 2018; Cantos 2013). Similarly, stylistic models have not been unaware of the social philosophy of their time. Rather, they have benefited from a long background in social philosophy: Rhetoric in the Sophists and Aristotle, Oratory in Cicero and Quintilian, Poetics in Jakobson, Determinism in Labov's audio-monitoring, Behaviourism in Bell's Audience Design, Contextualism in Biber's Register theory, Socio-constructionism in Coupland's Speaker Design, etc. (see Hernández-Campoy 2016). They have all contributed to account for the nature, functioning, and effectiveness of style-shifting processes in social interaction.

As stated in Hernández-Campoy/Cutillas-Espinosa (2012b:7) in the epistemic evolution of Sociolinguistics since its origins in the 1960s, there has been a shift from the early deterministic and system-oriented assumptions to the recent socio-constructionist and speaker-oriented approaches to inter- and intra-speaker variation. This has meant, as underlined in Hernández-Campoy/García-Vidal (2018a), moving the focus from collectivity to individuality, from the generality of the statistical mean to the singularity of mean deviation, from accumulative patterns of behaviour of the average speaker in large-scale aggregate data to authenticity in the individual usage of the 'case study', from reactive or responsive to agentive or creative, or from responsive to initiative or proactive.

Similarly, stylistic variation studies have also experienced the same epistemic evolution in the treatment of linguistic performance, rhetorical stance, and identity projection, among other effects. Traditional variationist conceptualisations of stylistic variation as a primarily responsive phenomenon, conditioned by factors external to the speaker, such as formality of the situation or audience, have been shown to be unable to account for *all* stylistic choices. Contrarily, more recent views of stylistic variation as creative and strategic, and as essential to identity projection and creation, and the furthering of speakers' specific situational goals, can be used to explain their stylistic choices and provide us with a wider complementary perspective of the style choices speakers may make (Hernández-Campoy/Cutillas-Espinosa 2010;2012b).

Therefore, style is a complex multidimensional phenomenon that cannot be modelled in a single unidimensional theory, so stylistic studies have to progress, as Rickford/Eckert (2001:2) state, by understanding the boundaries between the three main components of sociolinguistic variation –*stylistic*, *linguistic* and *social*– as more permeable within the study of speakers' agency and performance in society, and through multidimensional, multidisciplinary and interdisciplinary approaches. We thus need to focus on the socially constructionist potential of style-shifting in order to find out how sociolinguistic variation interfaces with other dimensions of meaning-making in discourse (Coupland 2007:ix). Using Lass' (1984:8) initial example, this means playing with new wine, some old wine, some old bottles and a good number of new bottles, epistemologically speaking, obviously.

References

Alcaraz Varó, Enrique. 1990. *Tres paradigmas de la investigación lingüística*. Alcoy: Marfil.

Andrews, Tom. 2012. What is Social Constructionism? *Grounded Theory Review* 11(1). http://groundedtheoryreview.com/2012/06/01/what-is-social-constructionism/ [accessed 2014-10-15].

Auer, Peter (ed.). 2007. *Style and Social Identities: Alternative Approaches to Linguistic Heterogeneity*. Berlin: Walter de Gruyter.

Bell, Allan. 1977/1979. *The Language of Radio News in Auckland. A Sociolinguistic Study of Style, Audience and Subediting Variation*. PhD dissertation, University of Auckland. Ann Arbor, MI: University Microfilms, 1979.

Bell, Allan. 1984. Language Style as Audience Design. *Language in Society* 13(2). 145-204.

Bell, Allan. 1991. *The Language of News Media.* Oxford: Blackwell.

Bell, Allan. 2001. Back in style: Reworking Audience Design. In: Eckert, Penelope/ Rickford, John (eds.), *Style and Sociolinguistic Variation*, 139-169. Cambridge: Cambridge University Press.

Bell, Allan. 2007a. Style in Dialogue: Bakhtin and Sociolinguistic Theory. In: Bayley, Robert/ Lucas, Ceil (eds.), *Sociolinguistic Variation: Theories, Methods, and Applications*, 90-109. Cambridge: Cambridge University Press.

Bell, Allan. 2007b. Style and the Linguistic Repertoire. In: Llamas, Carmen/ Mullany, Louise/ Stockwell, Peter (eds.), *The Routledge Companion to Sociolinguistics*, 95-100. London: Routledge.

Bell, Allan. 2014. *The Guidebook to Sociolinguistics*. Malden: Wiley-Blackwell.

Biber, Douglas. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.

Biber, Douglas/ Conrad, Susan. 2009. *Register, Genre and Style*. Cambridge: Cambridge University Press.

Biber, Douglas/ Finegan, Edward. 1989a. Drift and the Evolution of English Style: A History of Three Genres. *Language* 65(3). 487-517.

Biber, Douglas/ Finegan, Edward (eds.). 1994. *Sociolinguistic Perspectives on Register*. Oxford: Oxford University Press.

Bourdieu, Pierre/ Boltanski, Luc. 1975. Le fétichisme de la langue. *Actes de la recherche en sciences sociales* 4(1). 2-32.

Bucholtz, Mary. 2003. Sociolinguistic Nostalgia and the Authentication of Identity. *Journal of Sociolinguistics* 7(3). 398-416.

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

Burr, Vivien. 1995/2003. *Social Constructionism*. London: Routledge (2nd ed.).

Cantos Gómez, Pascual. 2013. *Statistical Methods in Language and Linguistic Research*. Sheffield: Equinox Publishing.

Coupland, Nikolas. 1981. *The Social Differentiation of Functional Language Use: A Sociolinguistic Investigation of Travel Agency Talk*. PhD. dissertation, University of Wales Institute of Science and Technology.

Coupland, Nikolas. 1985. Hark, Hark the Lark: Social Motivations for Phonological Style-shifting. *Language & Communication* 5(3). 153-172.

Coupland, Nikolas. 2001a. Language, Situation, and the Relational Self: Theorising Dialect-style in Sociolinguistics. In Eckert, Penelope/ Rickford, John (eds.), *Style and Sociolinguistic Variation*, 185-210. Cambridge: Cambridge University Press.

Coupland, Nikos. 2001b. Dialect Stylization in Radio Talk. *Language in Society* 30(3). 345-375.

Coupland, Nikoas. 2003. Sociolinguistic Authenticities. *Journal of Sociolinguistics* 7(3). 417-431.

Coupland, Nikoas. 2007. *Style: Language Variation, and Identity*. Cambridge: Cambridge University Press.

Craib, Ian. 1997. Social Constructionism as a Social Psychosis. *Sociology* 31(1). 1-15.

Cutillas-Espinosa, Juan Antonio/ Hernández-Campoy, Juan Manuel. 2006. Nonresponsive Performance in Radio Broadcasting: A Case Study. *Language Variation and Change* 18(3). 1-14.

Cutillas-Espinosa, Juan Antonio/ Hernández-Campoy, Juan Manuel. 2007. Script Design in the Media: Radio Talk Norms behind a Professional Voice. *Language & Communication* 27(2). 127-152.

Cutillas-Espinosa, Juan Antonio/ Hernández-Campoy, Juan Manuel/ Schilling-Estes, Natalie. 2010. Hyper-vernacularisation in a Speaker Design Context: A Case Study. *Folia Linguistica* 44(1). 1-22.

Eckert, Penelope. 2012. Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology* 41. 87-100.

Eckert, Penelope. 2018. *Meaning and Linguistic Variation: The Third Wave in Sociolinguistics*. Cambridge: Cambridge University Press.

Eckert, Penelope/ Rickford, John (eds.). 2001. *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press

Figueroa, Esther. 1994. *Sociolinguistic Metatheory*. Oxford: Pergamon.

Finegan, Esther/ Biber, Douglas. 1994. Register and Social Dialect Variation: An Integrated Approach. In: Biber, Douglas/ Finegan, Edward (eds.), *Sociolinguistic Perspectives on Register*, 315-347. Oxford/New York: Oxford University Press.

Finegan, Edward/ Biber, Douglas. 2001. Register Variation and Social Dialect Variation: The Register Axiom. In: Eckert, Penelope/ Rickford, John. (eds.), *Style and Sociolinguistic Variation*, 235-267. Cambridge: Cambridge University Press.

Gadet, Françoise. 2005. Research on Sociolinguistic Style. In: Ammon, Ulrich/ Dittmar Norbert/ Mattheier Klaus J./ Trudgill, Peter (eds.). *Sociolinguistics: An International Handbook of the Science of Language and Society*, 2 vols., 1353-1361. Berlin: Walter de Gruyter.

Giddens, Anthony. 1991. *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Cambridge: Polity.

Giles, Howard. 1979. Sociolinguistics and Social Psychology: An Introductory Essay. In: Giles, Howard/ St. Clair, Robert N. (eds.), *Language and Social Psychology*, 1-20. Oxford: Blackwell.

Hammersley, Martyn. 1992. *What's Wrong with Ethnography?* London: Routledge.

Hernández-Campoy, Juan Manuel. 2014. Research Methods in Sociolinguistics. *AILA Review* 27(1). 5-29.

Hernández-Campoy, Juan Manuel. 2016. *Sociolinguistic Styles*. Malden: Wiley-Blackwell.

Hernández-Campoy, Juan Manuel/ Almeida, Manuel. 2005. *Metodología de la investigación sociolingüística*. Granada: Comares.

Hernández-Campoy, Juan Manuel/ Cutillas-Espinosa, Juan Antonio. 2010. Speaker Design Practices in Political Discourse: A Case Study. *Language and Communication* 30(4). 297-309.

Hernández-Campoy, Juan Manuel/ Cutillas-Espinosa, Juan Antonio (eds.). 2012a. *Style-Shifting in Public: New Perspectives on Stylistic Variation*. Amsterdam: John Benjamins.

Hernández-Campoy, Juan Manuel/ Cutillas-Espinosa, Juan Antonio. 2012b. Introduction. In: Hernández-Campoy, Juan Manuel/ Cutillas-Espinosa, Juan Antonio (eds.), *Style-Shifting in Public: New Perspectives on Stylistic Variation*, 1-18. Amsterdam: John Benjamins.

Hernández-Campoy, Juan Manuel/ Cutillas-Espinosa, Juan Antonio 2012c. Speaker Design Strategies in Political Contexts of a Dialectal Community. In: Hernández-Campoy, Juan Manuel/ Cutillas-Espinosa, Juan Antonio (eds.), *Style-Shifting in Public: New Perspectives on Stylistic Variation*, 21-43. Amsterdam: John Benjamins.

Hernández-Campoy, Juan Manuel/ Cutillas-Espinosa, Juan Antonio. 2013. The Effects of Public and Individual Language Attitudes on Intra-speaker Variation: A Case Study of Style-shifting. *Multilingua* 32(1). 79-101.

Hernández-Campoy, Juan Manuel/ García-Vidal, Tamara. 2018a. Persona Management and Identity Projection in English Medieval Society: Evidence from John Paston II. *Journal of Historical Sociolinguistics* 4(1). 1-31.

Hernández-Campoy, Juan Manuel/ García-Vidal, Tamara. 2018b. Style-Shifting and Accommodative Competence in Late Middle English Written Correspondence: Putting Audience Design to the Test of Time. *Folia Linguistica Historica* 39(2). 383-420.

Hernández Campoy, Juan Manuel/ Jiménez Cano, José Maria. 2003. Broadcasting Standardisation: An Analysis of the Linguistic Normalisation Process in Murcia. *Journal of Sociolinguistics* 7(3). 321-347.

Jaffe, Alexandra (ed.). 2009. *Stance: Sociolinguistic Perspectives.* Oxford: Oxford University Press.

Johnstone, Barbara. 2000. The Individual Voice in Language. *Annual Review of Anthropology* 29. 405-425.

Johnstone, Barbara. 2001. The Individual. In: Duranti, Alessandro (ed.), *Key Terms in Language and Culture*, 122-125. Malden, MA: Blackwell.

Johnstone, Barbara. 2014. '100% Authentic Pittsburgh': Sociolinguistic Authenticity and the Linguistics of Particularity. In: Lacoste, Veronique/ Leimgruber, Jakob/ Breyer, Thiemo (eds.), *Indexing Authenticity: Sociolinguistic Perspectives*, 97-112. Berlin: De Gruyter.

Jones, Huw. 1981/1990. *Population Geography*. 2nd edn. London: Paul Chapman Publishing.

Koch, Peter/ Oesterreicher, Wulf. 1994. Schriftlichkeit und Sprache. In: Günther, Hartmut/ Ludwig, Otto (eds.), *Schrift und Schriftlichkeit/Writing and its Use. Ein interdisziplinäres Handbuch internationaler Forschung/An Interdisciplinary Handbook of International Research*, 587-604. Berlin/New York: Mouton de Gruyter.

Koch, Peter/ Oesterreicher, Wulf. 2012. Language of Immediacy – Language of Distance: Orality and Literacy from the Perspective of Language Theory and Linguistic History. In: Lange, Claudia/ Weber, Beatrix/ Wolf, Göran (eds.), *Communicative Spaces: Variation, Contact, and Change – Papers in Honour of Ursula Schaefer,* 441-473. Bern: Peter Lang.

Kristiansen, Gitte. 2008. Style-shifting and Shifting Styles: A Socio-cognitive Approach to Lectal Variation. In: Kristiansen, Gitte/ Dirven, René (eds.), *Cognitive Sociolinguistics*, 45-88. Berlin: Mouton de Gruyter.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.

Labov, William. 1966/2006. *The Social Stratification of English in New York City*. 2nd edn. Cambridge: Cambridge University Press.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 2001. *Principles of Linguistic Change II: Social Factors*. Oxford: Blackwell.

Lass, Roger. 1984. *Phonology: An Introduction to Basic Concepts*. Cambridge: Cambridge University Press.

Le Page, Robert B./ Tabouret-Keller, Andrée. 1985. *Acts of Identity: Creole-based Approaches to Language and Ethnicity*. Cambridge: Cambridge University Press.

Mackenzie, John Stuart. 1890/2006. *An Introduction to Social Philosophy*. New York: Elibron Classics.

Marková, Ivona. 1982. *Paradigm, Thought and Language*. London: John Wiley & Sons.

Meyerhoff, Miriam. 2006. *Introducing Sociolinguistics*. London: Routledge.

Myers-Scotton, Carol. 1998. A Theoretical Introduction to the Markedness Model. In: Myers-Scotton, Carol (ed.), *Codes and Consequences: Choosing Linguistic Varieties*, 18-38. Oxford/New York: Oxford University Press.

Olssen, Mark. 2010. Social Democracy, Complexity and Education: Sociological Perspectives from Welfare Liberalism. In: Apple, Michael W./ Ball, Stephen J./ Gandin, Luis A. (eds.), *The Routledge International Handbook of the Sociology of Education*, 75-89. London: Routledge.

Oyelaran, Olasope. 1970. Aspects of Linguistic Theory in Firthian Linguistics. *Word* 23(1-3). 428-52.

Podesva, Robert J. 2006. *Phonetic Detail in Sociolinguistic Variation: Its Linguistic Significance and Role in the Construction of Social Meaning*. PhD. dissertation, Stanford University.

Pride, John B. 1971.*The Social Meaning of Language*. London: Oxford University Press.

Rickford, John/ Eckert, Penelope. 2001. Introduction. In: Eckert, Penelope/ Rickford, John (eds.), *Style and Sociolinguistic Variation,* 1-18. Cambridge: Cambridge University Press.

Sankoff, David/ Laberge, Suzanne. 1978. The Linguistic Market and the Statistical Explanation of Variability. In: Sankoff, David (ed.), *Linguistic Variation: Models and Methods*, 239-250. New York: Academic Press.

Schilling, Natalie. 2013. Investigating Stylistic Variation. In: Chambers, J.K./ Schilling, Natalie (eds.), *The Handbook of Language Variation and Change*, 327-349. 2nd edn. Oxford: Blackwell.

Schwandt, Thomas A. 2003. Three Epistemological Stances for Qualitative Inquiry: Interpretivism, Hermeneutics and Social Constructionism. In: Denzin, Norman/ Lincoln, Yvonna (eds.), *The Landscape of Qualitative Research: Theories and Issues*, 292-331. Thousand Oaks, CA: SAGE.

Tajfel, Henri. 1978. Interindividual Behaviour and Intergroup Behaviour. In: Tajfel, Henri (ed.), *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations*, 27-60. London: Academic Press.

Tajfel, Henri. 1979. Individuals and Groups in Social Psychology. *British Journal of Social and Clinical Psychology* 18. 183-190.

Tannen, Deborah. 1984/2005. *Conversational Style: Analyzing Talk among Friends*. Oxford: Oxford University Press.

Trudgill, Peter. 1974. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Trudgill, Peter. 1980. Acts of Conflicting Identity: A Sociolinguistic Look at British Pop Songs. *York Papers in Linguistics* 9. 261-277.

Trudgill, Peter. 1983. *Sociolinguistics: An Introduction to Language and Society*. 2nd edn. Harmondsworth: Penguin.

Trudgill, Peter. 2003. *A Glossary of Sociolinguistics*. Edinburgh: Edinburgh University Press.

Tucker, G. Richard. 1997. The development of sociolinguistics as a field of study: concluding observations. In: Paulston, Christina B./ Tucker, G. Richard (eds.), *The Early Days of Sociolinguistics: Memoires and Reflections*, 325-362. Dallas, TX: Summer Institute of Linguistics.

Ure, Jean. 1982. Introduction: Approaches to the Study of Register Range. *International Journal of the Sociology of Language* 35. 5-23.

Williams, Glyn. 1992. *Sociolinguistics: A Sociological Critique*. London: Routledge.

Variation und Wandel der Pluralformen von *sein* in den Dialekten Salzburgs

Lars Bülow (Universität Salzburg)
lars.buelow@sbg.ac.at

Abstract

Der Beitrag befasst sich mit der Variation und Entwicklung der Pluralformen (Präsens, Indikativ, aktiv) des Verbs sein *in den verschiedenen Dialektregionen des Bundeslandes Salzburg. Es werden sowohl Ergebnisse aus einer* real-time- *als auch einer* apparent-time-*Studie präsentiert, die im Rahmen des SFB-Projekts (FWF F60) „Deutsch in Österreich" (DiÖ) durchgeführt wurden. Die Daten stammen aus verschiedenen Erhebungskontexten, z. B. dem „Sprachatlas Salzburg" (Scheutz 2017) und dem DiÖ-Projekt (Budin u. a. 2019). Neben der* h/s-*Anlautalternation sowie der* a/ai-*Stammvokalvariation in der 2.Ps.Pl. wird besonders die Variation und Entwicklung der Suffigierung der 1./3.Ps.Pl. aufgezeigt und diskutiert. Die Ergebnisse bieten breite Evidenz dafür, dass Wandel in den bairischen Dialekten Österreichs heute im Wesentlichen auf Advergenz zu überregionalen Varietäten (Regiolekte, Standardvarietäten) zurückzuführen ist und kaum mehr im Sinne ‚innersprachlicher' Ausgleichserscheinungen modelliert werden kann.*

Keywords: *Dialektwandel in Österreich, Advergenz, verbale Pluralmorphologie,* real-time-*und* apparent-time-*Studie*

1 Einleitung[1]

Das Verb *sein* ist bekanntlich ein sogenanntes Wurzelverb und zählt zur Klasse der besonderen Verben (vgl. Paul [25]2007: 279 f.). Seine Formen bilden heute in den Varietäten des Deutschen suppletive Paradigmen, welche diachron, diatopisch sowie auch diastratisch und diaphasisch betrachtet unterschiedlich ausgeprägt sein können und in deren Geschichte permanente Reorganisationsprozesse stattfinden (vgl. Koch 2004; Philipp/Weider 2002; Nübling 2000; Scheuringer 1990). Die paradigmatischen Reorganisationsprozesse sind innersprachlich zumeist durch phonetisch-phonologischen und/oder morphologischen Wandel einzelner Formen im Paradigma bedingt (vgl. u. a. Wurzel 1984), die zu Reanalysen und/oder Analogiebildungen in anderen Formen des Paradigmas führen.

Der Fokus dieses Beitrags liegt auf der Beschreibung und Diskussion der basisdialektalen Pluralformen des Verbs *sein* (Präsens, Indikativ, aktiv) in den verschiedenen Dialektregionen des Bundeslandes Salzburg.[2] Für diesen Beitrag gilt: „Auch das Bairische ging bzw. geht seinen eigenen Weg in der Reorganisation des Paradigmas" (Koch 2004: 121). Die genaue Betrachtung des Pluralparadigmas erscheint besonders lohnend zu sein, weil sich hier zunächst synchron sowohl diatopisch als auch im Dialekt-Standard-Gefüge interessante Variation hinsichtlich verschiedener Variablen zeigt (vgl. Scheutz 2017; Bülow/Wallner im Druck), die ein wesentlicher Indikator für rezenten Sprachwandel ist.

[1] Für hilfreiche Hinweise und die Bereitstellung von Daten möchte ich mich bei Stephan Elspaß, Günter Koch, Hannes Scheutz und Dominik Wallner herzlich bedanken.

[2] Die Daten des Sprachatlas Salzburg (vgl. Scheutz 2017) umfassen Altsalzburg und damit auch vier deutsche Ortschaften im Berchtesgadener Land (Schönau am Königssee, Petting, Teisendorf, Surheim). Diese Daten werden hier aufgrund ihrer Verfügbarkeit mitausgewertet und auf den Karten gezeigt, ohne dass in der Diskussion genauer auf sie eingegangen wird. Im Folgenden wird generalisierend von den Dialekten oder Dialektregionen „Salzburgs" oder „in Salzburg" gesprochen; gemeint ist dann das Bundesland Salzburg.

Ziel dieses Beitrags ist es, Variation und Wandel der folgenden drei Variablen von *sein* mithilfe eines *real*- und eines *apparent*-time-Vergleichs zu erfassen und mögliche Erklärungen zu diskutieren:

- *h*/*s*-Anlautalternation im Pluralparadigma (z. B. (*mia*) *han* vs. *san*)
- *a*/*ai*-Stammvokalvariation in der 2.Ps.Pl. (z. B. (*es*) *hadds* vs. *haidds*)
- Variation bei der Suffigierung der 1./3.Ps.Pl. (z. B. (*mia*) *han* vs. *hand*; (*se*) *hand* vs. *han*)

In dieser Untersuchung werden mehrere Variablen der Pluralformen von *sein* in den Blick genommen, um mögliche Evidenzen für horizontal-diatopische und/oder ‚vertikale', d. h. auf das Dialekt-Standard-Gefüge bezogene Wandelerscheinungen besser abzusichern. Diese gestalten sich in der vorliegenden diaglossischen Sprachsituation im bairischen Teil Österreichs (vgl. Auer 2005: 22) als komplexer Prozess, in denen horizontale und vertikale Einflüsse in der Regel nur analytisch auseinandergehalten werden können (vgl. Bülow/Scheutz/Wallner 2019): „Usually, these two developments go hand in hand, leading to leveling" (Auer 2018: 159).

Abschnitt 2 schildert zunächst die Dialektsituation in Salzburg, bevor in Abschnitt 3 genauer auf die sprachlichen Variablen und den Forschungsstand dazu eingegangen wird. Abschnitt 4 erläutert dann das methodische Vorgehen. In Abschnitt 5 werden die Ergebnisse dargelegt. In Abschnitt 6 werden diese dann ausführlich diskutiert. Der Beitrag endet mit einem kurzen Fazit (Abschnitt 7).

2 Salzburger Dialektregionen

Die Formen von *sein* sind für die bairischen Dialekte prinzipiell gut erfasst, sei es durch eine Vielzahl von Ortsgrammatiken (vgl. Schatz 1897; Lessiak 1903) oder die großen Dialektatlasprojekte, die seit den 1980er Jahren initiiert wurden (vgl. Sprachatlas von Niederbayern (SNiB), Sprachatlas von Oberbayern (SOB), Sprachatlas von Oberösterreich (SAO)). Die vorliegenden Analysen konzentrieren sich auf den Salzburger Raum. Dieser Fokus hat im Wesentlichen zwei Gründe: Erstens kann auf eine einzigartige zeitlich tiefe und räumlich engmaschige Datengrundlage zurückgegriffen werden (vgl. Abschnitt 4). Zweitens durchziehen Salzburg drei größere Dialektregionen (vgl. Wiesinger 1983), deren Betrachtung es ermöglicht, horizontalen Dialektkontakt in den Blick zu nehmen. Im Norden Salzburgs (im Flachgau) finden sich westmittelbairische Dialekte, ganz im Südosten (im Lungau) dominieren südbairische Dialektmerkmale, und die Gebiete dazwischen (Tennengau, Pongau und Pinzgau) zählen zum südmittelbairischen Übergangsgebiet (vgl. Abbildung 1).

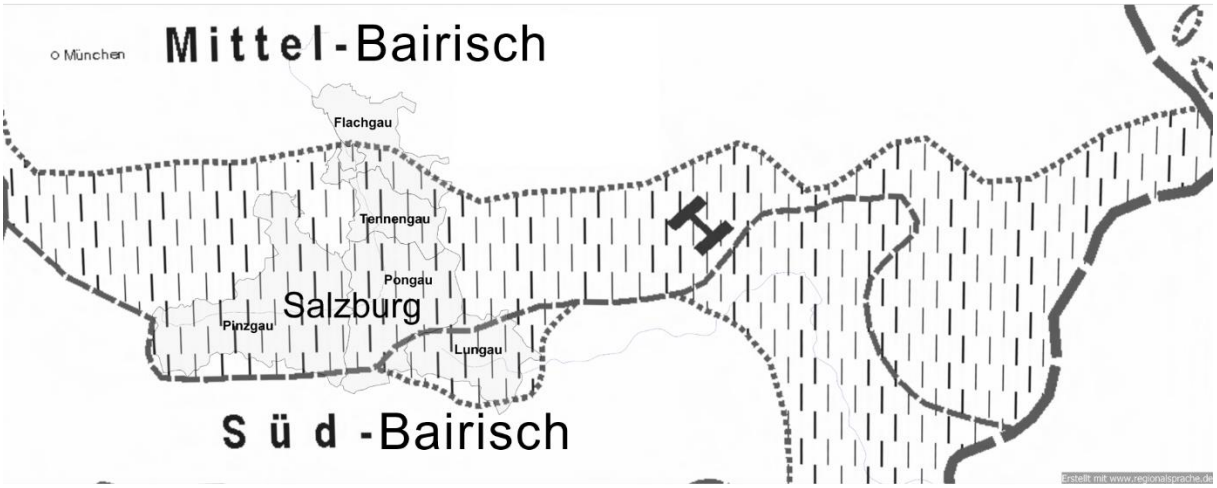


Abbildung 1: Bairische Dialektregionen in Salzburg in Anlehnung an Wiesinger (1983); die gestrichelten Areale sind – mit Ausnahme des Lungaus (Südbairisch) – dem Südmittelbairischen zuzuordnen

Ein wesentliches Ziel dieses Beitrags ist es zu untersuchen, ob die Variationsmuster bei der Pluralbildung von *sein* dialektregionenspezifisch sind. Im Folgenden werden zunächst die Variablen und Varianten dieser Untersuchung erläutert.

Obwohl die Formen von *sein* in allen gängigen direkten Dialekterhebungen abgefragt werden, gibt es erstaunlich wenige systematische Studien, die sich dezidiert mit der rezenten Entwicklung auseinandersetzen (vgl. z. B. Philipp/Weider 2002; Koch 2004; Mauser 2007). Weil es m.W. für Österreich diesbezüglich keine größeren und systematisch angelegten diachronen und/oder diatopischen Untersuchungen gibt, wird als Vergleichsgrundlage für die Erläuterung der einzelnen Variablen auf die von Koch (2004: 128) und Mauser (2007: 70) skizzierten Pluralparadigmen im Bairischen zurückgegriffen (vgl. Tabelle 1).

Tabelle 1: Pluralparadigmen von *sein* nach Koch (2004: 128) und Mauser (2007: 70)

Ps.Pl.	**Mhd. *sîn***	**Nhd. Standard *sein***	**Bair. *sa(e)*** ***s*-Anlaut**	**Bair. *sa(e)*** ***h*-Anlaut**
1.	*birn, sîn*	*sind*	*san(d)* / *sama*	*han(d)* / *hama*
2.	*birt, sît*	*seid*	*sats* / *saidds*	*hats* / *haidds*
3.	*sint*	*sind*	*san(d)*	*han(d)*

Die verschiedenen Paradigmen verdeutlichen sowohl Sprachwandel (vgl. Mhd. vs. Nhd.) als auch Unterschiede im Dialekt-Standard-Gefüge (vgl. Bair. vs. nhd. Standard). Die Unterschiede, die für die bairischen Dialekte beschrieben werden (vgl. z. B. bair. *s*-Anlaut vs. *h*-Anlaut und *a*-Stammvokal vs. *ai*-Stammvokal), deuten bereits die diatopische (und diastratische sowie diaphasische) Variation für die hier untersuchten Variablen an. Im Folgenden werden zunächst die beiden lautlichen Variablen genauer beschrieben (3.1, 3.2), bevor schließlich ausführlich auf die morphologische Variation im Paradigma fokussiert wird (3.3). An jede dieser drei Darstellungen knüpft jeweils eine Untersuchungshypothese an.

3.1 *h*/*s*-Anlautalternation

Die *h*/*s*-Variation im Anlaut der Pluralformen von *sein* ist ein distinktives Merkmal zur Abgrenzung der westmittelbairischen Dialekte von den übrigen bairischen Dialektregionen (vgl. Wiesinger 2004: 24). Während für die westmittelbairischen Basisdialekte der *h*-Anlaut

charakteristisch ist, zeigen ostmittelbairische, südmittelbairische und südbairische Dialekte *s*-Anlaut. Scheuringer (1990: 322) verortet den *h*-Anlaut im Dreieck zwischen Nürnberg, München und Linz. Das entspricht ungefähr der Isoglosse, die Wiesinger (2004: 22) auf der Grundlage von Daten aus dem Deutschen Sprachatlas (Karten 108–111) zeichnet (vgl. Abbildung 2).

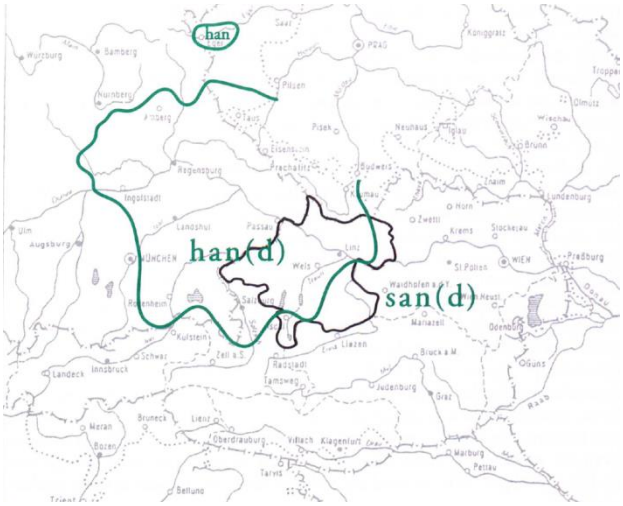


Abbildung 2: Diatopische Verteilung der *h*/*s*-Anlautalternation im Bairischen nach Wiesinger (2004: 22); die durchgezogenen grünen Linien umfassen die Gebiete, in denen der *h*-Anlaut dominiert[3]

Die Dominanz des *h*-Anlauts für die westmittelbairischen Dialekte wird auch in den Daten des SNiB (vgl. Koch 2007: 392), des SAO (vgl. Band II Karte 36) und in der Untersuchung von Mauser (2007: 70) deutlich.

Die Daten des SNiB deuten aber schon insofern eine soziolinguistische Dimension an, dass in den urbanen Zentren wie Passau bereits häufiger *s*-Anlaut zu finden ist (vgl. Koch 2007: 393). Dieser Befund spiegelt sich auch in den Daten von Mauser (2007) wider. Diese zeigen zwar größtenteils den *h*-Anlaut für den westmittelbairischen Teil Salzburgs, in städtischer Umgebung wie in Oberndorf bei Salzburg (zusammen mit Laufen hat Oberndorf ca. 14.000 Einwohner) erscheint aber auch *s*-Anlaut.

Für die Argumentation ist wichtig hervorzuheben, dass der *h*-Anlaut ein rein basisdialektales Merkmal ist, während der *s*-Anlaut auch dort als regiolektales Merkmal erscheint, wo im Basisdialekt der *h*-Anlaut dominiert. Diese vertikal-soziolinguistische Variation wird von Scheuringer (1990; 1993) für die in der westmittelbairischen Dialektregion liegende Kleinstadt Braunau in Oberösterreich (ca. 16.500 Einwohner) herausgearbeitet. Der *h*-Anlaut wurde in seiner Untersuchung nur noch von Personen verwendet, die der „Grundschicht" angehörten (in 42,6% der Fälle). Personen aus der „Mittelschicht" und „höheren Schicht" verwendeten fast ausnahmslos den *s*-Anlaut (vgl. Scheuringer 1990: 289–327; Scheuringer 1993: 76 f). Auch das Alter der Gewährspersonen schien in dieser Studie eine wichtige Rolle zu spielen. Die älteste Gewährspersonengruppe (> 65 Jahre) zeigte noch in 35,2% der Fälle den *h*-Anlaut, die Gewährspersonengruppe im mittleren Alter (35–65 Jahre) verwendete diesen noch zu 23,5%, die jüngste Gruppe (< 35 Jahre) gebrauchte nur noch in 2,3% der Fälle den *h*-Anlaut.

Im Zusammenhang mit der Entwicklung der *h*/*s*-Variation im Anlaut macht Koch (2004: 123; 2007: 392) auf einen weiteren Aspekt aufmerksam: die 1.Ps.Pl. von *sein* mit *h*-Anlaut (und

[3] Dass die durchgezogene schwarze Linie die Grenzen des an das Bundesland Salzburg östlich angrenzenden Bundeslands Oberösterreich zeigt, ist dem Umstand geschuldet, dass sich der Aufsatz von Wiesinger (2004) der Dialektgeographie Oberösterreichs widmete.

-ma-Suffix; vgl. Tabelle 1) entspricht ausdrucksseitig weitestgehend der Pluralform von *haben* (*mia hama*). Es gibt zwar qualitative Unterschiede in der Aussprache des *a*-Stammvokals (vorderes [a] vs. hinteres [ɑ]), die dafür sorgen, dass die Formen der 1.Ps.Pl. von *sein* und *haben* nicht gänzlich homophon sind (vgl. Koch 2004: 128); dieser Kontrast – obwohl er ein Minimalpaar bilden kann – ist im Grunde aber nicht besonders ausgeprägt. Koch (2004: 130) argumentiert daher, dass die Sprecher/innen langfristig betrachtet eher zum *s*-Anlaut tendieren könnten, um die Formen dieser beiden frequenten Verben für die Hörer unterscheidbar zu halten. Der *s*-Anlaut ermöglicht im Sinne der Hörerbedürfnisse einen starken phonetischen Kontrast zum *h*-Anlaut. Für die Salzburger Dialekte muss allerdings beachtet werden, dass Pluralformen mit *-ma*-Suffix nur vereinzelt im salzburgisch-bayerischen Grenzgebiet belegt sind; im Südmittelbairischen und Südbairischen spielt die Form *hama* (1.Ps.Pl.) keine Rolle, auch weil hier natürlich der *s*-Anlaut sowohl basisdialektal als auch regiolektal gut verankert ist.

Für die hier analysierten Daten erscheint eine andere Entwicklungsrichtung plausibler, nämlich die Dialekt-zu-Standard-Advergenz, die für eine allmähliche Tendenz in Richtung des im Standard üblichen *s*-Anlauts sprechen würde. Die Dialekt-zu-Standard-Advergenz ist als dominante Entwicklungsrichtung rezenten Dialektwandels in den bairischen Teilen Österreichs beschrieben (vgl. das Modell von Auer 2005: 22) und wird u. a. in den *apparent-time*-Ergebnissen von Scheuringer (1990; 1993) deutlich.[4] Vor dem Hintergrund einer allgemeinen Dialekt-zu-Standard-Advergenz wird erwartet, anhand von *real*- und *apparent-time*-Effekten eine Entwicklung zum *s*-Anlaut in den Daten beobachten zu können.

Hypothese 1: In den westmittelbairischen Gebieten Salzburgs findet ein Wandel vom *h*- zum *s*-Anlaut statt.

3.2 *a*/*ai*-Stammvokalvariation

Während die *h*/*s*-Anlautvariation bereits Gegenstand von einigen Untersuchungen war, wurde die *a*/*ai*-Variation im Stammvokal der 2.Ps.Pl. von *sein* bisher kaum eingehend analysiert bzw. diskutiert. Die *a*/*ai*-Stammvokalvariation lässt sich als Merkmal insbesondere zur großräumigen Unterscheidung von mittel- und süd(mittel)bairischen Dialekten heranziehen. Der SAO (Band II Karte 36) zeigt für die mittelbairischen Dialekte in Oberösterreich und im angrenzenden Salzburg fast ausschließlich den Monophthong (z. B. *hadds* / *sadds*). Lediglich in der süd-westlichen und süd-östlichen Peripherie der Stadt Salzburg im Kontakt zum südmittelbairischen Übergangsgebiet sind Diphthonge (z. B. *haidds* / *saidds*) im SAO verzeichnet. Auch die Verhältnisse für den südbairischen Teil Salzburgs scheinen relativ klar zu sein: Mauser (1998: 318) hat für den Lungau ausschließlich Diphthonge dokumentiert (*saidds*). Für die südmittelbairischen Dialektregionen Salzburgs sind bezüglich dieses Merkmals bisher keine Auswertungen durchgeführt worden.

Auch für die Entwicklung der Stammvokalvariation ist die vertikale Dimension in die Betrachtungen einzubeziehen. Im Vergleich zur *h*/*s*-Anlautvariation ist die Situation für den Stammvokal der 2.Ps.Pl. im Dialekt-Standard-Kontinuum allerdings komplexer. Im Standard ist ein Diphthong anzusetzen (*ihr seid*), während im Regiolekt ein Monophthong (*es sadds*) vorherrscht (vgl. Abbildung 3).

[4] Dieser Entwicklungstrend wird auch durch die Studien von Bülow u. a. (2019), Bülow/Wallner/Scheutz (2019) und Bülow/Wallner (im Druck) bestätigt.

Standard	/ai/-Diphthong	
Regiolekt	/a/-Monophthong	
Basisdialekt	/a/-Monophthong	/ai/-Diphthong
Dialektregion	**westmittelbairisch**	**süd(mittel)bairisch**

Abbildung 3: Variation des Stammvokals im Dialekt-Standard-Kontinuum (vereinfacht)

Die regiolektale Form *sadds* scheint sich insbesondere in den städtischen ‚Umgangssprachen' von z. B. Wien, Linz, St. Pölten und Salzburg durchgesetzt zu haben, von wo sie ins jeweilige Umland ausstrahlt. Sie hat im Alltag ein höheres Prestige als die Standardform *seid*. Mit Blick auf die Dialekt-Standard-Advergenz wird daher angenommen, dass sich für die 2.Ps.Pl. Formen mit Monophthong in der südmittelbairischen Übergangszone des Bundeslands Salzburg ausbreiten; im südbairischen Raum werden aufgrund der Beharrsamkeit dieser Dialekte (vgl. Mauser 1998) vorrangig Diphthonge erwartet.

Hypothese 2: In den südmittelbairischen Gebieten Salzburgs findet ein Wandel vom *ai*-Diphthong zum *a*-Monophthong im Stammvokal der 2.Ps.Pl. statt.

3.3 Morphologische Variation im Pluralparadigma

Für die Varietäten des Deutschen müssen drei Pluralparadigmen unterschieden werden: Dreiformenplural, Zweiformenplural und Einformenplural (vgl. Tabelle 2).

Tabelle 2: Pluralparadigmen in deutschen Varietäten am Beispiel des Verbs *nehmen*

Pluralparadigmen	**Varietät**	**'nehmen'**	**Kategorie**	**Suffix**	**Label**
Dreiformenplural	Mittelhochdeutsch (Mhd.)	*nëm-en*	1.Ps.Pl.	*-en*	A
		nëm-et	2.Ps.Pl.	*-et*	B
		nëm-ent	3.Ps.Pl.	*-ent*	C
Zweiformenplural	Standard	*nehm-en*	1.Ps.Pl.	*-en*	A
		nehm-t	2.Ps.Pl.	*-t*	B
		nehm-en	3.Ps.Pl.	*-en*	A
	bairische Umgangssprache	*nem-en*	1.Ps.Pl.	*-en*	A
		nem-dds	2.Ps.Pl.	*-dds*	B
		nem-en	3.Ps.Pl.	*-en*	A
Einformenplural	Ostalemannisch	*nem-et*	1.Ps.Pl.	*-et*	B
		nem-et	2.Ps.Pl.	*-et*	B
		nem-et	3.Ps.Pl.	*-et*	B

Der Dreiformenplural ist dabei als Ausgangsbasis anzusetzen. Im Mhd. hatten die schwachen und starken Verben für alle drei Personen unterschiedliche Suffixe bzw. Morpheme. Nur die sogenannten Präteritopräsentia wie *wollen*, *können* und *müssen* zeigten schon im Mhd. einen Zweiformenplural, bei dem die 1./3.Ps.Pl. ausdrucksseitig dasselbe Suffix (*-en*) aufwiesen (vgl. Paul [25]2007: M94). Diese häufig genutzte Verbklasse könnte somit als wichtiges Vorbild auf

dem Weg zum Zweiformenplural im Standard bei den schwachen und starken Verben gedient haben. Während sich im Standard der Zweiformenplural durchgesetzt hat, gibt es in den deutschen Dialekten alle drei Paradigmentypen. Der Einformenplural ist beispielsweise in bestimmten alemannischen Dialekten präsent (vgl. Christen/Glaser/Friedli 2013: 296 f.); der Zwei- und der Dreiformenplural kommen in den bairischen Dialektregionen vor (vgl. Bülow/Scheutz/Wallner 2019).

Für die Basisdialekte in Salzburg ist zunächst davon auszugehen, dass diese den mhd. Dreiformenplural fortsetzen (vgl. Mauser 2007: 67; Wiesinger 1989), wobei sich heute „nur noch die konservativsten Teillandschaften durch Trennung von {en}-Endungen in der 1. Person des Plurals und {ent}-Endungen in der 3. Person auszeichnen“ (Scheuringer 1993: 78). Das können zumindest Mausers (1998) Befunde für den Lungau belegen. Andere empirische Belege deuten allerdings auf eine komplizierte Situation für die west- und südmittelbairischen Dialekte Salzburgs hin (vgl. Mauser 2007; Scheuringer 1993: 78), in denen verschiedene Varianten sowohl des Drei- als auch des Zweiformenplurals auch für das Verb *sein* unterschieden werden müssen (vgl. Tabelle 3).

Tabelle 3: Varianten des Drei- und Zweiformenplurals in den Salzburger Basisdialekten

		Dreiformenplural				**Zweiformenplural**			
NUM	PS	Variante 1		Variante 2		Variante 3		Variante 4	
PL	1. (*mi(a)*)	*han* *san*	A	*hand* *sand*	C	*han* *san*	A	*hand* *sand*	C
	2. (*es*)	*hadds* / *haidds* *sadds* / *saidds*	B	*hadds* / *haidds* *sadds* / *saidds*	B	*hadds* / *haidds* *sadds* / *saidds*	B	*hadds* / *haidds* *sadds* / *saidds*	B
	3. (*se*)	*hand* *sand*	C	*han* *san*	A	*han* *san*	A	*sand* *sand*	C

Variante 1 entspricht dem Dreiformenplural des mhd. Ausgangstyps. Statt des mhd. Suffixes *-et* für die 2.Ps.Pl. hat sich durch Klitisierung mit dem alten Pronomen (*-et + es = -edds*) allerdings die Form *-(e)dds* durchgesetzt. Lediglich im sprachlich archaisch geprägten Lungau ist mitunter noch *-(e)t* belegt (vgl. Mauser 1998: 318). Da *-(e)dds* im gesamten Untersuchungsgebiet relativ stabil erscheint, wird auf dieses Suffix für die folgenden Varianten nicht weiter eingegangen. Die für diesen Beitrag relevante Variation findet im Bereich der 1./3.Ps.Pl. statt (vgl. Tabelle 3).

Bei Variante 1 unterscheiden sich die Suffixe für die 1. und 3. Person dadurch, dass der Auslaut für das Suffix der 3.Ps.Pl. dentalhaltig, der der 1. Ps.Pl. hingegen dentallos ist. Bei dem Dreiformenplural in Variante 2 wird das dentalhaltige Suffix für die 1.Ps.Pl. verwendet, während das dentallose Suffix nun für die 3.Ps.Pl. steht. Variante 3 stimmt strukturell im Wesentlichen mit dem Typ des Zweiformenplurals überein, der im Regiolekt und im Standard Verwendung findet (vgl. Tabelle 2 und 3). Weder in der 1.Ps.Pl. noch in der 3.Ps.Pl. erscheint hier ein dentalhaltiges Suffix. In Variante 4 hingegen ist sowohl für die 1.Ps.Pl. als auch für die 3.Ps.Pl. das dentalhaltige Suffix anzusetzen.

Ausgehend vom mhd. Referenzsystem und in Anlehnung an die Arbeiten von Rabanus (2005; 2008) werden die verschiedenen Suffixtypen mit Großbuchstaben identifiziert (vgl. Tabelle 2 und 3). Dentallosen Suffixen für die 1./3.Ps.Pl. wird der Buchstabe A zugewiesen, dentalhaltigen Suffixen der Buchstabe C. Die Suffixe für die 2.Ps.Pl. erhalten den Buchstaben B. Die Pluralparadigmen-Variante 1 entspricht demnach zum Beispiel der Notation ABC, die Variante 4 weist CBC auf (vgl. Tabelle 3). Variante 4 ist diejenige, die Scheuringer (1993: 78) zufolge

im „salzburgischen Einflußbereich“ in den westmittelbairischen und südmittelbairischen Basisdialekten dominiert.

Auch für die Verwendung der Pluralparadigmen ist die vertikale Relation im Dialekt-Standard-Kontinuum zu berücksichtigen. Im Gegensatz zu den meisten Verben im Standard entspricht das Pluralparadigma von *sein* nicht der Variante 3 (ABA) des Zweiformenplurals, sondern der Variante 4 (CBC) mit Dentalsuffix in der 1./3.Ps.Pl. (*wir*/*sie sind*). Die Untersuchungen von Scheuringer (1990; 1993) in Braunau lieferten zudem Evidenz für die Bestimmung der regiolektalen Variante. Demnach dominiert in den „mittelbairischen Verkehrssprachen heute überwiegend […] Gleichklang zwischen 1. und 3. Person mit Ausgleich nach dem {en}-Morphem“ (Scheuringer 1993: 78). Diese Beobachtung würde bedeuten, dass für den Regiolekt *mia san* und *se san* (Variante 3, ABA) angenommen werden muss (vgl. dazu auch Moosmüller 1991: 50 f.; Vergeiner 2019: 156), zumal *h*-Anlaut bei *sein* laut Scheuringer (1993: 78) auch in der 1.Ps.Pl. nur zusammen mit dentalhaltigem Suffix erscheint. Der *h*-Anlaut ist wie oben beschrieben auf den Basisdialekt beschränkt.

Aufgrund der anzunehmenden Dialekt-zu-Standard-Advergenz wird mit Ausgleichserscheinungen zugunsten des Zweiformenplurals der Variante 3 (ABA) in den Basisdialekten Salzburgs gerechnet, wobei sich die südbairischen Dialekte des Lungaus konservativer verhalten dürften und nach wie vor zur Variante 1 (ABC) tendieren könnten (vgl. Bülow/Scheutz/Wallner 2019).

Hypothese 3: In den Basisdialekten Salzburgs findet ein Wandel zum Paradigma des Zweiformenplurals der Variante 3 (ABA) statt.

4 Methode

Die in Abschnitt 3 aufgestellten Hypothesen werden sowohl mithilfe einer *apparent-time*- als auch einer *real-time*-Studie (*cross-sections*) untersucht. Synchron kontrastiert werden dabei Gewährspersonen (GP) zweier Generationen, einer älteren (65 bis 85 Jahre) und einer jüngeren (20 bis 35 Jahre). Diese Daten stammen aus Dialektaufnahmen von 2016/17, die zur Erstellung eines ‚sprechenden Sprachatlas‘ von Salzburg angefertigt wurden (vgl. Scheutz 2017). Des Weiteren fließen hier Daten aus einem Teilprojekt des SFB „Deutsch in Österreich“ (DiÖ) ein, die ebenfalls 2016/17 aufgenommen wurden.[5] Für die *real-time*-Studie werden Daten aus den 1970/80er Jahren mit den oben genannten Aufnahmen für den „Sprachatlas Salzburg“ von 2016/17 verglichen.

4.1 Material und Stimuli

Alle Daten wurden direkt von dafür trainierten Exploratoren mithilfe traditioneller Dialektfragebücher erhoben. Die Dialektfragebücher weisen große inhaltliche Überschneidungsbereiche auf; sie sind an die Dialektfragebücher des SNiB oder SAO angelehnt. Die Erhebungen aus den 1970/80er Jahren wurden von Herbert Tatzreiter, Werner Bauer, Franz Patocka und Hermann Scheuringer durchgeführt. Die Transkripte wurden noch während der Befragung in Teuthonista vor Ort angefertigt, dabei sind mehrere GP für ein Fragebuch befragt worden.

Hannes Scheutz und Dominik Wallner haben die Audioaufnahmen von 2016/17 (Sprachatlas Salzburg und DiÖ-Projekt) durchgeführt. Diese dauerten jeweils zwischen zwei und drei Stunden. Jede GP hatte alle Fragen des Fragebuchs zu beantworten. Die Mehrzahl der Items

[5] Es handelt sich um das hauptsächlich an der Universität Salzburg angesiedelte Teilprojekt „Variation und Wandel dialektaler Varietäten in Österreich (in *real* und *apparent time*) (F 6002-G23)“ des vom FWF geförderten Spezialforschungsbereichs „Deutsch in Österreich (DiÖ)“ (SFB F 60) (vgl. Budin u. a. 2019).

besteht aus Übersetzungs- und Ergänzungsaufgaben. Das Verb *sein* wurde in allen Fragebüchern sowohl im Paradigma als auch in Satzkontexten abgefragt. In Tabelle 4 sind exemplarisch die abgefragten Satzkontexte aus dem DiÖ-Dialektfragebuch wiedergegeben. Dabei wird deutlich, dass *sein* insbesondere in der 1./3.Ps.Pl. abgefragt wurde, nicht aber in der 2.Ps.Pl.

Tabelle 4: Sätze aus dem DiÖ-Fragebuch, in denen Pluralformen von *sein* abgefragt wurden

Nr.	**Item**	**Person**
43	Das waren die Allerklügsten!	3.Ps.Pl.
67	Wo sind Mutters Schuhe?	3.Ps.Pl.
82	Die Schlitten im Katalog sind aber ziemlich teuer.	3.Ps.Pl.
87	Das waren aber schöne Tage.	3.Ps.Pl.
135	Die Äpfel waren schon faulig.	3.Ps.Pl.
337	Wir sind heute nach Wien gefahren.	1.Ps.Pl.
338	Wir sind im Hotel.	1.Ps.Pl.
368	Die sind am besten!	3.Ps.Pl.
377	Diese Würste sind gesotten besser als gebraten.	3.Ps.Pl.
533	Die Vögel, die dort sitzen, sind aber dick.	3.Ps.Pl.
543	Jetzt höre ich zu fragen auf, weil wir fertig sind.	1.Ps.Pl.

4.2 Gewährspersonen und Aufnahmeorte

Sowohl für die *real-time*- als auch für die *apparent-time*-Studie kann auf Daten aus einem relativ engmaschigen Ortsnetz zurückgegriffen werden (vgl. Abbildung 4), wodurch sich Veränderungen in den jeweiligen Dialektregionen gut beobachten lassen. Das Netz für die 1970/80er Jahre besteht aus 57 Orten, jeweils 9 Orte liegen im westmittelbairischen und südbairischen Gebiet, 39 Orte in der südmittelbairischen Übergangszone (vgl. oben Abbildung 1). Für den Sprachatlas Salzburg wurden 63 GP aus 32 Orten aufgenommen. Hier liegen 10 Orte im Westmittelbairischen (davon 4 auf bayerischer Seite des Grenzgebiets), 4 Orte im südbairischen Lungau und 18 Orte in der südmittelbairischen Übergangszone. Im Kontext des DiÖ-Projekts sind 20 GP aus 5 Orten aufgenommen worden. Diese verteilen sich über die Regionen Flachgau (Berndorf), Tennengau (Rußbach), Pongau (Hüttschlag), Pinzgau (Maria Alm) und Lungau (Lessach).

Abbildung 4: Aufnahmeorte 1970/80er Jahre (links), Sprachatlas Salzburg 2016/17 (Mitte), DiÖ-Projekt (rechts); für die Auflistung der Ortspunkte siehe Anhang (Tabelle 5 bis 7)

Alle GP wurden anhand soziodemografischer Merkmale ausgewählt: In den 1970/80er Jahren wurden ausschließlich typische NORMs und NORFs (**N**on-mobile **O**ld **R**ural **M**ales / **F**emales, > 65 Jahre) befragt, also ältere ‚ortsfeste' Gewährsmänner und -frauen. Neben diesen NORMs und NORFs wurden 2016/17 je Ortspunkt auch eine ortsansässige, jüngere GP (< 35 Jahre) ohne tertiären Bildungsabschluss aufgenommen. Zudem wurde auf ein ausgewogenes Geschlechterverhältnis geachtet. Für die DiÖ-Aufnahmen sind jeweils je eine NORF, ein NORM sowie ein jüngerer Sprecher und eine jüngere Sprecherin (beide ebenfalls ortsfest) befragt worden.

5 Ergebnisse

Im Folgenden wird zunächst die Variation und der Wandel der Pluralformen von *sein* variablenweise dargestellt, bevor schließlich die Einzelergebnisse im Diskussionsteil (vgl. Abschnitt 6) zusammenfassend betrachtet werden.

5.1 *h*/*s*-Anlautalternation

Die Daten aus den 1970/80er Jahren zeigen für die *h*/*s*-Anlautvariation eine klare diatopische Verteilung (vgl. Abbildung 5), $\chi^2 = 22.326$, $n = 57$, $df = 2$, $p < .001$, Cramer's $V = .626$.

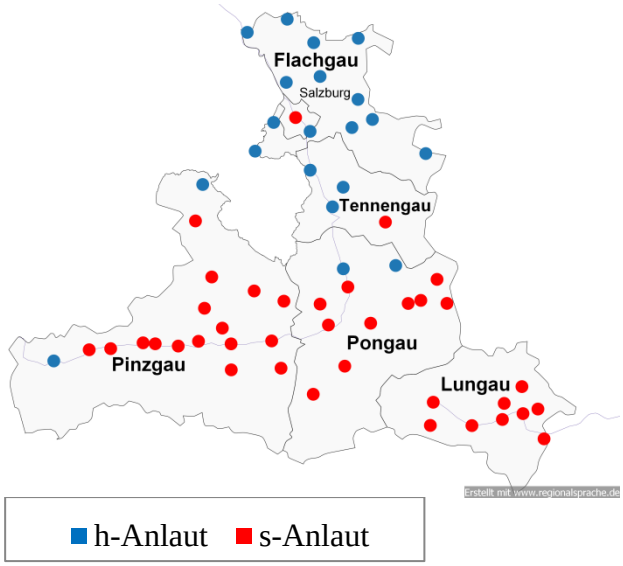


Abbildung 5: Verteilung der Anlautvariation in den 1970/80er Jahren

In den westmittelbairischen Dialekten dominiert klar der *h*-Anlaut (mit Ausnahme der Stadt Salzburg), während im südbairischen Raum ausnahmslos *s*-Anlaut auftritt. Für das südmittelbairische Übergangsgebiet zeigt sich eine eindeutige Nord-Süd-Verteilung: Im Tennengau, das an die westmittelbairischen Dialekte angrenzt, erscheint *h*-Anlaut; im südlicheren Pinzgau und im Pongau überwiegt deutlich der *s*-Anlaut.

Vergleicht man die Ergebnisse der NORM/Fs aus den 1970/80er Jahren mit denen der NORM/Fs, die 2016/17 für den Sprachatlas Salzburg befragt wurden, zeigt sich kein signifikanter Unterschied, $\chi^2 = .859$, $n = 88$, $df = 1$, $p = .354$, Cramer's $V = .099$. Die signifikante Verteilung der Anlautvariation über die drei Dialektregionen bleibt dabei bestehen, $\chi^2 = 14.849$, $n = 31$, $df = 2$, $p = .001$, Cramer's $V = .692$. Das Raumbild ähnelt dem für die 1970/80er Jahre.

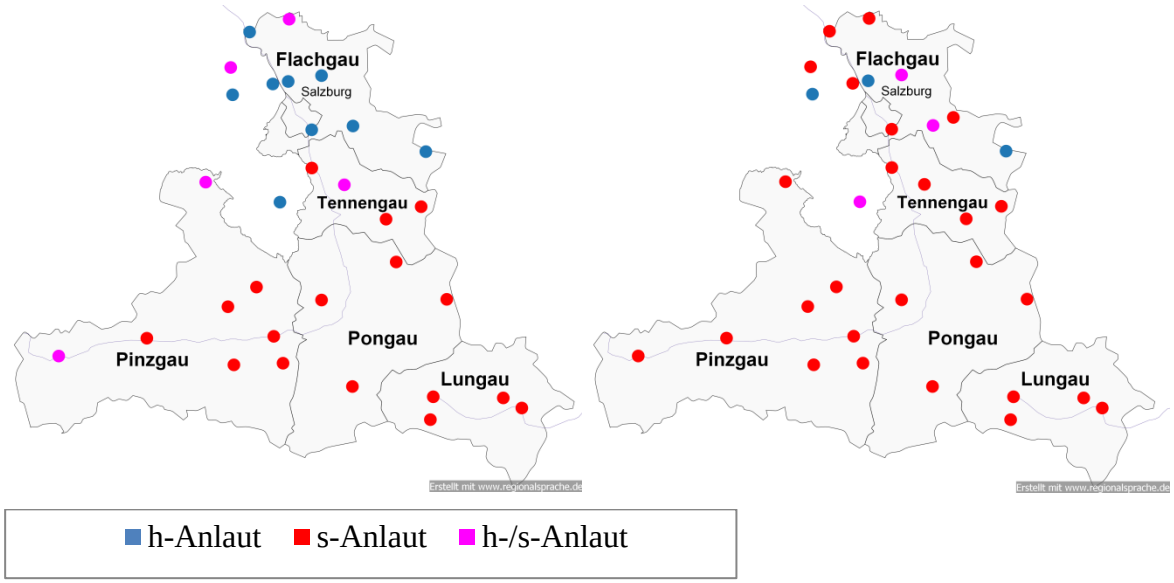


Abbildung 6: Verteilung der Anlautvariation 2016/17 (Sprachatlas Salzburg) für ältere (links) und jüngere GP (rechts)

Für die Daten des Sprachatlas Salzburg von 2016/17 ergibt sich allerdings ein *apparent-time*-Effekt (vgl. Abbildung 6), $\chi^2 = 3.946$, $n = 63$, $df = 1$, $p = .047$, Cramer's $V = .25$. Die jüngeren GP verwenden tendenziell eher den *s*-Anlaut. Für diese Gruppe zeigt sich nun auch nicht mehr die vormals signifikante diatopische Verteilung über die drei größeren Dialektregionen, $\chi^2 = 5.744$, $n = 32$, $df = 2$, $p = .057$, Cramer's $V = .47$. Für diese Daten wurde zudem die intra-individuelle Variation im Paradigma untersucht. Von 63 GP variieren immerhin 8 (5 ältere, 3 jüngere; entsprechen in Abbildung 6 den lila Punkten); diese verwenden sowohl den *h*- als auch den *s*-Anlaut im Paradigma (z. B. *san – sadds – hand*).

In den DiÖ-Aufnahmen tritt der *h*-Anlaut sowohl im Paradigma als auch im Satzkontext nur im westmittelbairischen Berndorf auf, hier insbesondere bei den beiden älteren GP (in 96% der Fälle). Die jungen Berndorfer verwenden den *h*-Anlaut hingegen kaum noch (nur in 13% der Fälle). In den süd(mittel)bairischen Orten erscheint ausschließlich *s*-Anlaut.

5.2 *a*/*ai*-Stammvokalvariation

Die Daten aus den 1970/80er Jahren zeigen ebenfalls eine signifikante diatopische Verteilung der *a*/*ai*-Stammvokalvariation (vgl. Abbildung 7), $\chi^2 = 12.555$, $n = 57$, $df = 2$, $p = .002$, Cramer's $V = .47$.

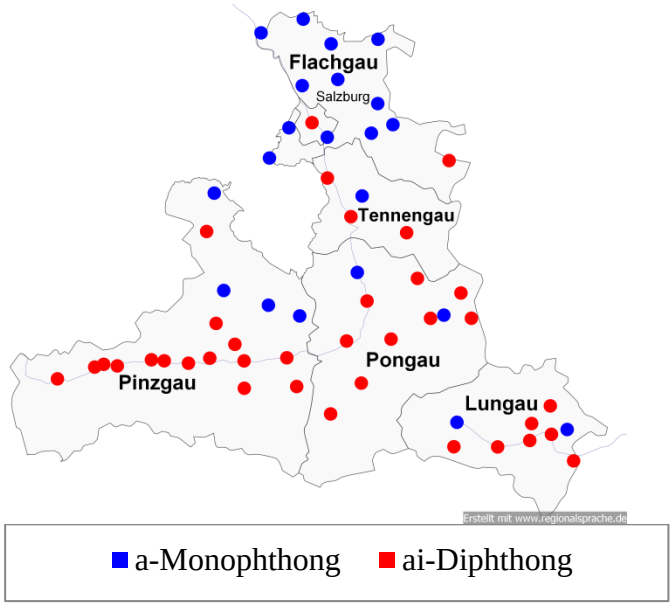


Abbildung 7: Verteilung der Stammvokalvariation in den 1970/80er Jahren

Im westmittelbairischen Flachgau dominiert der Monophthong (in der Stadt Salzburg wird aber diphthongiert), im südbairischen Lungau überwiegt deutlich der Diphthong (mit Ausnahme der zwei Orte Zederhaus und Sauerfeld, die entsprechend eine blaue Markierung haben). In der südmittelbairischen Übergangszone hat der Diphthong (78% der Fälle) ebenfalls ein Übergewicht, wobei im Nord-Westen des Pinzgaus in 4 Orten der Monophthong erscheint. Im Tennengau wird ebenfalls der Diphthong bevorzugt.

Im Vergleich der NORM/Fs (1970/80er vs. 2016/17er Sprachatlas) zeigt sich zwar keine statistisch signifikante Korrelation, $\chi^2 = .185$, $n = 88$, $df = 1$, $p = .667$, Cramer's $V = .046$; der Diphthong scheint sich allerdings in Richtung Norden in die westmittelbairischen Dialekte auszubreiten. Im Flachgau macht der Monophthong bei den NORM/Fs, die 2016/17 für den Sprachatlas Salzburg befragt wurden, nur noch 62,5% der Belege aus, während im Lungau nun in 100% der Fälle der Diphthong erscheint. Die diatopische Verteilung der Stammvokalvariation ist für diese GP knapp nicht mehr signifikant, $\chi^2 = 5.896$, $n = 31$, $df = 2$, $p = .052$, Cramer's $V = .44$).

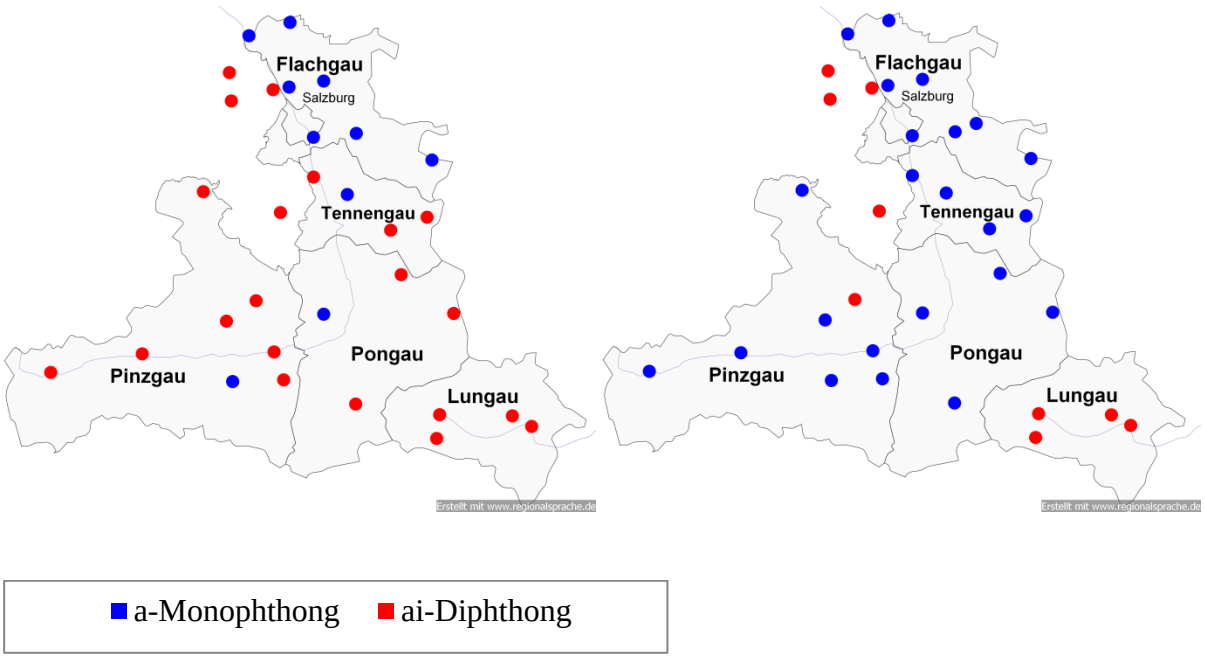


Abbildung 8: Verteilung der Stammvokalvariation 2016/17 (Sprachatlas Salzburg) für ältere (links) und jüngere GP (rechts)

Auch für die Stammvokalvariation ergibt sich ein interessanter *apparent-time*-Effekt mit Blick auf die Daten für den Sprachatlas 2016/17, $\chi^2 = 9.908$, $n = 63$, $df = 1$, $p = .002$, Cramer's $V = .397$. Während sich der Diphthong im *real-time*-Vergleich der älteren GP noch auszubreiten scheint, favorisiert die jüngere Generation 2016/17 den Monophthong. Dieser wird von den jüngeren GP nun auch im Südosten der südmittelbairischen Übergangszone verwendet. Der Diphthong behält seine Dominanz nur noch im südbairischen Lungau, wo er in 83% der Fälle erscheint. Die diatopische Verteilung des Merkmals über die Dialektregionen ist innerhalb der jüngeren Generation signifikant, $\chi^2 = 9.354$, $n = 32$, $df = 2$, $p = .009$, Cramer's $V = .541$.

Für die DiÖ-Daten zeigt sich eine klare Verteilung für die Ortspunkte: In Berndorf (Flachgau) und Maria Alm (Pinzgau) erscheint ausschließlich der Monophthong, während die GP in Rußbach (Tennengau), Hüttschlag (Pongau) und Lessach (Lungau) alle den Diphthong bevorzugen.

5.3 Morphologische Variation im Pluralparadigma

Was für die beiden phonetisch-phonologischen Variablen gilt, zeigt sich auch für die Verwendung der Varianten im Pluralparadigma. Die Daten aus den 1970/80er Jahren lassen ein Raumbild erkennen (vgl. Abbildung 9).

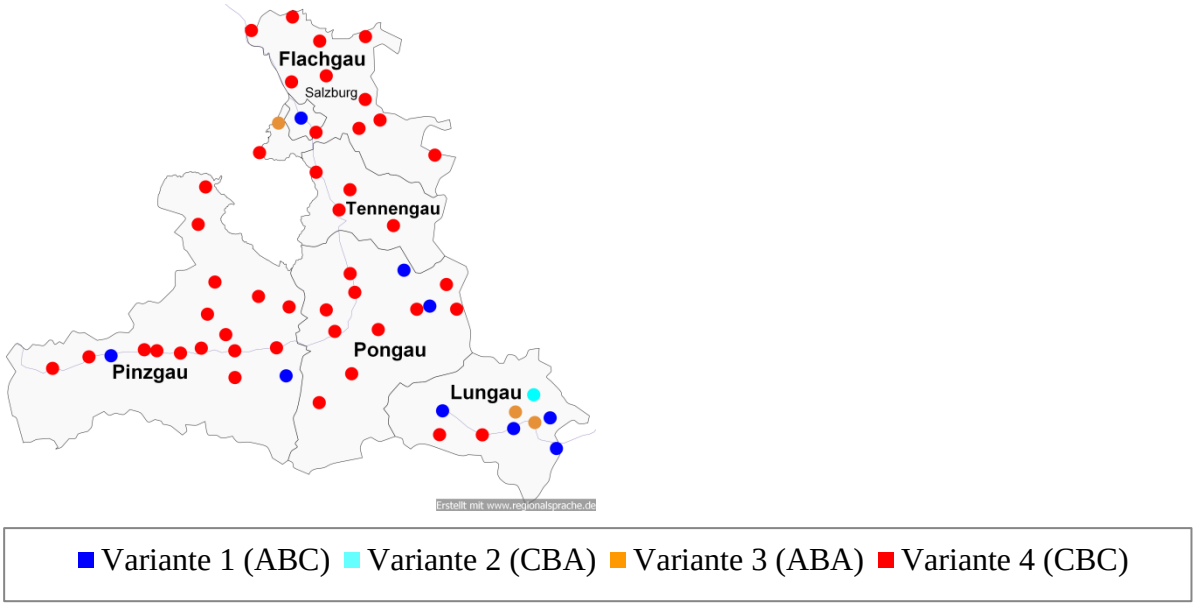


Abbildung 9: Verteilung der Pluralparadigmenvarianten von *sein* in den 1970/80er Jahren

Die Verwendung der Varianten korreliert mit den drei Dialektregionen, $\chi^2 = 17.413$, $n = 57$, $df = 6$, $p = .008$, Cramer's $V = .39$. In den westmittelbairischen Dialekten wird ausschließlich der Zweiformenplural der Variante 4 (CBC) verwendet, was mit den Beobachtungen von Scheuringer (1993: 78) übereinstimmt. Diese Variante (82% der Fälle) überwiegt auch deutlich in der südmittelbairischen Übergangszone. Im südbairischen Lungau ist die Verteilung weniger deutlich. Neben der Variante 4 (CBC) wird gleichermaßen die Variante 1 (ABC) des Dreiformenplurals verwendet.

Während sich für die beiden phonetisch-phonologischen Variablen kein Unterschied zwischen den NORM/Fs der 1970/80er Jahre und denen von 2016/17 gezeigt hat, kann für die Verwendung der Pluralparadigmen eine signifikante Korrelation konstatiert werden, $\chi^2 = 16.847$, $n = 85$, $df = 3$, $p = .001$, Cramer's $V = .45$. Die NORM/Fs der Sprachatlas-Erhebung von 2016/17 verwenden nun häufiger den Zweiformenplural der Variante 3 (ABA) (vgl. Abbildung 10). Varianten 4 und Variante 1 werden dafür weniger häufig genutzt. Mit dieser Veränderung geht auch einher, dass sich für die Sprachatlasdaten von 2016/17 keine Korrelation der Verwendung der Paradigmenvarianten zu den Dialektregionen ergibt, $\chi^2 = 9.245$, $n = 28$, $df = 6$, $p = .16$, Cramer's $V = .406$.

Der Trend, dass sich der Zweiformenplural der Variante 3 (ABA) ausbreitet, setzt sich auch für die jüngere Generation fort. Diese nutzen Variante 3 sogar in 71% der Fälle. Die Verwendung des Zweiformenplurals der Variante 4 (CBC) beschränkt sich auf die südmittelbairische Übergangszone, insbesondere auf den Pinzgau. Der Dreiformenplural tritt bei den jüngeren GP nur einmal in Form der Variante 2 (CBA) auf (vgl. Abbildung 10).

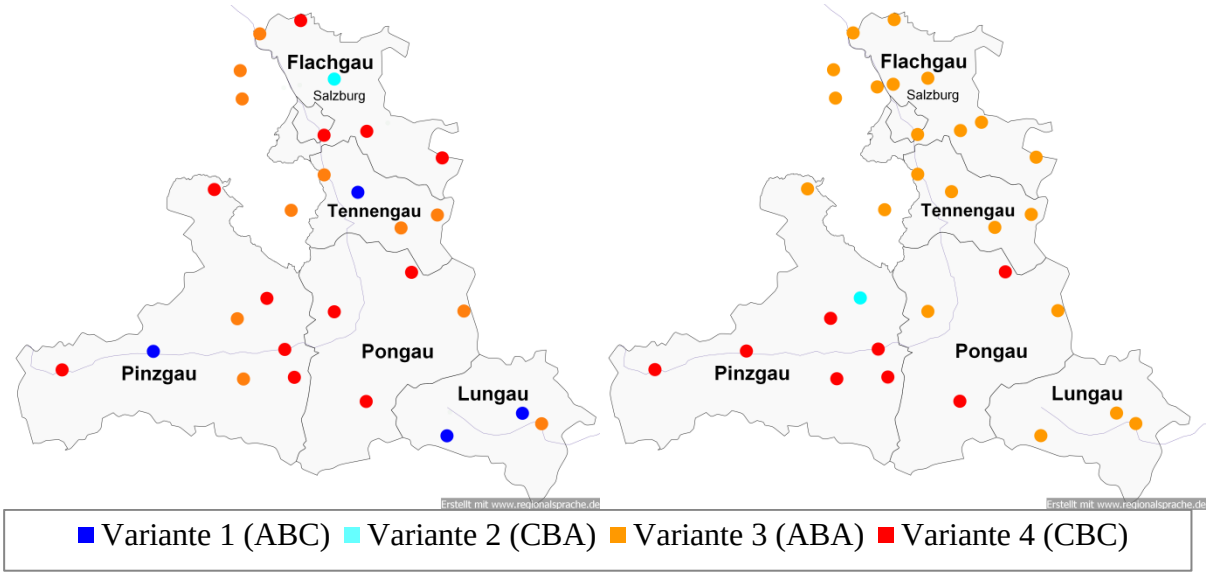


Abbildung 10: Verteilung der Pluralparadigmenvarianten 2016/17 (Sprachatlas Salzburg) für ältere (links) und jüngere GP (rechts)

Damit zeigt sich für diese Variable zwar ein signifikanter *apparent-time*-Effekt, $\chi^2 = 8.336$, $n = 59$, $df = 3$, $p = .040$, Cramer's $V = .38$, allerdings besteht keine Korrelation zwischen der Verwendung der Pluralparadigmenvarianten und den Dialektregionen, $\chi^2 = 5.920$, $n = 31$, $df = 4$, $p = .201$, Cramer's $V = .31$.

Die DiÖ-Daten weisen mit Blick auf die Verwendung der Suffixe für die 1./3.Ps.Pl. viel intra-individuelle Variation auf. Diese Variation ist insbesondere in den Satzkontexten virulent, hier verwenden 14 der 20 GP für die 1.Ps.Pl. sowohl das Suffix *-(e)n* als auch das Suffix *-(e)nd*. Für die 3.Ps.Pl. verwenden sogar 19 von 20 GP beide Suffixe. Am meisten variieren die GP aus dem südmittelbairischen Dialektraum (Maria Alm, Hüttschlag und Rußbach) während sich die GP aus dem Westmittelbairischen (Berndorf) und Südbairischen (Lessach) zumindest im Hinblick auf die 1.Ps.Pl. tendenziell stabiler verhalten.

6 Diskussion

Sowohl die Ergebnisse der *real-time*- als auch die der *apparent-time*-Analyse verdeutlichen für die Pluralformen des Verbs *sein* Dialekt-zu-Standard-Advergenz. Damit bestätigt sich der in den Hypothesen antizipierte Trend hinsichtlich der Hauptentwicklungsrichtung für den rezenten Dialektwandel in den bairischen Dialektregionen Österreichs. Die Daten zeigen allerdings auch, dass der Wandel aufgrund der unterschiedlichen Ausgangsbedingungen in den verschiedenen Dialektregionen differenziert bewertet werden muss.

Mit der Hypothese 1 wurde für die westmittelbairischen Dialekte ein Wandel vom *h*- zum *s*-Anlaut erwartet. Diese Entwicklung bestätigt sich insbesondere durch den *apparent-time*-Effekt, der sich für die Daten des Sprachatlas Salzburg und die DiÖ-Aufnahmen von 2016/17 nachweisen lässt. Die jüngeren Gewährspersonen (GP) zeigen nun in der gesamten Dialektregion Salzburg *s*-Anlaut. Der *h*-Anlaut ist nur noch bei den älteren GP im Westmittelbairischen (im Norden des Bundeslandes Salzburg) präsent.

Während die südmittelbairischen Dialekte (in der Mitte und im Westen Salzburgs) sowie die südbairischen Dialekte (in dem im Südosten Salzburgs gelegenen Lungau) beim Anlaut stabil sind, zeigt sich in diesen Gebieten allerdings ein zum Regiolekt gerichteter Wandel des Stammvokals der 2.Ps.Pl. Auch für die *a*/*ai*-Variation im Stammvokal besteht ein *apparent-time*-Effekt, welcher die Hypothese 2 bestätigt. Die jüngeren GP tendieren deutlich zur

Verwendung des Monophthongs und damit zur Form *sadds*, die auch umgangssprachlich markiert ist. Der Befund wird weiters dadurch untermauert, dass der *a*-Monophthong für die westmittelbairischen Dialekte ein stabiles Merkmal bleibt. So erscheint beispielsweise in den DiÖ-Aufnahmen von Berndorf (in dem im Norden des Bundeslandes Salzburg gelegenen Flachgau) ausschließlich der Monophthong.

Für die Verwendung des Pluralparadigmas kann sowohl ein *real-time*- als auch ein *apparent-time*-Effekt nachgewiesen werden. Das heißt konkret für die *real-time*-Analyse, dass die älteren, ortsfesten GP (NORM/Fs) aus den 1970/80er Jahren die Variante 1 (ABC; vgl. Tabelle 3) des Dreiformenplurals und die Variante 4 (CBC) des Zweiformenplurals häufiger verwenden als ihre Pendants aus den Jahren 2016/17. Die NORM/Fs von 2016/17 gebrauchen insbesondere im Westmittelbairischen deutlich öfter die Variante 3 (ABA) des Zweiformenplurals. Dieser Trend setzt sich für die jüngeren GP fort; diese tendieren in allen Dialektregionen klar zur Variante 3. Damit bestätigt sich auch die Hypothese 3. Dass der antizipierte Wandel aber noch nicht abgeschlossen ist, zeigt sich u. a. daran, dass die Variante 4 im Westen des Pinzgaus noch einigermaßen präsent ist und ebenfalls Variante 1 noch vereinzelt im südbairischen Lungau erscheinen kann. Auch Variante 2 (CBA) ist belegt, im Salzburger Sprachatlas sowohl bei einer älteren GP aus Seekirchen (Flachgau) als auch einer jüngeren GP aus Maria Alm (Pinzgau, vgl. in Abbildung 10 die hellblauen Punkte auf den Karten links und rechts). Im DiÖ-Datensatz ist Variante 2 bei einer jüngeren GP aus Hüttschlag verwendet worden. Der Wandel führt dazu, dass sich die signifikante diatopische Gliederung, die für die 1970/80er Jahre nachgewiesen werden konnte, für die Daten von 2016/17 nicht mehr bestätigt.

Mit der Entwicklung hin zu den regiolektaleren Formen geht für alle Variablen einher, dass die diatopischen Unterschiede zwischen den Dialektregionen und damit auch die Differenzierungsmöglichkeiten zwischen zum Beispiel den west- und ostmittelbairischen oder auch zwischen den süd- und südmittelbairischen Dialekten weiter nivelliert werden. Diatopisch betrachtet heißt das letztlich auch, dass sich Homologie mit den progressiveren ostbairischen Dialektformen von *sein* ergibt.

Dieser Entwicklungsprozess verläuft allerdings nicht abrupt und ist auch noch nicht vollständig abgeschlossen, wie insbesondere die Variationsbreite in den DiÖ-Daten im Hinblick auf die Verwendung der Suffixe für die 1./3.Ps.Pl. nahelegt. Bülow/Scheutz/Wallner (2019) zeigen diese Variationsbreite auch für andere Verben im Salzburger Raum. Die Pluralbildung kann je nach Verb und GP im Rahmen der ober skizzierten strukturellen Möglichkeiten sehr unterschiedlich ausfallen. Die klare Dominanz von Variante 4 (CBC) für *sein* im Westbairischen und im südmittelbairischen Übergangsgebiet in den Daten der 1970/80er Jahre zeigt sich beispielsweise für andere Verben weniger deutlich.

Die Daten zeigen insgesamt sowohl für die phonetisch-phonologische als auch für die morphologische Ebene verschiedene Arten von Variation (z. B. diatopische Variation, intra-individuelle Variation) und Wandel (*apparent*- und *real-time*-Effekte), die systemlinguistisch nur schwer gefasst werden können. Der Begriff Allomorphie beispielsweise ist für Morph-Alternanz innerhalb eines Systems reserviert. Der von Haas (2004: 10–13) und Koch (2006) eingeführte Terminus „Heteromorphie“[6] erfasst zwar Variation zwischen verschiedenen Systemen, kann aber nicht die Frage lösen, wo bzw. wann das eine System anfängt und das andere System aufhört.

Die vorliegende Analyse ermöglicht lediglich allgemeine Aussagen über die Entwicklung der Pluralformen von *sein* in den größeren Dialektregionen Salzburgs. Um die konkrete sprachliche

[6] Koch (2006: 130) definiert ein Heteromorph „als ‚ein Morph, das innerhalb eines Diasystems gegenüber einem grammatisch entsprechenden Morph eine andere Ausdrucksseite aufweist‘. Daraus lässt sich ableiten, dass sich zwei Morphe unterschiedlicher Systeme ‚heteromorph‘ zueinander verhalten“.

Orientierung (horizontal wie vertikal) der Sprecher/innen besser einschätzen und insbesondere das Ausmaß und den Einfluss des Kontakts mit ‚höherschichtigen' Varietäten untersuchen zu können, müssen in zukünftigen Untersuchungen weitere ‚äußere' Einflussfaktoren wie etwa Verkehrswege, die Bedeutung des Tourismus oder andere wirtschaftliche Strukturen am Ort einbezogen werden.

7 Fazit

Für die hier untersuchten Variablen der Pluralformen von *sein* wurde anhand von *real*- und *apparent-time*-Effekten Advergenz zu überregionalen Varietäten (regiolektal, standardnah) festgestellt. Die Varianten, die sich bei den jüngeren GP im Westmittelbairischen, Südmittelbairischen und Südbairischen durchsetzen (*s*-Anlaut, *a*-Stammvokal und Zweiformenplural der Variante 3 (ABA)), entsprechen gleichzeitig den progressiveren und prestigereichen Varianten der ostmittelbairischen Dialekte. Das wiederum kann makrostrukturell als Evidenz dafür gewertet werden, dass sich dialektale Neuerungen in den bairischen Dialekten Österreichs tendenziell von Osten nach Westen ausbreiten (vgl. Moosmüller 1991; Wiesinger 2004; Moosmüller/Scheutz 2013).

Literatur

Auer, Peter. 2005. Europe's Sociolinguistic Unity, or: A Typology of European Dialect / Standard Constellations. In: Delbecque, Nicole/ van der Auwera, Johan/ Geeraerts, Dirk (Hrsg.), *Perspectives on Variation*, 7-42. Berlin: de Gruyter.

Auer, Peter. 2018. Dialect Change in Europe – Leveling and Convergence. In: Boberg, Charles/ Nerbonne, John/ Watt, Dominic (Hrsg.), *The Handbook of Dialectology*, 159-176. Oxford: Wiley.

Budin, Gerhard/ Elspaß, Stephan/ Lenz, Alexandra N./ Newerkla, Stefan M./ Ziegler, Arne. 2019. The Research Project (SFB) 'German in Austria'. Variation – Contact – Perception. In: Bülow, Lars/ Fischer, Ann-Kathrin/ Herbert, Kristina (Hrsg.), *Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptualisierung*, 7-35. Berlin u. a.: Lang.

Bülow, Lars/ Scheutz, Hannes/ Wallner, Dominik. 2019. Variation and change of plural verbs in Salzburg's base dialects. In: Schallert, Oliver/ Dammel, Antje (Hrsg.), *Morphological Variation. Theoretical and empirical perspectives*, 95–134. Amsterdam: Benjamins.

Bülow, Lars/ Vergeiner, Philip C./ Scheutz, Hannes/ Wallner, Dominik/ Elspaß, Stephan. 2019. Wie regelhaft ist Lautwandel? Grundsätzliche Überlegungen anhand einer Fallstudie zu mhd. ô in den bairischen Dialekten Österreichs und Südtirols. *Zeitschrift für Dialektologie und Linguistik* 86 (1), 5–50.

Bülow, Lars/ Wallner, Dominik. In Druck. Inter- and intra-individual variation in Bavarian base dialects. The case of the verb *sein*. In: Lenz, Alexandra N./ Maselko, Mateusz (Hrsg.), *Variationist Linguistics meets Contact Linguistics*. Wiener Arbeiten zur Linguistik 2. Göttingen: Vienna University Press.

Christen, Helen/ Glaser, Elvira/ Friedli, Matthias. 2013. *Kleiner Sprachatlas der deutschen Schweiz*. Frauenfeld u. a.: Huber.

Haas, Walter. 2004. Sprachatlanten als Darstellungsmittel der Dialektgeographie. Aus Anlaß des Erscheinens des Sprachatlasses von Bayerisch-Schwaben. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 126. 1-22.

Koch, Günter. 2004. Irregularisierung und Nivellierung der Paradigmen von *haben* und *sein*. In: Kanz, Ulrich/ Wildfeuer, Alfred (Hrsg.): *Kreuther Kräuterbuschen. Beiträge zur 9. Bayerisch-österreichischen Dialektologentagung in Wildbad Kreuth*, 121-139. Regensburg: vulpes.

Koch, Günter. 2006. Diatopische Heteromorphie: Die Endung des Partizips Präteritum der schwachen Verben im alemannischen Sprachraum. In: Klausmann, Hubert (Hrsg.), *Raumstrukturen im Alemannischen*, 129-140. Graz: Neugebauer.

Koch, Günter. 2007. Formengeographie I: Verbum. In: Eroms, Hans-Werner (Hrsg.). 2003 ff. *Sprachatlas von Niederbayern*. (Bayerischer Sprachatlas: Regionalteil 5). Heidelberg: Winter.

Lessiak, Primus. 1903. Die Mundart von Pernegg in Kärnten. In: *Beiträge zur Geschichte der deutschen Sprache und Literatur* 28. 1-227.

Mauser, Peter. 1998. *Die Morphologie im Dialekt des Salzburger Lungaus*. Wien: Lang.

Mauser, Peter. 2007. Dialektale Wortformen und Formsysteme. In: Scheutz, Hannes (Hrsg.), *Drent und herent. Dialekte im salzburgisch-bayerischen Grenzgebiet*, 57-78. Salzburg u. a.: EuRegio.

Moosmüller, Sylvia. 1991. *Hochsprache und Dialekt in Österreich. Soziophonologische Untersuchungen zu ihrer Abgrenzung in Wien, Graz, Salzburg und Innsbruck*. Wien: Böhlau.

Moosmüller, Sylvia/ Scheutz, Hannes. 2013. Chain shifts revisited: The case of monophthongisation and *E*-merger in the city dialects of Salzburg and Vienna. In: Auer, Peter/ Caro Reina, Javier/ Kaufmann, Göz (Hrsg.), *Language variation – European Perspectives* IV, 173-186. Amsterdam u. a.: Benjamins.

Nübling, Damaris. 2000. *Prinzipien der Irregularisierung. Eine kontrastive Analyse von zehn Verben in zehn germanischen Sprachen*. Tübingen: Niemeyer.

Paul, Hermann. 2007. *Mittelhochdeutsche Grammatik*. 25. Aufl., neu bearbeitet von Thomas Klein, Hans-Joachim Solms und Klaus-Peter Wegera. Mit einer Syntax von Ingeborg Schöbler, neubearbeitet und erweitert von Heinz-Peter Prell. Tübingen: Niemeyer.

Philipp, Marthe/ Weider, Erich. 2002. sein *und* haben *im elsass-lothringischen Mundartraum. Ein organisiertes Chaos*. Stuttgart: Steiner.

Rabanus, Stefan. 2005. Dialektwandel im 20. Jahrhundert: Verbalplural in Südwestdeutschland. In: Eggers, Eckhard/ Schmidt, Jürgen Erich/ Stellmacher, Dieter (Hrsg.): *Moderne Dialekte – Neue Dialektologie. Akten des 1. Kongresses der internationalen Gesellschaft für Dialektologie des Deutschen (IGDD)*, 267-290. Stuttgart: Steiner.

Rabanus, Stefan. 2008. *Morphologisches Minimum. Distinktionen und Synkretismen im Minimalsatz hochdeutscher Dialekte*. Stuttgart: Steiner.

SAO = Gaisbauer, Stephan/ Scheuringer, Hermann (Hrsg.). 2003 ff. *Sprachatlas von Oberösterreich*. Linz: Adalbert-Stifter-Institut des Landes Oberösterreich.

Schatz, Josef. 1897. *Die Mundart von Imst. Laut- und Flexionslehre*. Straßburg: Trübner.

Scheuringer, Hermann. 1990. *Sprachentwicklung in Bayern und Österreich. Eine Analyse des Substandardverhaltens der Städte Braunau am Inn (Österreich) und Simbach am Inn (Bayern) und ihres Umlandes*. Hamburg: Buske.

Scheuringer, Hermann. 1993. Ziele und Methoden dialektologischer Stadtsprachenforschung (dargestellt am Beispiel des Vergleichs von Braunau am Inn, Österreich, und Simbach am Inn, Deutschland). *Dialectologia et Geolinguistica* 1, 70-91.

Scheutz, Hannes. 2017. *Sprachatlas Salzburg*. <https://www.sprachatlas.at/salzburg/> [zuletzt aufgerufen am 25. April 2018].

SNiB = Eroms, Hans-Werner (Hrsg.). 2003 ff. *Sprachatlas von Niederbayern*. Heidelberg: Winter. (Bayerischer Sprachatlas. Regionalteil V).

SOB = Eichinger, Ludwig M. (Hrsg.). 2008 ff. *Sprachatlas von Oberbayern*. Heidelberg: Winter. (Bayerischer Sprachatlas. Regionalteil VI).

Vergeiner, Philip C. 2019. *Kookkurrenz – Kovariation – Kontrast. Formen und Funktionen individueller Dialekt-/Standardvariation in universitären Beratungsgesprächen*. Berlin u. a.: Lang.

Wiesinger, Peter. 1983. Die Einteilung der deutschen Dialekte. In: Besch, Werner/ Knoop, Ulrich/ Putschke, Wolfgang/ Wiegand, Herbert E. (Hrsg.): *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Zweiter Halbband, 807-900. Berlin, New York: de Gruyter.

Wiesinger, Peter. 1989. *Die Flexionsmorphologie des Verbums im Bairischen*. Wien: Österreichische Akademie der Wissenschaften.

Wiesinger, Peter. 2004. Die Dialektgeographie Oberösterreichs und ihre Geschichte. In: Gaisbauer, Stephan/ Scheuringer, Hermann (Hrsg.): *Linzerschnitten*, 15-61. Linz: Adalbert-Stifter-Institut des Landes Oberösterreich.

Wurzel, Wolfgang U. 1984. *Flexionsmorphologie und Natürlichkeit. Ein Beitrag zur morphologischen Theoriebildung*. Berlin: Akademie Verlag.

Anhang

Tabelle 5: Aufnahmeorte 1970/80er Jahre

Nr.	Ort	Nr.	Ort
1	Wald im Pinzgau	30	Forstau
2	Bramberg am Wildkogel	31	Zederhaus
3	Hollersbach	32	Muhr
4	Stuhlfelden	33	St. Michael im Lungau
5	Uttendorf	34	Maria Pfarr
6	Niedernsill	35	Unternberg
7	Piesendorf	36	Lessach
8	Unken	37	Tamsweg
9	St. Martin bei Lofer	38	Sauerfeld
10	Leogang	39	Kendlbruck
11	Viehhofen	40	St. Georgen
12	Zell am See	41	Dorfbeuern
13	Bruck	42	Mattsee
14	Fusch	43	Strasswalchen
15	Maria Alm am Steinernen Meer	44	Anthering
16	Dienten am Hochkönig	45	Seekirchen am Wallersee
17	Taxenbach	46	Thalgau
18	Rauris	47	Wals-Siezheim
19	Bad Hofgastein	48	Salzburg
20	Mühlbach	49	Elsbethen
21	St. Veit am Pongau	50	Faistenau
22	Großarl	51	Fuschl
23	Pfarrwerfen	52	Grossgmain
24	Bischofshofen	53	Hallein
25	Wagrain	54	Strobl
26	St. Martin am Tennengebirge	55	St. Kolomann
27	Altenmarkt	56	Golling
28	Radstadt	57	Abtenau
29	Filzmoos		

Tabelle 6: Aufnahmeorte Sprachatlas Salzburg 2016/17

Nr.	Ort	Nr.	Ort
1	Wald im Pinzgau	17	Dorfbeuern
2	Stuhlfelden	18	Petting
3	Unken	19	Teisendorf
4	Maishofen	20	Surheim
5	Fusch	21	Anthering
6	Maria Alm	22	Seekirchen am Wallersee
7	Taxenbach	23	Elsbethen
8	Rauris	24	Faistenau
9	Schönau am Königsee	25	Fuschl am See
10	Mühlbach	26	Hallein
11	Hüttschlag	27	St. Kolomann
12	Zederhaus	28	Strobl
13	Muhr	29	Abtenau
14	Mariapfarr	30	Rußbach
15	Tamsweg	31	St. Martin am Tennengebirge
16	St. Georgen	32	Forstau

Tabelle 7: Aufnahmeorte DiÖ-Projekt

Nr.	Ort	Nr.	Ort
1	Maria Alm	4	Rußbach
2	Hüttschlag	5	Lessach
3	Berndorf		

Zahalit – how Israeli soldiers speak

Philipp Striedl (Ludwig-Maximilians-Universität München)
p.striedl@web.de

Abstract

This article is an introduction to linguistic phenomena which have their origins in the Israeli army and are subsumed under the term Zahalit*. The first part illustrates what* Zahalit *looks like with several examples. The examples are taken from written sources as well as from interviews with Hebrew speakers conducted by the author. Characteristic traits of* Zahalit *include abbreviations and acronyms along with metaphoric and metonymic constructions. Some of its linguistic properties, as well as common perceptions of Hebrew speakers of* Zahalit*, are displayed and the argument is made for these phenomena to be viewed as part of a single variety of Modern Hebrew. In the second part, the main functions of* Zahalit *are reviewed. It satisfies not only the communicational and professional needs of the Israel Defense Forces' (IDF) soldiers, but carries an implicit ideology and can fulfill political, as well as social and psychological functions. In fact, its functions can be understood as a continuum between professional language, secret language, and slang. Details regarding the social implications of* Zahalit *in Israel are given in part three. The role of the IDF in Israeli society and its impact on everyday life are discussed. Linguistic spillovers from the military to civilian society can be explained by the fuzzy boundaries of these two spheres and the influence that the IDF has on an individual's everyday life.*

Keywords: *Israeli army, Zahalit, linguistic variation, sociolinguistics*

1 What does *Zahalit* look like?

In Modern Hebrew (MH) one can come across linguistic constructions which have been originally used exclusively by soldiers in the Israeli army. Many of these constructions have made their way into everyday speech and can even be found in Israeli newspapers and books. Fiction writer Ilan Heitner occasionally uses such soldierly expressions in civil contexts in his books:

(1) עוד חצי שנה למנייאק (Heitner 2018: 122)
'od ḥatsi shana la-manayeḳ
another half year to-release
'half a year until release (from the military service)'

When used in its original context in the army, this expression means 'half a year until release (from the military service)'. It is uncertain how this expression developed. Rosenthal (2015: 124) points out that the original Arabic derogatory lexeme *manayeḳ* 'fucker' usually refers to military police officers in the soldiers' slang. On the day of one's release from the army, one has to pass the military police at the gates of the military compound. *Manayeḳ* possibly became conventionalized with this background in the soldiers' slang as a synonym for 'end of the military service'.

In the fictional text from which (1) was taken, the first-person narrator refers to the time remaining until the end of his family's one-year-long journey. The author's choice of this variant from soldiers' slang instead of a more common wording like *'od ḥatsi shana le-sof ha-ṭiyul* 'half a year until the end of the journey' conveys additional information about the narrator. First of all, people who use soldier slang have likely served in the Israeli army themselves (and

so has the narrator of the story). Furthermore, the choice of words suggests that he conceptualizes this journey with his family in a way similar to his experience during the military service – consecutive challenges to master in a limited period of time. This interpretation is supported by the narrator's repeated description of everyday tasks as military missions throughout the book.

The use of soldierly linguistic variants by the novel's protagonist achieves several literary effects. The ironic exaggeration in the comparison of day-to-day family life with military routine adds humor, though is quite difficult to decipher for anyone unfamiliar with Israeli military culture.

1.1 What does *Zahalit* mean?

In this article, I use the Hebrew term *Zahalit* to refer to the domain of the Modern Hebrew language to which all the linguistic phenomena presented here belong. I follow Rosenthal's (2014) use of this term according to his pioneer research on soldier slang in MH. As will be elaborated in 1.3, this designation is well-known to the wider non-expert public in Israel and therefore, is best suited for the discussed phenomena.

The term *Zahalit* itself is one of many characteristic examples of these phenomena and can only be rendered in English with further explanation. It was formed with the acronym for the Israeli army's official name *tsva' ha-hagana le-yisra'el* 'Israel Defense Forces', pronounced *Zahal*[1]. Together with the derivational suffix *-it*, which forms adjectives and designations of languages like *'anglit* 'English', *Zahalit* can be used to refer to the special kind of Hebrew originally used within the Israeli Army (IDF).

Acronyms and abbreviations are abundant in military vocabulary around the world (cf. Möller 2018 and Thorne 2006: 22), but they are also frequently used in Hebrew in general. Due to its main word formation strategies which are based on three root consonants, Hebrew is prone to the use of abbreviations just as if they were natural lexemes. *Zahal* fits perfectly in the noun pattern CaCaC (C stands for consonant) and possesses the phonological quality of a natural Hebrew word like *nahal* 'river'. Abbreviations can also be found in huge quantities in the Rabbinical literature and other contexts. The frequent use of abbreviations in both written and spoken language, in addition to the derivation of nouns, adjectives and verbs from abbreviations is typical for *Zahalit*.

1.2 Linguistic spillover from the military into the civil domain

Example (2) is taken from Leshem's bestselling novel *'im yesh gan 'eden* 'if there is a paradise' which tells the fictional story of a unit of young Israeli soldiers during the end of the South Lebanon conflict in the year 2000. This story is told in a realistic fashion from the unit commander's perspective and is filled with military terminology and slang.

[1] Instead of the standard transliteration *tsahal*, [z] is used here because the term *Zahal* is already established.

(2) כשמישהו היה מזכיר בטעות את אחותי ויקי, הייתי יוצא מאיפוס (Leshem 2005:22)

kshe-mishehu haya mazkir be-ṭa'ut 'et 'aḥot-i ẏiḳi
when-somebody be.PST.3MS mention.PRE.MS by-mistake ACC sister-my Viki

hayiti yotse' me-'ipus [2]
be.PST-1S exit.PRE.MS of-calibration
'when somebody mentioned my sister Viki by mistake I lost my mental balance'

In (2), the construction *yotse' me-'ipus,* which literally means 'leave the calibration', is particularly notable. *'ipus* is a technical term which means 'calibration' and is typically used in the context of handling weapons and watches. This example shows how this technical military term metaphorically extended to refer to one's mental constitution in general. The broader meaning of *'ipus* is so widespread that the *Free Online Hebrew Dictionary* (www.milog.co.il, accessed 21.01.2019) lists *'izun nafshi* 'mental balance' as a definition in the entry for '*ipus.*

A similar spillover of a term from the technical to the human semantic domain is described by Klemperer (1947:235) for the German lexeme *Einstellung* 'configuration (originally of a machine)', which became conventionalized with the additional meanings 'mental state' or 'worldview' at the beginning of the twentieth century and can be used nowadays to refer to one's political opinions, for example, as in *politische Einstellung*.

Rosenthal (2018) cites many examples for the use of Zahalit in civilian contexts and argues that spillovers take place in the semantic domains of *politics*, *management*, *sport*, *relationships*, *ways of behavior* and *words of empowerment.*

(3) כוננות ספיגה: ההגנה תתרסק להפועל גם מול מכבי תל אביב? (Rosenthal 2018)

konenut sfiga: ha-hagana ti-traseḳ le-hapo'el gam mul
state.of.alert absorption: DEF-defense 3FS-be.crushed.FUT to-Hapoel also in.front.of
makabi tel 'aviv?
Makabi Tel Aviv[3]

'State of alert: Will Hapoel's defense also be crushed in front of Makabi Tel Aviv?'

Example (3) is a newspaper headline about a football match between the teams *Hapoel* and *Makabi Tel Aviv*. Rosenthal (2018) explains how the meaning of *konenut sfiga* broadened as it was used often in an ironic manner in non-military contexts to indicate readiness to react to one's opponent. It was originally the IDF's designation for a defensive state of alert when attacks from the enemy were anticipated.

The above examples show that Zahalit is used in Israeli popular culture in contexts which have nothing to do with the army per se. The examples are instances of linguistic spillovers from Zahalit into MH. According to Rosenthal (2018), the influence of Zahalit on MH is considerable.

[2] PST.3MS = past tense, third person, masculine, singular; PRE = present tense; ACC = accusative
[3] DEF = definite; 3FS = third person, feminine, singular; FUT = future tense

1.3 How do Hebrew Speakers perceive Zahalit?

When I asked Israelis questions about Zahalit and its usage in everyday life, they all knew the linguistic phenomena to which I was referring. Most had already heard the term before and some even mentioned *Zahalit* before I had introduced the term into our conversation.

Hemmingby (2011: 3) claims that foreigners may have the impression of it as a "language within the language" and Israeli native speakers of Hebrew even use this concept when referring to Zahalit. Fania, aged 57, who served in the army, put it as follows:

> "In the IDF in Israel, there are a lot of words which belong to the army. Outsiders don't understand, but a part of the army words got transferred to the civil language […] and a lot of abbreviated words – acronyms – came from the army to the civil society, too. But the army invents new words all the time – I already don't know a part of the army words anymore."[4]

She argued further that Israelis who haven't been in the army, like Israeli Arabs or Ultraorthodox Jews, have difficulties understanding Zahalit.

Iris, a 53-year-old mother who never served in the IDF herself, told me that her son, who was doing his military service at the time of the interview, used many abbreviations and she had to ask him frequently about his special language usage because she sometimes has difficulties understanding him. She described Zahalit as follows:

> "They have abbreviations. They have an army language, slangs or… slang. Like kind of a special army language. I have to ask him [her son] all the time what does it mean?"

When I asked her about the use of Zahalit in everyday life, she declared:

> "I think that many things [expressions in Zahalit] – when they come back to the civilian sphere – then they already don't say them because it is only when you are in the army. There are things that stay but in general, they lose them."

In contrast, Yaniv, aged 28, who completed his military service some years ago, told me that he still uses Zahalit, but mainly when talking to his army friends.

When I asked what is characteristic of Zahalit in their opinion, all of the informants mentioned the great number of abbreviations. Yaniv added that there is a lot of imagery (*dimuyim*) in Zahalit and that a lot is based on the special speech conventions used for two-way radio communication in the IDF.

I found that people feel confident judging whether or not a word or construction is likely to have originated in the army. In many cases, I asked Hebrew speakers specifically if they would classify certain phenomena as belonging to Zahalit. For example, at least five informants judged *'ipus* as belonging to Zahalit originally, although it is used in civilian contexts as well (see example (2)).

[4] All interviews were conducted in Hebrew. The citations in English are my own translations.

The aforementioned utterances are evidence that at least the Israelis I asked are aware of the peculiarities of Zahalit and its special status within MH, as well as the difficulties that outsiders are likely to have understanding it. As all of my informants were neither linguists nor experts on the subject, it can be expected that many more Hebrew speakers are aware of the shared origins of the linguistic phenomena subsumed here under the term *Zahalit*. According to Krefeld (2015: 23), one can speak of a linguistic non-standard variety if co-occurring variants are perceived as belonging together or as one complex. If this is also the case with Zahalit, one can reasonably argue that it is a linguistic variety of MH.

1.4 What are the linguistic properties of Zahalit?

Zahalit is Modern Hebrew; its phonology, morphology, syntax, and orthography are nearly identical with that of Modern Hebrew. One of its most notable characteristics is its productivity and therefore, its changeability. Many new lexemes can be formed from one single lemma due to the variety of derivational strategies available in MH. For example, the blend *ḥapash*, 'simple soldier', is composed of the two words *ḥayal* 'soldier' and *pashuṭ* 'simple'. Further forms like *ḥapsh-an* 'passive, lazy person with lack of initiative' or the abstract noun *ḥapsh-an-ut* 'laziness, passivity' (Rosenthal 2015: 65) were derived and successfully conventionalized; even a verbal form *mitḥapshen* is used. The morphological structure of the new verb was adapted to the verbal pattern called *hitpa'el* which often carries a reflexive meaning or is used to indicate "initiate one's own activity upon oneself" (Bolozky 1982: 77). Therefore, *mitḥapshen* could be translated as 'laze about'.

When used in contexts outside the IDF, Zahalit occurs mainly as lexical variation, just like the above examples illustrate. It is hard to find utterances outside the IDF which are purely Zahalit, though they are possible nonetheless. Rosenthal at times uses catchy headlines composed entirely in Zahalit when he writes about the topic on his blog (www.ruvik.co.il). These examples are only understandable with some background knowledge and show how far Zahalit can deviate from MH. These examples are, however, highly artificial and would not plausibly be used in this manner.

Within the IDF, a higher frequency of Zahalit lexemes is likely to be used – especially in professional contexts which require specialized vocabulary. Zahalit is highly differentiated between the different units in the army (Rosenthal 2016). Another domain where Zahalit deviates considerably from MH is in two-way radio communication inside the IDF. In order to conceal transmission content from outsiders, a large amount of conventionalized lexical variants is used.

Zahalit deviates from the common morphosyntactic strategies of MH in some ways. One example is the use of the MH past tense – the historic perfective form – to convey an imperative meaning in commands.

(4) חמישים שניות הייתם כאן בחזרה. זוז (cited as in Rosenthal 2014: 217)
ḥamishim shniyot hayi-tem kan beḥazara. zuz
fifty seconds be.PST-2PL here back move
'In fifty seconds, you are going to be back here – move!'

In (4), the speaker commands someone to run to a designated place and to be back within 50 seconds. The command phrase is kept as succinct as possible. In this specific context, it is clear to the recipients that *ḥamishim shniyot* 'fifty seconds' indicates the available time for the task.

Usually, a temporal designation of this sort would include the temporal adverbial *'aḥarey* 'after' in MH, which is omitted here. The most significant variant in (4) is the past form *hayitem* which is used instead of a future form like *ti-hiy-u* (2-be.FUT-P[5]). This is widespread in commands in Zahalit. The past form here more resembles the verbal systems of Classical Arabic or earlier stages of Hebrew, which mark the imperfective and perfective aspects instead of marking tenses.

Without the availability of proper corpora of Zahalit, it is hard to describe its grammatical properties on the level of syntax and morphology in detail. The main obstacle for data collection of Zahalit inside the IDF is the seclusion of military institutions from the public due to matters of security. For more examples with their translation and further remarks about the conventionalization and the use of metaphors in Zahalit see Striedl (2018).

In conclusion, Zahalit does not necessarily stand out in terms of its linguistic properties, but rather its semantic and etymological peculiarities, the effects it can have on its recipients and its diverse functions for the speakers.

2 Zahalit's functions

Every language variety adapts to the communicational needs of its speakers and fulfills important functions (cf. Labov 1972). In the following, the main functions of Zahalit will be reviewed. It should be kept in mind that the borders between the functional domains are fluid and that one and the same phenomenon in Zahalit can fulfill several functions at the same time.

Figure 1 was designed to illustrate the functional continuum of Zahalit and its understandability for non-expert speakers of MH. There is one axis at the top which represents the understandability of the italicized Zahalit phenomena in accordance with their contexts of usage (in bold letters). The bottom axis divides the contexts from less to more specific, ranging from situations in everyday life to situations in specific units within the IDF. The degree of variation from MH increases in more specific contexts. One elliptic form, Zahalit slang, is adjacent and/or overlaps with almost all entries to illustrate its huge scope, comprising of phenomena from diverse categories which can be used in nearly any context.

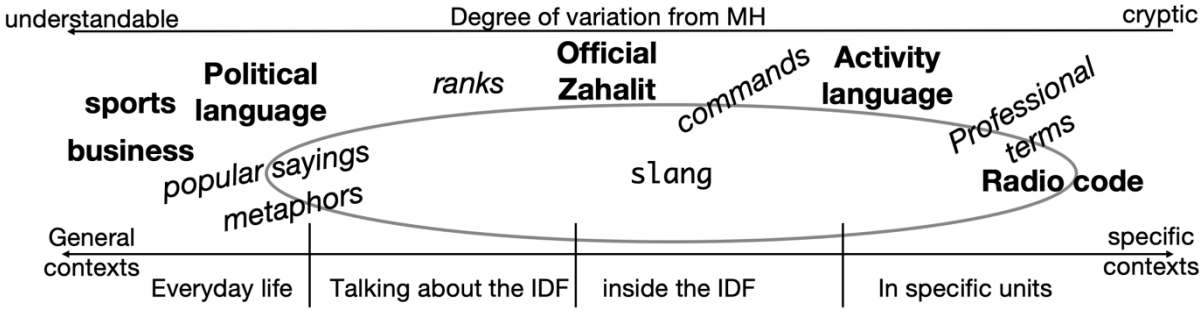


Figure 1: The functional continuum of Zahalit

[5] P= plural

2.1 As a professional language

MH itself was first established at the beginning of the twentieth century as the language of the Jewish population in the British mandate of Palestine (Shapira 2012:144). At almost the same time, the military institutions of the Jewish settlement became more and more organized, and the first Hebrew speaking soldiers developed special vocabulary and language conventions for their communicative needs (Rosenthal 2014: 37). A professional Hebrew language for the army was created in the same way as for other domains of modern life. Shortly after the IDF was officially instituted as Israel's army with the foundation of the Israeli state in 1948, the first dictionary of Hebrew military terms, which reflects Zahalit's function as a professional language, was published (see Akavia 1951).

2.2 As a secret language

Parts of Zahalit were originally developed as a secret language to restrict access to official military communication and render it difficult for outsiders to decipher. For this purpose, the IDF have been using a special code for their two-way radio communication in which the strategy of *Neosemantisierung* (Siewert 2018: 19) is, therefore, characteristic. *Neosemantisierung* occurs when one lexeme from the standard variety is used with a different meaning. For example, *gafrur* 'match' is used as a metaphor to mean 'soldier'. This strategy is common in military contexts and a similar system was used by German pilots during the Second World War (cf. Siewert 2018: 20). Over time, a lot of the coded vocabulary lost its secrecy and lists of coded words with their meanings can be found online. Some of these words were used outside of radio communication and have made their way into soldiers' slang and even civilian spheres (cf. Rosenthal 2014: 46 and section 1.3).

This special kind of slang is also depicted in TV programs like the popular series *Ramzor*. The main character, Hefer, says *tamtini ḳṭana* 'wait a minute' in one scene when he is confronted with a problem and goes into an army-like mindset (Ramzor season one, part 9, minute 16). This utterance is part of the radio code. *Tamtini* 'wait.2F.IMP', which originally belongs to a higher register of Hebrew was conventionalized as 'wait' whereby *ḳṭan-a* 'small-F' means 'minute'.

Even if the way soldiers speak is not completely secret, it can still function as an authentication strategy. This is underlined by a report about attempts to infiltrate the IDF via smartphone applications. The report detailed how IDF soldiers were approached by false actors using nearly authentic slang (Föderl-Schmidt 2018).

2.3 From the sociological point of view

Rosenthal (2014) shows how Zahalit mirrors the hierarchy inside the IDF in both directions top-down and bottom-up. With the official language, which comprises the commands and the ranks, the formal structure of the IDF is expressed from above. On the other hand, in the soldiers' slang, there are many pejorative and ridiculing lexemes which express a negative attitude towards the authorities within the institution.

Hierarchical differences between the soldiers are expressed, for example, according to one's *pazam*, which is an acronym of *perak zman miz'ari,* which is literally for 'minimal period of time' but is used as 'the remaining period of time until the end of the service'. My informant Yaniv explained that you can't ask *'ad matai* 'until when' unless you "are the most *pazam*", which means 'the most senior soldier in the speech situation' (cf. Rosenthal 2015: 124, see Möller 2018: 152-154 for the description of a similar system in the army of the GDR). Other

hierarchical structures inside the IDF based on gender, ethnicity, religiosity and the contrast between combat soldiers and non-combat soldiers are expressed with Zahalit (Rosenthal 2014).

Like every slang, Zahalit creates a sense of belonging which can be invoked again in the civilian sphere in its use with army friends or even when politicians utilize it to address a particular audience (see Stenström 2009: 2 for a general description of slang).

2.4 From the psychological point of view

It is most obvious with regard to slang that Zahalit has psychological functions due to its use to express a soldier's identity as both a living being, not a machine, and as a young person. In Zahalit, there are taboos surrounding death and injured comrades. In the radio code, a dead soldier is called *harduf* 'Oleander' and an injured soldier *peraḥ* 'flower', whereas enemies are called *melukhlakhim* 'dirty ones'. These terms disguise the drastic events in a combat situation using linguistic means.

Many expressions from the soldiers' slang are filled with humor and serve to soften the often-harsh reality in the army. Small acts of rebellion with linguistic means are possible, for example, when one refers to a senior officer as *falafel* because of the symbol made of round leaves resembling the food on his/her epaulettes. As displayed in Striedl (2018: 180-182), metaphors in Zahalit are often taken from the source domains *flora, food, childhood,* and *Jewish culture*. Using familiar terms like *mamtaḳim* 'sweets' for 'cartridges' or *'aba'* 'father' for a 'commander' functions to trivialize the military environment.

3 Why is Zahalit special?

The IDF is omnipresent in Israeli society. The *ramaṭkal* (abbreviation for) 'Chief of staff of the IDF' is as prominent as the most important politicians and can even challenge their power.

One can follow the IDF's past and present activities on a daily basis in nearly every Israeli newspaper and on every Israeli news channel. The IDF itself has been broadcasting its own radio program on the stations *Galats* and *Galgalats* for many years and these programs are very popular. The IDF also has its own YouTube channel[6], as well as accounts on Facebook, Instagram, and Twitter, where they address a young target group.

The structure of the military service in Israel is legislated primarily according to the 1986 Defence Service Law. With only a few exceptions (for Arabic Israelis and Ultra-Orthodox Jews), the military service is mandatory for men for a period of 32 months and for unmarried women for two years. New immigrants under 22 years old have to serve, too. After completing the military service, many Israelis are assigned to reserve duty, which includes military training and participation in military operations on a regular (monthly or yearly) basis. The possibility of frequently returning home during obligatory military service aids considerably in blurring the boundaries between the civilian and military sphere.

The military institutions in Israel have been crucial throughout the consolidation of the state and the shaping of its ideology. From the beginning of the 19th century, the use of Hebrew among the Jewish population in mandatory Palestine carried ideological weight and functioned as a symbol for the Zionist-nationalist ambitions of the political leaders, as well as for the emerging Israeli culture (Shapira 2012: 143-144). This is also true for Zahalit, which became the means with which to express the ideology of the new state and the values of the IDF. The Israeli army grew to be a nearly mythological symbol and a powerful institution at the same

[6] On its channel (https://www.youtube.com/user/idfnadesk) one can find introductory clips which explain aspects of the army life and even terms in Zahalit.

time. For generations of native Israelis from the 1930s onwards, the *mitos ha-tsabar ha-loḥem* 'the myth of the fighting cactus pear' (Almog 1997: 187) has been a cultural ideal. *Tsabar* is a common metaphor for 'native Israeli'. In his portrait of the cultural prototype *tsabar*, which has an illustration of a soldier on its cover, Almog (1997) shows how influential the pioneer-fighter ideal was for Israeli culture in the 20th century. He frequently cites the *tsabar*'s special use of Hebrew, which is closely related to Zahalit.

Whether Israel is a militaristic society or not has been a recurring question in many political and sociological analyses over the past sixty years. The different scholarly approaches to this question are reviewed in Sheffer and Barak (2016). Notwithstanding, it is unquestionable that the IDF exerts a huge impact on individuals' lives and their worldview for a considerable amount of time. Thus, my informant Yaniv recapitulated his experience in the army as follows: "It's as if the entire army is brought into your life after that, too."[7]

The symbolic weight of the IDF in the ideological framework of the state, its presence in the media and the blurring of the civilian and the military sphere in everyday life cannot be found easily in other societies today, which contributes to Zahalit's uniqueness.

4 Outlook

Zahalit is far from well-researched and a thorough analysis of its linguistic properties is needed. To do so requires more linguistic data in the context of its usage and ultimately, a compiled corpus of Zahalit.

From a historical linguistic perspective, it would be interesting to investigate how and to what extent the military institutions in Israel participated in the shaping of MH and its propagation. Another extensive research question is: which role does the IDF play today with regard to the dynamics of MH and particularly MH slang?

References

Almog, Oz. 1997. הצבר- דיוקן [The Sabra – A Profile]. Tel Aviv: Am Oved Publishers.

Akavia, Avraham. 1951. *Milon le-munhe tsava/ dictionary of military terms (English-hebrew)*. Haifa: Hotsaat Magen.

Bolozky, Shmuel. 1982. Strategies of Modern Hebrew verb formation. In: Golomb, David, Ramras-Rauch, Gila (eds.), *Hebrew Annual Review* 6, 69-79. Columbus, Ohio: The Ohio State University.

Föderl-Schmidt, Alexandra. 2018. Israelische Soldaten fallen auf Flirt-App herein. https://www.sueddeutsche.de/politik/israel-israelische-soldaten-fallen-auf-flirt-app-herein-1.4040172 [accessed 2019-01-17].

Heitner, Ilan. 2018. *7250*. Ben Shemen: Modan.

Hemmingby, Cato. 2011. *Modern hebrew conflict and military terminology*. University of Oslo: Master's Thesis.

Klemperer, Victor. 1947. *LTI Notizbuch eines Philologen*. Berlin: Aufbau Verlag.

[7] The original in Hebrew: "כאילו, כאילו כל הצבא מושלך גם לחיים אחר כך"

Krefeld, Thomas. 2015. Über die Perzeption von Variation und den leichtfertigen Umgang mit dem Ausdruck <Varietät>. Vortrag an der Uni Wien. https://www.romanistik.uni-muenchen.de/personen/professoren/krefeld/download/perzeption_variation_varietaet.pdf [accessed 2019-01-16].

Labov, William. 1972. Academic Ignorance and Black Intelligence. *Atlantic Monthly* 229, 59-67.

Leshem, Ron. 2005. אם יש גן עדן [If there is a paradise]. Or Yehuda: Zmora-Bitan.

Möller, Klaus-Peter. 2018. Außer Kontrolle. Die DDR-Soldatensprache als ein Dokument der Verweigerung. In: Hardy, Stéphane, Herling, Sandra, Siewert, Klaus (eds.), *Kontrollierte Kommunikation*, 161-192. Hamburg/Münster: GSV Geheimsprachen Verlag.

Rosenthal, Ruvik. 2014. השתקפות ההיררכיה הצבאית בשפת הצבא הישראלית [The reflection of the military hierarchy in the language of the Israeli army]. Bar-Ilan University: PhD Thesis.

Rosenthal, Ruvik. 2015. שירת הפזמניק. המילון הצבאי הלא רשמי [Soldier's Muse. The Unofficial Dictionary of the Israeli Army]. Jerusalem: Keter.

Rosenthal, Ruvik. 2016. Hierarchy and the language of the Israeli army: A sociolinguistic case study. Proceedings of the workshop 'The Lingustic and Cultural Manifestations of War and Terrorism' held at the University of Lodz.

Rosenthal, Ruvik. 2018. אחי, תקפיץ אותי לקניון, יש מבצע פיצוץ. http://www.ruvik.co.il/פינת-עיון/2018/זליגת-הצבאי-לאזרחי.aspx [accessed 2019-01-08].

Siewert, Klaus. 2018. Typen und Muster der Codierung in deutschen Geheimsprachen. In: Hardy, Stéphane/ Herling, Sandra/ Siewert, Klaus (eds.), *Kontrollierte Kommunikation*, 9-23. Hamburg/Münster: GSV Geheimsprachen Verlag.

Shapira, Anita. 2012. *Israel: A History*. Waltham, Massachusetts: Brandeis University Press.

Sheffer, Gabriel/ Barak, Oren. 2016. Militarism and Civil-Military Relations in Israel: A New Approach. In: Ben-Rafael, Eliezer/ Shoeps, Julius H/ Sternberg, Yitzhak/ et al. (eds.), *Handbook of Israel: Major Debates*, Volume 1, 590-608. Berlin/ Boston: De Gruyter.

Stenström, Anna-Brita/ Jörgensen, Anette Myre. 2009. Youngspeak in a multilingual perspective: Introduction. In: Stenström, Anna-Brita/ Jörgensen, Anette Myre (eds.), *Youngspeak in a multilingual perspective.* Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Striedl, Philipp. 2018. Zahalsprache. Sprachliche Charakteristik der militärischen Institutionen in Israel. In: Hardy, Stéphane/ Herling, Sandra/ Siewert, Klaus (eds.), *Kontrollierte Kommunikation*, 161-192. Hamburg/Münster: GSV Geheimsprachen Verlag.

Thorne, Steve. 2006. *The language of war*. London: Routledge.

Trollspeak: who do Russian trolls tweet like?

Martin Eberl (Ludwig-Maximilians-Universität München)

martin.eberl@lmu.de

Abstract

The present study is concerned with the link between political affiliation and linguistic variation on lexical as well as sublexical levels on Twitter, with a focus on tweets regarding the 2016 US elections. Linguistic variation depends on a variety of factors and reasons, including social identities. As previous research has shown, political affiliation and associated sociolinguistic variation can be both reflected in speech and in writing. Variables in the latter may even include sublexical features such as punctuation. This study examines two datasets of tweets, one from supporters of the presidential candidates Hillary Clinton and Donald Trump, and the other from Russian operatives engaging in a misinformation campaign, as to their lexical variation in the text, variation in the use of emojis and hashtags as potentially metatextual[1] features and variation in the use of punctuation as a sublexical feature. The results show that the Russian tweets have their own, distinct features, while tweets from Trump and Clinton supporters are remarkably alike.

Keywords: *sociolinguistics, social media, twitter, variation, metatextual features*

1 Introductory remarks

Social media has become increasingly political since its inception. Twitter, for example, "has emerged as a key platform on which anyone with a smartphone can engage in political discourse" (Nguyen 2017). This increasing political aspect of Twitter and other platforms involves not only the offices of political parties and individual candidates running specific social media campaigns, but also discussions between normal users of the platform sharing their political opinions. As in other spheres, they seem to seek out those that share their opinions, leading to a polarization of the userbase (Conover 2013). But does this polarization show itself only in who the users talk to and what or whom they talk about? Or does it also reveal itself in how we talk, i.e. in sociolinguistic patterns?

In Labov's (2010) framework, for instance, political affiliation may reveal itself in sociolinguistic patterns, or, conversely, it may be presupposed by the hearer due to the speech of the interlocutor. However, such variation is not limited to speech, but may also show itself in written discourse Schnoebelen (2012) has shown that stylistic variation may occur even in metatex-tual features such as Twitter emoticons. In a similar vein, Tatman and Paullada (2017) have shown, for instance, that such variation extends even to sublexical features such as punctuation. In their data, supporters of the Black Lives Matter movement tend to use less punctuation with relatively uniform distribution across tweets, whereas supporters of the Blue Lives Matter movement use more punctuation, which is mostly centered around the beginning and end of tweets. In addition, they found distinguishing lexical features such as the 'fist' emoji being used exclusively by Black Lives Matter tweeters, while the 'white star' emoji was used solely by those in favour of

[1] For a debate on whether eomjis and hashtags are textual or metatextual features, see, among others, Zappavigna 2015 and Rambukanna (2015). I follow the latter in arguing that they should be treated "as both text and metatext simultaneously" (2015, 161). As such, even when hashtags are included in running text rather than attached separately, they serve functions beyond the utterance itself, such as topic marking, the reference to and activation of interpersonal relationships and structuring the text (Zappavigna 2015, 274). Therefore, hashtags, as used on Twitter and other social media platforms, cannot be simply textual.

Blue Lives Matter. Further research (Golbeck et al. (2011), Hu et al. (2013)) has shown that personality traits may be predicted from Twitter data and that language on Twitter shows distinct features when compared to other computer-mediated communication. Taken together, these findings indicate that sociolinguistic studies of Twitter data, as far as it is available, are a promising endeavour.

As luck would have it, a large trove of Twitter data from a specific politically motivated group was made available recently. On February 14th, 2018, NBC news published a database of 203,451 tweets that Twitter claimed to have engaged in 'malicious activity' during the 2016 presidential election. Said malicious activity was the undue influence of the presidential election. The users behind these tweets, then, comprise one political group with a specific agenda and should show similar sociolinguistic patterns.

I examine the tweet content for patterns of variations on three distinct levels: lexical variation in the text, variation in the use of emojis and hashtags as metatextual features and variation in the use of punctuation as a sublexical feature. The same analysis is then carried out on a set of tweets by users who have been identified as politically conservative and politically liberal from roughly the same time period. Taken together, these three groups should be identifiable through linguistic patterns.

In analysing these three datasets, the present study therefore seek to answer this central question: is political affiliation a strong enough factor to influence sociolinguistic variables on social media platforms such as Twitter?

2 The data

The first dataset comprises the tweets from so-called "Russian trolls", i.e. those tweets removed by Twitter on suspicion of having been written and disseminated by Russian operatives trying to sow discord before, during and after the election of 2016. This dataset was collected by the American news company NBC and made publicly available for research purposes. NBC describes the Russian operation as thus:

> [T]hese accounts, working in concert as part of large networks, pushed hundreds of thousands of inflammatory tweets, from fictitious tales of Democrats practicing witchcraft to hardline posts from users masquerading as Black Lives Matter activists. Investigators have traced the accounts to a Kremlin-linked propaganda outfit founded in 2013 known as the Internet Research Agency (IRA). (Popken, 2018)

After Twitter had deleted the tweets in question, along with the accounts who had written them, "three sources familiar with Twitter's data systems cross-referenced the list of names released by Congress, excluding any account that Twitter later restored, to create a partial database of tweets that could be recovered from the suspended accounts". The restored dataset comprises 203,541 tweets from a total of 453 different accounts. For the purposes of this study, however any retweets were excluded, reducing the number to 55,889 analysable tweets. The tweets are dated from June 14th, 2014 to July 26th, 2017.

The second dataset has been sampled directly from Twitter using the public Twitter API and the *rtweet* package for R by Michael Kearney (2017). The goal was to identify political supporters of either Donald Trump or Hillary Clinton who had tweeted in favour of their candidate during the election process. This proved to be a significant problem, given that they could not be identified automatically, for instance through their use of hashtags, so as not to affect the later results of the linguistic and metalinguistic analysis. Thus, the individual

accounts had to be identified manually, before their Twitter timeline was sampled through the *rtweet* package. The criteria included their vocal support for one candidate, having tweeted multiple times in favour of their candidate, and activity throughout the election process. In addition, members of any of the various political campaigns as well as politicians and journalists were deliberately excluded from the dataset, as were accounts used by more than one person. This manual selection process identified 536 total accounts of interest. 76,878 tweets were sampled from these accounts. Excluding retweets, this left 39,152 tweets for analysis. Of these, 20,775 were made by Clinton supporters and 18,377 were made by Trump supporters. The maximum amount of tweets coming from a single user was set at 160 tweets, so as not to give a few users the ability to dominate the dataset. The median amount of tweets were 73 for the Trump supporters and 76 for Clinton supporters. The tweets were written between September 1st, 2010 and July 27th, 2017.

The figure below shows some exemplary tweets from the first datasets. Note that the tweeters' usernames, their Twitter handles and the time of the tweets have been withheld for these examples due to privacy concerns.[2]

Once gathered, the individual tweets were tokenized for the following analysis by using the *stringr* package for R (Wickham 2019). An additional pre-processing step was not necessary due to the output format of *rtweet*. Tweets were tokenized into individual words for analysis of lexical and metatextual features first, with a second tokenization step into individual characters to later determine the position of sublexical features.

Using the *str_split* feature of the *stringr* package allowed for exact tokenization on the basis of new lines and spaces as well as punctuation. In rare cases, it failed to properly extract words when written without spaces (either due to space issues or due to typos). The ad-ditional *str_extract_all* function was used to extract all strings beginning with a #, which allowed for the listing of hashtags. Emojis had to be identified using different methods for the Russian dataset when compared to the Trump and Clinton datasets. The former included encoding of emojis in 4 byte strings (starting with D), whereas *rtweet* gave the unicode for emojis (starting with <U+, e.g. <U+0001F602>). Given that they were encoded consistently throughout both datasets themselves, however, identifying them could also be achieved through the *str_extract_all* function.

[2] I am aware that there is a debate on how best to handle reproducing tweets in a scholarly context. Some researchers prefer giving the full account information for copyright reasons, while others prefer removing any information through which the authors can be identified due to privacy concerns, going so far as to replace lexical items in the tweets themselves. In my opinion, privacy concerns should take precedent as far as possible, but the linguistic examples themselves should not be altered. I have therefore opted for the middle ground to preserve the linguistic structure of the tweets themselves while making the authors not immediately identifiable.

Figure 1: Exemplary Tweets

3 Lexical variation

The first factor investigated in regard to the lexical variation of the three sets of tweets was their type-token-ratio. As the table below shows, there is very little difference between the ratios of the individual sets, with Russian tweets showing a very slightly lower type-token-ratio overall than that of Trump supporters, who in turn show a very slightly lower type-token-ratio than Clinton supporters.[3] In a sample of the Russian dataset roughly matching the size of the other two datasets, the TTR was 0.11 as given in the parentheses.

Table 1: Type-token ratios of all three datasets

Russian	Pro-C	Pro-T
0.09 (*0.11)	0.11	0.10

The uniformity of the type-token-ratio may suggest that this datapoint is varied less according to the individual users or their political stances, but is more dependent on the limitations of Twitter as a messaging platform, with the text being capped at 140 characters. Previous studies on TTR have shown the length of the text to affect the average type-token ratio (Caruso et al. 2014, 139), so it would seem logical to assume that tweets show similar type-token ratios. The tweet length of the various datasets is summed up in figure 2 below. It shows a density plot of all three datasets across the length of tweets in characters, meaning that higher values of density show a higher number of tweets with that particular length.[4]

[3] TTR is calculated for each dataset as a whole rather than individual tweets.

[4] Note that due to technical issues with the readout of tweets (such as encoded emojis accounting for more) characters than their graphical equivalents), tweets in the datasets can be up to 150 characters. Since this affects all three datasets equally, the ratios between the datasets remain unaffected.

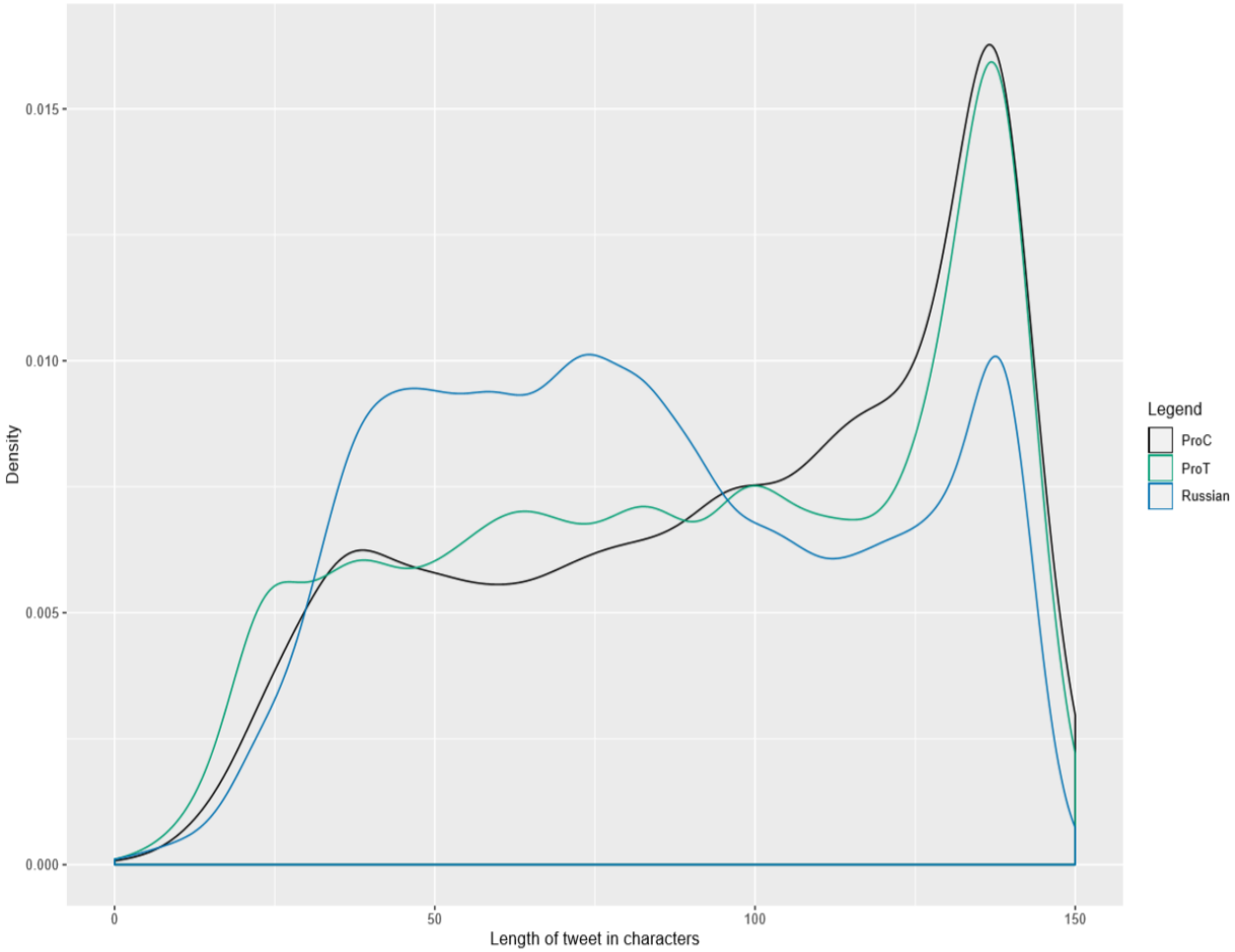


Figure 2: Length of the tweets in characters

As is evident, the character length of Trump and Clinton supporters is very similar, with a peak towards the maximum tweet length of 140 characters, which suggests that the users adjusted an originally longer message to the platform's character limit. The Russian tweeters seem to have hit this limit far less often, and show a similar peak towards shorter messages between 35 and 75 characters. This may suggest that their tweets are less spur-of-the-moment and may have been pre-planned. It also explains their lower type-token ratio as explained above, however.
A second variable to be explored were the most frequently used nouns and adjectives in the three datasets. These were identified through a manual check of the word frequency lists generated for each dataset. The six most common ones for each are shown in the table below, with their absolute frequency:

Table 2: Frequent Nouns

Russian	**Pro-C**	**Pro-T**
trump (9,399)	hillary (1,171)	trump (4,552)
clinton (4,345)	trump (928)	people (1,064)
politics (3,709)	day (680)	hillary (738)
hillary (3,561)	time (587)	don (680)
obama (2,540)	president (516)	trumptrain (653)
midnight (2,342)	people (472)	potus (617)

Table 3: Frequent Adjectives

Russian	**Pro-C**	**Pro-T**
black (889)	good (457)	new (842)
new (889)	great (437)	great (620)
american (734)	new (429)	good (396)
good (682)	happy (269)	happy (292)
great (586)	real (227)	best (224)
white (575)	proud (175)	free (213)

Quite unsurprisingly, the candidate's names are among the most frequent nouns, with the candidate the tweeter identified with leading the count and the opposing candidate closely behind. Russian tweeters also mentioned the sitting president, Barack Obama, frequently. *Time*, in the tweets of Pro-Clinton tweeters, seemed to mostly refer to the time for a female president having arrived. In the end, however, the frequent nouns offer little insight into sociolinguistic tendencies.

The frequent adjectives are slightly more interesting. Note that they are exclusively positive for both Pro-Clinton and Pro-Trump tweets, and very similar among these two groups. One might have thought for *great* to be more favoured among Pro-Trump tweeters, but the Pro-Clinton tweeters seem to have countered the idea of 'making America great again' with the idea that 'America is already great'. Russian tweeters, aside from the same positive adjectives *new, good* and *great* that the other groups use, quite ironically make frequent use of *american* as well. This is usually connected with 'American values' that the voter should keep in mind. Also note that they frequently use both *black* and *white*, which may reflect an effort to stoke racial tensions ahead of the election. Once again, the differences are more pronounced between the Russian tweets and the other two sets than they are between Clinton and Trump supporters.

4 Use of hashtags and emojis

Aside from the lexical variation, I also examined the variation in metatextual features such as hashtags and emojis. Their patterns reflected some of the tendencies already observed above. The following table shows the most frequent hashtags among the three datasets:

Table 4: Most frequent hashtags among the three datasets

Russian	**Pro-Clinton**	**Pro-Trump**
#politics	#ImWithHer	#Trump2016
#news	#StrongerTogether	#MAGA
#Merkelmussbleiben	#quote	#MakeAmericaGreatAgain
#TrumpForPresident	#DemConvention	#TrumpTrain
#IslamKills	#Hillary2016	#quote
#PJNET	#GOPdebate	#Trump
#tcot	#mapoli	#POTUS
#Brussels	#Election2016	#NeverHillary
#StopIslam	#HillYes	#maga

As can be gleaned from the table, variation between highly frequent tokens of Clinton and Trump supporters is, as has been the case with the nouns above, mostly limited to the candidates' names and their slogans. While Pro-Trump tweeters prefer #Trump2016 and #MAGA, Pro-Clinton tweeters opt for her slogans of #ImWithHer and #StrongerTogether. It may be surprising to some observers that only one of the most frequently used hashtags is a negative hashtag oriented at the political opponent, namely #NeverHillary, which is, of course, an adaptation of the #NeverTrump slogan that Republicans used to show their party-internal opposition to nominating Donald trump as their candidate. Pro-Clinton supporters also seemed to be more interested in electional events such as the Democratic national convention

(#Dem-Convention) and the Republican debates (#GOPdebate). The variation in Russian lexical items and hashtags hints at a broader array of topics, including highly controversial debates and islamophobic concepts. Also note that it is not limited to discussions surrounding the U.S. elections, but also includes other international political hashtags such as the German #Merkelmussbleiben. In addition, Russian users also use more hashtags overall than the other two groups: whereas Clinton supporters use 0.51 hashtags per tweet and Trump supporters use 0.64, the Russian tweets contained 0.86 hashtags per tweet on average[5]. Once more, we find that Russian tweeters seem to be concerned with a variety of political topics, whereas Pro-Clinton and Pro-Trump users show little variation.

Emojis vary a bit more between the datasets. In general, Russian emoji use is much lower, clocking in at only 0.06 emojis per tweet, whereas Pro-Clinton users employed 0.29 emojis per tweet and pro-Trump users employed 0.33 emojis per tweet[6].

Table 5: Emojis per tweet

Russian	**Pro-C**	**Pro-T**
0.06	0.29	0.33

In qualitative terms, there are also significant differences between the emojis pro-candidate tweeters used versus those used by the Russian tweeters. The ever-present 'tears of joy'-emoji, for instance, was the most common emoji among both pro-Trump and pro-Clinton tweeters. Russian tweeters, however, have a stronger preference for the 'sad' emoji. The table below sums up the six most common emojis of the various datasets:

Table 6: Frequency of emojis

Russian	**Pro-C**	**Pro-T**
sad (269)	tears of joy (399)	tears of joy (575)
vehicle (108)	black heart (192)	dash symbol (369)
chick (79)	fire (162)	steam loc (173)
flag (76)	clapping (132)	right arrow (160)
pointdown (65)	cryface (115)	rightward hand (123)
elephant (55)	camera (91)	fire (111)

In addition to the difference between 'tears of joy' and 'sad', there is almost no overlap between the frequent emojis. The 'steam loc' emoji, for instance, is only used by Trump supporters in parallel to the *trumptrain* being one of the most common nouns. On the other hand, only Pro-Clinton users seemed to have been on 'fire' and 'clapping' frequently. It is here, then, that we may observe the sociolinguistic patterns that differ between the latter two groups. There is also a possible argument to be made about the 'sad' emoji being used, once more, to emphasize

[5] The total numbers are 10,567 hashtags for Pro-Clinton tweeters, 11,773 for Pro-Trump tweeters and 48,216 for the Russian tweeters.

[6] The total numbers of emojis identified in the datasets were 1,836 for the Russian dataset, 6,866 for the Pro-Trump tweeters and 5,405 emojis in the Pro-Clinto dataset.

negative emotions and connotations with the state of the Union, as it were, to further sow discord among the electorate.

5 Sublexical variation

As a final variable in this study, the use and position of both punctuation marks and hashtags was examined. The first category included exclamation marks, question marks, quotation marks, periods and commas across the length of the tweets[7]. The three plots in figure 3 below show the general punctuation patterns of these signs across the three datasets:

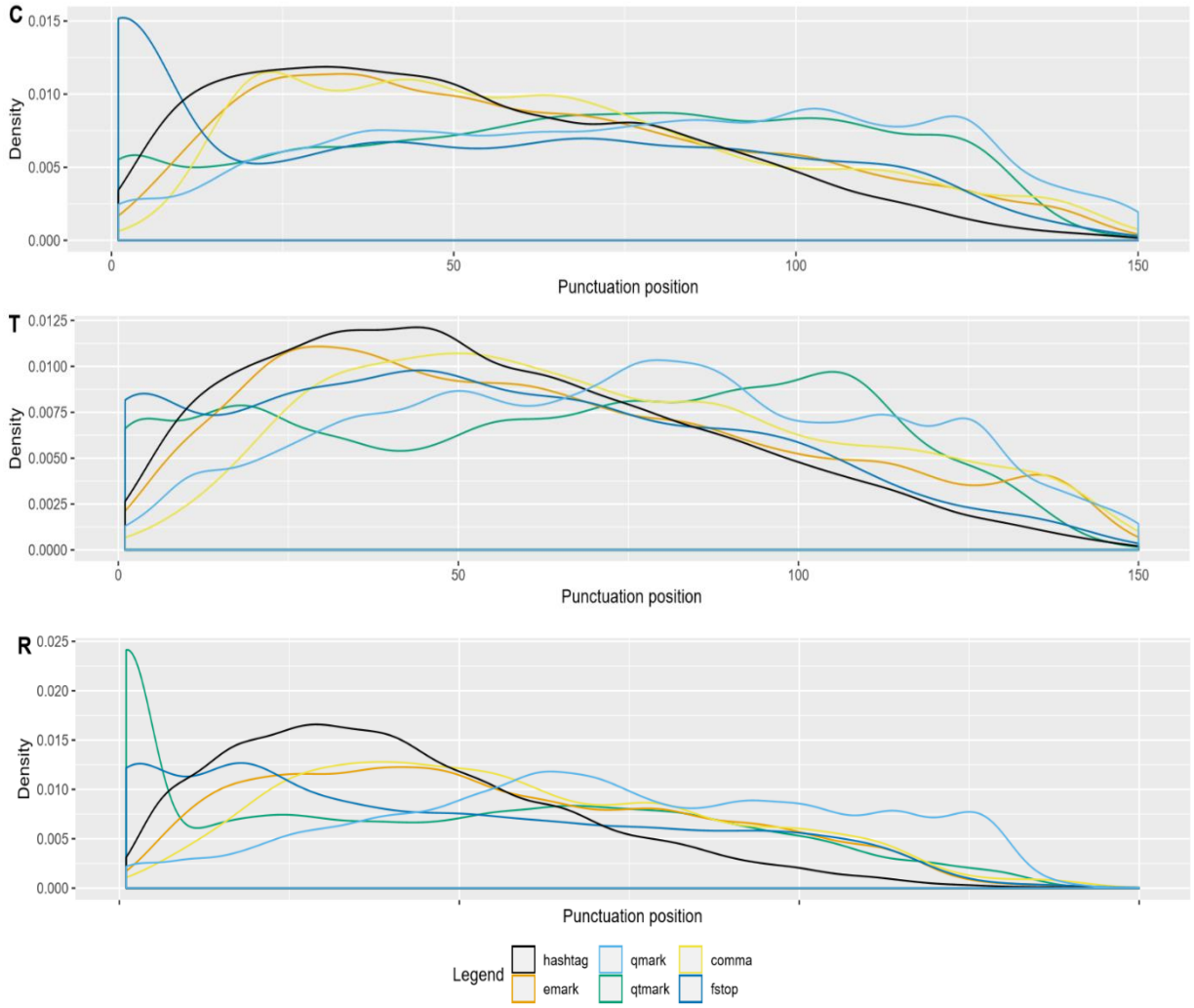


Figure 3: Position of various punctuation marks across tweets

[7] Due to technical processes in the sampling of the tweets through Twitter's API and the way that the *rtweet* package works, the maximum length of the tweets in the graphs below is 150 characters instead of 140 characters. This does not, however, affect the general patterns observed.

From this data, it seems that the general position of punctuation marks and hashtags is similarly spread out across pro-Trump and pro-Clinton tweeters as well as the Russian tweeters. Figure 4 below focuses more closely on the position of individual signs as they appear in Russian tweets, pro-Clinton tweets and pro-Trump tweets.

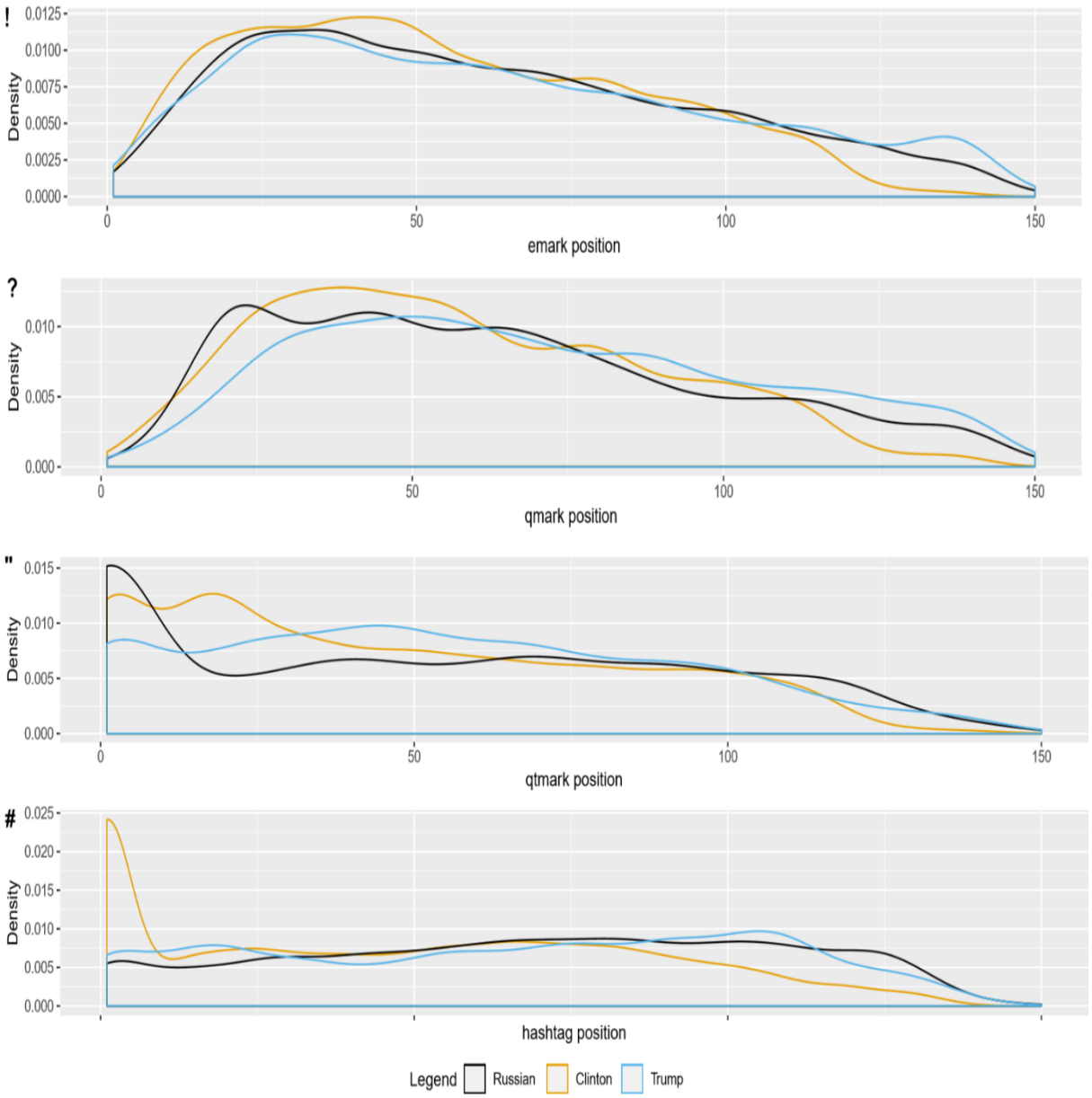


Figure 4: Position of specific punctuation marks across the datasets

As is evident, the use and position of both exclamation marks and question marks match closely across all datasets. Deviations in more discourse-related punctuation include the prominent fronting of hashtags in the Russian tweets. This adds up with the lower overall use of hashtags by Russian tweeters to suggest a different use of the metatextual feature as such.

In addition, there seems to have been a propensity by Clinton supporters for longer quotes (mean character count 53, mean word count 9.5) over Trump supporters (means 33 and 55.7), who are also slightly more likely to use quotation marks for single noun phrases. Russian quotes have a mean of 45 characters and 8 words) and thus lie between.

6 Conclusion

The results of earlier studies pointing to different use of sociolinguistic variants could only be partly replicated. While there is some variation on the lexical level, the tweets of Clinton and Trump supporters are remarkably alike on a structural level, including the use of metatextual hashtags and punctuation. As a result, the patterns observed by the Russian troll tweeters do not match either of the groups, but show some of their own, distinct features.

These patterns paint a similar picture across the three categories of features surveyed. In regard to the lexical variation, Russian tweets used slightly different adjectives, playing up racial tensions between 'black' and 'white'. They also referred back to the sitting president Obama far more than the other two groups. The variety of their lexical items shows a broader picture less concerned with the individual candidates and more concerned with American and international politics at large. In terms of metatextual features, they were more likely to use hashtags and showed a higher propensity for fronting them in their tweets, both of which are likely intended to get hashtags trending and influence current discourse topics on Twitter. On the sublexical level of punctuation, there is once again little variation: exclamation marks, question marks and commas closely match, suggesting that these variables are influenced more by the structural environment and structural limitations of Twitter rather than sociolinguistic factors and influences.

The usage of emojis was the only variable in which all three groups differed significantly, and, maybe more importantly, a variable in which the pro-Clinton and pro-Trump groups showed some variance. This may suggest that emojis are a feature that more closely reflect socio-cultural group identity than surface text features do.

The central question of whether political affiliation is enough of a social factor to drive sociolinguistic variation on Twitter could thus not be answered in the affirmative. If it were a strong factor, we would expect to find more variation between the tweets of Trump and Clinton sup-porters. Whether or not the differences between the US-based tweets and the Russian tweets may be said to be due to political affiliation on a larger scale or on other extralinguistic factors remains up for debate. More analysis will have to be done on these and other similar features. Given that Nate Silver's platform, FiveThirtyEight, has, more recently, released an even bigger dataset of Russian tweets, the data is available to do so.

References

Caruso, Assunta/ Folino, Antonietta/ Parisi, Francesca/ Trunfio, Roberto. 2014. A statistical method for minimum corpus size determination. In: Née, Emilie (ed.), *JADT 2014: Proceedings of the Twelfth International Conference on Textual Data Statistical Analysis, June 3-6, 2014*, 135–146. http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/11-JADT2014.pdf [accessed 2019-12-17]

Conover, Michael D./ Ratkiewicz, Jacob/ Francisco, Matthew R./ Gonçalves, B./ Menczer Filippo/ Flammini, Alessandro. 2011. Political polarization on twitter. In: Adamic, Lada A./ Baeza-Yates, R./ Counts, Scott (eds.), *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011.* Palo Alto, CA: The AAAI Press.

Golbeck, Jennifer/ Robles, Christina/ Edmondson, Michon/ Turner, Karen. 2011. Predicting personality from twitter. In: Proceedings of *2011 IEEE International Conference on Privacy, Security and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, 149–156. Los Alamitos, CA, The Institute of Electrical and Electronics Engineers, Inc.

Hu, Yuheng/ Talamadupula, Kartik/ Kambhampati, Subbarao. 2013. Dude, srsly? the surprisingly formal nature of twitter's language. In *Proceedings of the Seventh International AAAI Con-ference on Weblogs and Social Media, Cambridge, Massachusetts, USA, 8 - 11 July 2013*. Palo Alto, CA: AAAI Press.

Kearney, Michael W. 2017. rtweet: Collecting twitter data. https://cran.r-project.org/package=rtweet. [accessed 2019-12-17]

Labov, William. 2010. *Principles of linguistic change, cognitive and cultural factors*, vol. 3. Malden, MA: John Wiley and Sons.

Nguyen, Michelle. 2017. Twitter's role in politics.
URL: https://northwesternbusinessreview.org/twitters-role-in-politics-b3ed620465c9.

Popken, Ben. 2018. Database of 200,000 russian troll tweets.
URL: https://www.nbcnews.com/ tech/social-media/n844731.

Rambukanna. Nathan. 2015. From #racefail to #ferguson: The digital intimacies of race-activist hashtag publics. *The Fibreculture Journal*, 26. 159-188. http://fibreculturejournal.org/wp-content/pdfs/FCJ-194Rambukanna.pdf [accessed 2019-12-17]

Schnoebelen, Tyler. 2012. Do you smile with your nose: Stylistic variation in twitter emoticons. *University of Pennsylvania Working Papers in Linguistics*, 18(2):14.

Tatman, Rachael/ Paulada, Amandalynne. 2017. Social identity and punctuation variation in the #bluelivesmatter and #blacklivesmatter twitter communities.
URL: http://www. rctatman.com/files/Tatman_2017_SocialIdentityAndPunctuation.pdf.

Wickham, Hadley. 2019. stringr: Simple, consistent wrappers for common string operations.
URL: https://cran.r-project.org/package=stringr.

Zappavigna, Michele. 2015. Searchable talk: the linguistic functions of hashtags. *Social Semiotics*, 25 (3):274–291. ISSN 1035-0330. doi: 10.1080/10350330.2014.996948.

Diminutives in Southern Hemisphere Englishes[1]

Alexandra Chudar (Minsk State Linguistic University)
aleksandra.chudar@gmail.com

Abstract

This article deals with the differences in diminutive formation and usage in Australian (AusE), New Zealand (NZE), and South African (SAE) Englishes (Southern Hemisphere Englishes; SHEs). Diminutives in -ie characteristic for all three varieties have been analyzed. The study shows that South African English contains a lot of diminutives borrowed from Afrikaans and is therefore different from Australian and New Zealand varieties where almost all diminutives are formed from English bases. Other differences in the composition of diminutives in each SHEs variety reflect the distinguishing features of local flora and fauna, as well as phenomena characteristic of the local social environment.

1 Introduction

English diminutives are a productive language phenomenon. They can potentially be derived from many words; however, this potential is limited by some restrictions. The formation and functioning of diminutive items is determined by both structural characteristics of the language, which imply some restrictions on the ways of diminutive formation, and the context of diminutive usage. This makes diminutives useful material for studying language variation, as well as the factors that cause this variation.

According to Schneider (2003), diminutives are words that express smallness and/or some kind of attitude. In this work I deal with diminutives in SHEs – Australian, New Zealand and South African Englishes that are considered a separate group of English varieties due to some common linguistic traits, as well as the geographical position of the countries, and the way English spread there (Watts/Trudgill 2002, Schneider 2011, etc.).

Of all English varieties, New Zealand and Australian Englishes are thought to be most prone to diminutive formation, which is confirmed by a great number of works on the topic (Wierzbicka 1984; Sussex 2004; Bardsley/Simpson 2009; Kidd/Kemp/Quinn 2011, etc.). The extent of diminutive "richness" of South African English, and the specific features of its diminutives are still to be analyzed. Therefore the main aim of this paper is to find the diminutives characteristic of each of the three varieties of SHEs, as well as the diminutives that are not shared by the varieties, and to find some explanation for the identified similarities and differences.

2 Data

The source of research material for this study is the Corpus of Global Web-Based English (GloWbE)[2]. I have chosen GloWbE, because it provides the possibility of studying regional variation – it includes texts from 20 varieties of English and allows analysis and comparison of data in all varieties.

This corpus is one of the largest (1.9 bln words) web-based corpora; it allows a wide range of studies on morphological, syntactic and semantic variation among different varieties of

[1] This study was funded by a grant from Belarusian Republican Foundation for Fundamental Research (G18M-062).

[2] https://corpus.byu.edu/glowbe/

English. As for the composition/content of the corpus, about 60 % of texts are taken from informal blogs, the other 40 % come from other web-based materials (newspapers, company web-sites, etc.; Davies/Fuchs 2015). Moreover, it is a web-based corpus, which means that the language of texts included in the corpus represents informal communication and presumably gives an opportunity for a better search for diminutives.

There is one more reason to use corpora for studying diminutives: a lot of words in English can form the base for a diminutive, which is why it is quite difficult to make a comprehensive list of diminutives. In this case, a corpus-driven study (used in this work) can be applied, as it may allow a productive search for diminutives with different frequencies characteristic of a particular variety of English.

3 What is a diminutive?

Diminutives can generally be described as lexical units that denote smallness and/or express some kind of attitude (Schneider 2003: 4). Thus, there are two main criteria that can be used to categorize a lexical unit as a diminutive. The first is a formal one; it suggests that to be defined as a diminutive, a lexical unit should be a derivative formed with the help of a diminutive marker (most commonly, a suffix; Crystal 1991: 104; McArthur 1992: 314). The majority of scholars though point to the fact that the meaning of the derivative should be taken into account first (Jurafsky 1996; Schneider 2003; Carter/McCarthy 2006; Tagabileva 2016, etc.).

The second criterion is therefore a semantic one. Káňa (2018) singles out two main diminutive meanings – smallness and evaluation (e.g. endearment or irony). He states that "real" diminutives have a (potentially) evaluative meaning. If an evaluative meaning is absent, the lexical unit is treated as a "quasi" diminutive (this is the case with the majority of technical terms in Czech he considers in his work – e.g. *lékárnička* [pharmacy^{DIM}] 'first-aid kit' or *pusinka* [mouth^{DIM}, kiss^{DIM} 'meringue']). Other linguists (Heltberg 1964; Strang 1968) support the idea that the "realness" of the diminutive is in its ability to express evaluation.

Some scholars, on the contrary, state that "real" diminutives are those only in which "the idea of smallness is still present" (Hasselrot/Politzer 1959: 334), at least to some extent. Torrend (1891) and Petersen (1910) support this view. Heltberg (1964: 95-96) also calls them "pure" diminutives, while Siatkowska (1967) gives them the name "proper" diminutives.

For this article, I will define diminutiveness in a broad sense and therefore include into the research material both diminutives denoting smallness and those that express some kind of attitude, e.g. *cutties* 'small cuts', *thickie* 'a person lacking in intelligence'.

4 Southern Hemisphere Englishes: history and linguistic peculiarities

In this work, I consider diminutives in inner circle varieties, according to the model of concentric circles introduced by Kachru (1992). The inner circle includes such countries as the United States, Canada, Great Britain, Australia and New Zealand, where English is the first (native) language of a large proportion of the population. Following Bolton (2009), Melchers and Shaw (2013), I also include South Africa into the inner circle, though this is a debatable issue that requires further consideration (Bauer 2002, Cichocka 2006).

The formation of SHEs started later than the formation of the Northern Hemisphere varieties (e.g. Irish, Scottish, American and Canadian Englishes) as part of the second wave of British

colonization that began at the end of the 18th century. Besides the time of formation and the geographical position, SHEs share a lot of structural similarities which are most probably due to the similarity in British settlers moving to the new territories. As Trudgill (2000: 302) notes:

> "These similarities, we maintain, are due to the fact that South African and Australian English are also the result of dialect mixture and that, furthermore, they arose from similar mixtures of similar dialects in roughly similar proportions occurring at a similar period in history."

Speaking of the mixture of similar dialects, all three SHEs varieties developed, first of all, due to British migrants (primarily from south and south-east of Britain; largely working-class people, including convicts), as well as those coming from Scotland and Ireland. The majority of them were people from lower or middle social class, in some cases they came from rural areas. Diagram 1 (based on the data from Trudgill 2000) represents the proportion of settlers coming from different anglophone areas in 1800–1876 to the territories of SHEs.

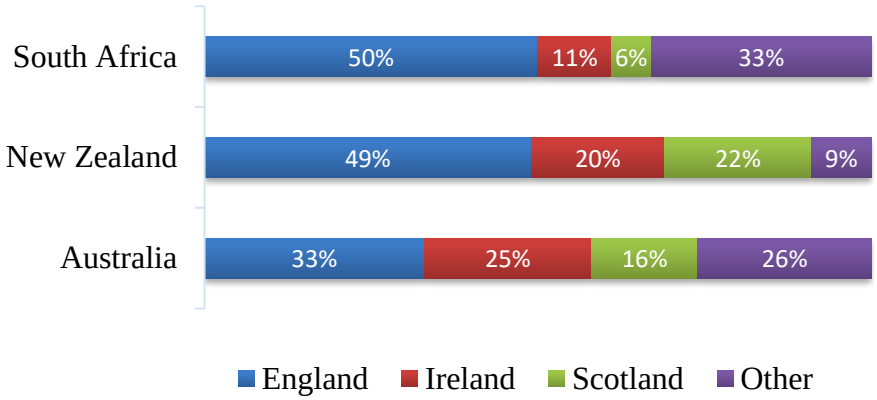


Diagram 1: The proportion of settlers coming from different Anglophone areas in 1800–1876

As for the common linguistic traits of the SHEs varieties, very often scholars mention their phonetic similarity (non-rhotic; short front vowel shift; Trudgill 2004; Watson, Harrington/Evans 2008; Millar 2012, etc.). There are almost no distinctive features of SHE grammar, as compared to British English; speaking of the rare differences, Brandford (1994) notes that it is difficult to say whether they are unique of the variety or just characteristic of spoken English in general.

Along with phonetics and phonology, lexis is another often discussed part of the language system, if one speaks of SHEs. Concepts that denote the local flora and fauna, peculiarities of the local topography, as well as customs and specific features of the local people's life – all of this had to be reflected in the lexis of settlers. As a result, words to express these features were borrowed from the indigenous languages into the developing varieties (Leitner/Sieloff 1998; Macalister 2006). In some cases, lexical units from British English acquired new meanings in the SHEs varieties, while some English words no longer used in their source varieties continued their functioning in SHEs (Trudgill 2004: 84).

Several linguists claim that one more distinctive feature of the SHEs is an extensive usage of diminutives (Bauer 2002; Trudgill 2004). Of all three varieties, Australian English is most prone to diminutive formation. Sussex (2004), for example, notes that this variety contains about 5000 diminutive items. There are several possible reasons for this.

Egalitarianism is often considered one of the main causes of the high frequency of diminutives in Australian English. Researchers (Fiske/Hodge/Turner 1987; Mcleay 1997, etc.) note that equality and anti-authoritarianism, which are part of the egalitarian society, are the authentic values of the Australian culture. As a result, informality and mateship are often emphasized, which is reflected in the local language varieties as well. One more reason for the abundance of diminutives is the ludic character (playfulness) of Australian English, which is due to the cultural value of not taking anything too seriously (Thornhill 1992: 133). To provide some evidence to this fact, scholars most often turn to the Australian vocabulary, as, following Collins and Blair (2001: 3), lexicon is "arguably, the most transparent reflection of speakers' attitudes, values and self-perception". Among such lexical means characteristic of Australian English linguists enumerate the tradition of creating nicknames, rhyming slang, etc., including diminutives. Apart from decreasing the formality of communication and making it non-serious, diminutives can serve as a group marker; in case of Australian English, they mark in-group membership of Australian society (Kidd et al. 2016).

Influence of other languages and their varieties is also often mentioned as predetermining such abundance of diminutives in Australian and other Englishes of the Southern Hemisphere. In this respect, scientists first of all refer to Cockney, the dialect of English that is traditionally spoken by working-class people in London. In the case of Australian English, the Yorta Yorta (Bangerang) language spoken by indigenous people could also have some impact on the development of the diminutives, because, as Curr (1887: 569) notes, diminutives were very common in this language. Regarding South African English, Donaldson (1993: 87) notes that a lot of diminutives in SAE are loanwords from Afrikaans.

5 Results

The main aim of this study was to single out the diminutives characteristic of each SHEs variety. In what follows I will describe the application of a corpus-driven approach that addresses this objective. Section 5.1. will discuss diminutives that can be found in several of the SHEs varieties, their semantic groups and frequencies of the most common ones. Section 5.2 will deal with variety-specific diminutives.

5.1 Cross-varietal comparison

Variation studies are one of the core possibilities provided by the web interface for GloWbE. One of its key features is that the interface allows the researcher to compare the frequency of items in instances per million (ipm) across the different varieties and thus to highlight those that are specific to a certain variety in comparison to the rest. As a result, the interface provides a list of most characteristic words for a particular variety. Moreover, there are other settings that can be modified according to the research task. For example, it is possible to set a minimum frequency for words included in the search results.

In my study, I contrasted the pairs of GloWbE subcorpora – a particular inner circle country (e.g. Australia) vs. other inner circle countries (excluding Australia). I carried out several queries with different minimum frequency (starting with 1) to find out which frequency provides the largest number of diminutives including both high-frequent ones as well as those that are not very widespread to get a more comprehensive picture of diminutive functioning in the varieties under discussion.

As *-ie* suffix is thought to be most commonly used suffix for diminutive formation, and because its use is not restricted geographically (Bauer 2002), this paper will deal with *-ie* only.

For each variety, a list of 1000 words with *-ies* ending was extracted from the corpus. Then, with the help of online dictionaries (Oxford, Cambridge, Collins online dictionaries), all non-diminutives (such as full-form words, misspellings) were manually removed from the lists. After that, the lists of diminutives typical of each variety were made. The number of diminutives and the subcorpora size for each SHEs variety are shown in Table 1.

Table 1: Subcorpora size and the number of diminutives found for each SHEs variety[3].

	Australian English	New Zealand English	South African English
Number of different types of diminutives	149	176	169
Number of tokens of diminutives (normalized frequency of instances per million words)	8 455 (57.05 ipm)	3 480 (42.76 ipm)	1 997 (44.02 ipm)
Size of subcorpus (words)	148 208 169	81 390 476	45 364 498

The table demonstrates that though speaking about diminutives a lot of scholars consider primarily Australian English, other Southern Hemisphere varieties also support this "pro-diminuitve" trend, as the number of diminutives is quite comparable for each variety, despite different subcorpora sizes.

As for the similarity of diminutives from SHEs, 81 items from my sample are similar for all three varieties or can be found in two of them. Table 2 presents these diminutives and lists them according to the varieties compared. For easier understanding, the diminutives are also organized according to their semantic properties. In case of polysemous diminutives (marked with *), several meanings were found in the corpus, e.g. *grandies* for 'grandchildren' or 'grandparents'; *catties* for 'catfish' (AusE) or 'catapult' (SAE); only the most common one was included in the table.

[3] The first row of the table reflects the diversity of diminutives in SHEs; the second one shows the diminutive richness of texts in the varieties under consideration.

Table 2: Diminutives shared by SHEs varieties

	All 3 varieties (6)	Items shared by Australian English and New Zealand English (52)	Items shared by Australian English and South African English (13)	Items shared by New Zealand English and South African English (10)
People	Aussies, touchies	cockies*, Dutchies, grandies*, greenies, groundies*, kindies*, leaguies, littlies, loopies, Ozzies, polies*, pollies, premmies, rellies, sparkies, subbies, surfies, truckies, tradies, towies*, warmies, wharfies	livies*, newies*, pommies, tweeties*, twinnies*, windies	Ausies, loosies, noddies*, troopies
Animals	Flatties	fluffies*, freshies*, mossies, pippies, salties, sheepies*, spotties*, tunnies, youngies	catties*	
Food	Brekkies	bickies*, bikkies*, cuppies, lollies, middies*, mushies, stubbies*, tinnies*	nibblies	bubblies*, slushies
Clothes	Woolies	boardies, jarmies, trackies, undies	sunnies*, woollies	beanies
Other	Bommies	bombies, gollies, leafies, leggies, ouchies, pokies*, possies*, wedgies, whoopsies	headies, steelies. tanties	bivvies, blinkies*, quickies

The table shows that in terms of diminutives Australian and New Zealand Englishes are much closer to each other than to South African English, though there are some diminutives characteristic of all three varieties as well. New Zealand and South African Englishes share the smallest number of diminutives. This closeness of Australian and New Zealand varieties can be attributed first of all to the geographical proximity of the two countries and to the additional influence of Dutch via Afrikaans on the vocabulary of South African English, which led to more unique diminutives in this variety (see also section 5.2).

Corpora provide an opportunity not only to find diminutives characteristic of a variety, but also to identify the ones most specific (i.e. more frequent in comparison) to each of the three varieties. The most frequent diminutives found in more than one variety are included in Table 3.

Table 3: The most frequent diminutives characteristic of several varieties.

Diminutive	Frequency in varieties (in ipm)
Aussies ‘australians’	15.70 (AusE); 5.87 (NZE); 7.72 (SAE)
woolies ‘woolen clothes’	2.17 (AusE); 0.18 (NZE); 2.65 (SAE)
pollies ‘politicians’	4.67 (AusE); 1.35 (NZE)
pokies ‘poker machine’ or ‘place with poker nachines’	3.20 (AusE); 2.16 (NZE)
lollies ‘lollipops’	2.97 (AusE); 2.48 (NZE)
undies ‘underwear’	2.58 (AusE); 1.23 (NZE)
greenies ‘supporters of a green party’	2.02 (AusE); 1.27 (NZE)
tradies ‘tradesmen’	1.41 AusE); 0.28 (NZE)

5.2 Zooming in on variety-specific items

In this section, I will turn to those diminutives that are present in one variety only[4].

Language contacts and influence of other (first of all, local) languages is one of the main reasons that predetermine existence and usage of variety-specific diminutives (cf. section 4). In my sample it is primarily relevant for South African English – while almost all diminutives in Australian and New Zealand Englishes are derived from English bases, some diminutives in South African English are loanwords from Afrikaans:

Bakkies – bak – Eng. ‘container’

Bergies – berg – Eng. ‘mountain’

Boeries – boerewors – Eng. ‘sausage’

Bossies – bos – Eng. ‘bush’

Broekies – broek – Eng. ‘trousers’

Dingetjies – ding – Eng. ‘thing’

Dorpies – dorp – Eng. ‘town/village’

Eendjies –eend – Eng. ‘duck’

Koppies – kop – Eng. ‘head’

Maties – maat – Eng. ‘mate’

Potjies – pot – Eng. ‘pot’

Vygies – vyg – Eng. ‘fig’

Waterblommetjies – waterblom – Eng. ‘waterflower’

I should note however the difficulty to state what exactly was borrowed from Afrikaans – the words already put into the diminutive form or just the words in their initial form that later became bases for diminutive formation in English. This is a result of English and Afrikaans being related languages with quite similar system of diminutive formation (Donaldson 1993).

Language contact is, however, not the only explanation for the differences in diminutives. The peculiarities of local environment, first of all, the names of the representatives of local flora and fauna, are also reflected in the sets of diminutives. Examples include:

[4] Differences in the sets of diminutives for the discussed varieties include i.a. differences in spelling, eg. *bombies* (AusE, 0.05) / *bommies* (NZE, 0.07) ‘bomboras’; *Aussies* (all SHEs, 24.1) / *Ausies* (SAE, 0.73) ‘Australians’; *sammies* (NZE, 0.06) / *sarmies* (SAE, 0.26) ‘sandwiches’ which, however, will not be discussed in this paper.

Humpies ‘humpback salmon’ (AusE)

Jewies ‘jewfish’ (AusE)

Queenies ‘queenfish’ (AusE)

Blowies ‘blowfish’ (AusE)

Ellies ‘elephants’ (SAE)

Apies ‘apes’ (SAE)

Raggies ‘ragged tooth sharks’ (SAE)

Another group includes names of local teams and organizations:

Vinnies ‘St Vincent de Paul Society’ (charity organization in Australia)

Swannies ‘Sydney Swans’ (sports team, Australia)

Scarfies ‘scarves’ (students of Otago University, New Zealand)

One more group consists of diminutives naming people characteristic of the location the variety is spoken in. These diminutives can be based on either some peculiarities of appearance or occupation:

Darkies ‘a person with brown or black skin’ (SAE)

Whities ‘a white person’ (SAE)

Beachies ’a person who spends much time at the beach’ (AusE)

Boaties ‘boatsmen’ (NZE)

The total number of non-shared and shared diminutives for each SHEs variety is presented in the following table.

Table 4: The number of non-shared vs. shared diminutives for SHEs

	Australian English	New Zealand English	South African English
Shared	71 (47.7%)	68 (38.6%)	29 (17.2%)
Non-shared	78 (52.3%)	108 (61.4%)	140 (82.8%)
Total	149	176	169

According to the table above, the largest number of non-shared diminutives is found in South African English, which is in line with the previously discussed data suggesting more closeness of Australian and New Zealand Englishes to each other than to South African variety.

6 Discussion

Despite the fact that Australian, New Zealand and South African Englishes are united under one name of Southern Hemisphere Englishes and very often discussed together, the degree of their similarity varies significantly. Of all SHEs, Australian and New Zealand Englishes have a lot of similar features, which is described in studies on phonology (Bauer 1999; Schneider 2011), grammar (Smith 2009; Elsness 2009), and lexis (Bardsley/Simpson 2009; Hay 2008). South African English, though sharing a lot of characteristics with two other SHEs varieties, still possesses of several traits at each level of language that make this variety distinctive to some extent.

The findings revealed in this study of diminutives are in line with previous research of the three varieties. First of all, this research confirms that diminutives are characteristic of all SHEs and that this is relevant for NZE or SAE no less than for AusE, as despite the different subcorpora sizes, the numbers of diminutives found for each SHE variety are quite comparable.

Secondly, speaking of the degree of similarity between the varieties, this study also supports the fact that there is more closeness between AusE and NZE than between these varieties and SAE. The majority of shared diminutives are shared exactly between AusE and NZE; there are as well diminutives common for AusE and SAE or NZE and SAE, and for all three varieties, but these numbers are much lower.

As for the non-shared diminutives, one can speak of several semantic domains these diminutives prevail in. These include words denoting the representatives of the local flora and fauna, elements of the local topography, items naming the distinctive features of the local social system. Some of these diminutives are borrowings from the local languages; some of them are diminutives derived from common English words, but present in one variety only. The reason why the non-shared diminutives prevail in these domains may be that the objects denoted by the words from these domains probably form the basis of the distinctiveness of the location in question.

In general, speaking of SHEs, linguists note that of all language levels, phonology of these three varieties has received much greater attention than the other ones. Studies of diminutives can thus contribute to the research into the peculiarities of SHEs lexical systems and into the lexical variation in the English language. On the one side, diminutives are a highly productive lexicon-grammar phenomenon, and their usage is determined by the structural characteristics of the language. On the other side, these linguistic items convey a wide array of pragmatic effects, and their functioning depends on some extralinguistic phenomena. As a result of such multifariousness, diminutives can be a good material for investigations into the linguistic and extralinguistic phenomena that predetermine variation on different levels of the language system.

7 Conclusion

This paper set out to explore similarities and differences in the diminutives found in each variety of the SHEs and to give some interpretation to the obtained results. The corpus-driven analysis has shown that of all three varieties, the highest number of diminutives is shared by Australian and New Zealand Englishes. Diminutives in South African English are, however, somewhat different. The main reason for it is the influence of Dutch-Afrikaans – while the bases of Australian and New Zealand English diminutives are mostly English, some diminutives in South African English are loanwords from Afrikaans.

Differences in the sets of diminutives for SHEs cannot be attributed only to language contact. In each variety, there are also variety-specific diminutives denoting features of the local environment, as well as the diminutives naming local teams and organizations or typical representatives of the local population.

At the same time, this study is subject to several limitations that could be addressed in future research. Firstly, this is a corpus-driven study that focused on the comparison of the sets of SHEs diminutives derived from the corpus; a future corpus-based study of diminutives (with the lists of diminutives retrieved from the dictionaries) could contribute to the sets of the non-shared items found in the three varieties. Secondly, a deeper analysis of the diminutive semantics may provide a more complex overview of the meanings included into the category. And thirdly, the studies of diminutive pragmatics are completely left outside the scope of this paper but could well be one of the future aspects of researching diminutives in SHEs providing more insights into the cross-varietal variation in SHEs.

References

Bardsley, Dianne/ Simpson, Jane. 2009. Hypocoristics in New Zealand and Australian English. In: Peters, Pam/ Collins, Peter (eds.), *Comparative Studies in Australian and New Zealand English: Grammar and beyond*, 49-69. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Bauer, Laurie. 1999. On the Origins of the New Zealand English Accent. *English World-Wide: A Journal of Varieties of English* 20(2). 287-307.

Bauer, Laurie. 2002. *An introduction to international varieties of English*. Edinburgh: Edinburgh University Press.

Bolton, Kingsley. 2009. Varieties of World Englishes. In: Kachru, Braj/ Kachru, Y./ Nelson, Cecil L. (eds.), *The Handbook of World Englishes*, 289-312. UK: Blackwell Publishing.

Branford, William. 1994. English in South Africa. In: Burchfield, Robert (ed.), *The Cambridge History of the English Language*, 430-496. Cambridge: Cambridge University Press.

Cambridge Dictionary. https://dictionary.cambridge.org/ [accessed 2018-04-15].

Carter, Ronald/ McCarthy, Michael. 2006. *Cambridge Grammar of English: a Comprehensive Guide*. Cambridge: Cambridge University Press.

Cichocka, Aleksandra. 2006. Global Expansion of English: The South African Case. *Werkwinkel: Journal of Low Countries and South African Studies* 1(1). 253-267.

Collins Dictionary. https://www.collinsdictionary.com/ [2018-04-20].

Collins, Peter/ Blair, David. 2001. Language and identity in Australia. In: Collins, Peter/ Blair, David (eds.), *English in Australia*, 1-13. Amsterdam/Philadelphia: John Benjamins Publishing.

Crystal, David. 1991. *A dictionary of linguistics and phonetics. 3rd edition*. Cambridge, MA: Basil Blackwell.

Curr, Edward M. 1887. *The Australian race: its origin, languages, customs, place of landing in Australia and the routes by which it spread itself over that continent*. Vol.3. Melbourne: John Ferres, government printer/London: Trübner and Co.

Davies, Mark/ Fuchs, Robert. 2015. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* 36(1). 1-28.

Donaldson, Bruce C. 1993. *A Grammar of Afrikaans*. Berlin/New York: Mouton de Gruyter.

Elsness, Johan. 2009. The perfect and the preterite in Australian and New Zealand English. In: Peters, Pam/ Collins, Peter/ Smith, Adam (eds.), *Comparative Studies in Australian and New Zealand English: Grammar and beyond*, 89-114. Amsterdam/Philadelphia: John Benjamins Publishing.

Fiske, John. 1983. Surfalism and Sandiotics: The Beach in Oz Culture. *Australian Journal of Cultural Studies* 1(2). 120-149.

Fiske, John/ Hodge, Bob/ Turner, Graeme. 1987. *Myths of Oz: Reading Australian Popular Culture*. Sydney: Allen & Unwin.

Gregova, Renáta. 2013. Evaluative Morphology and (Mor)phonological Changes in Diminutives and Augmentatives of a Sample of Indo-European, Niger-Congo and Austronesian Languages. *International Journal of Linguistics, Literature and Translation* 1(1). 1-14.

Hay, Jennifer. 2008. *New Zealand English*. Edinburgh: Edinburgh University Press.

Heltberg, Kristine. 1964. O deminutywach i augmentatywach. *Prace Filologiczne* 18(2). 93-102.

Jurafsky, Daniel. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72(3). 533-578.

Káňa, Tomáš. 2018. Terminology in and around Diminutives. https://slavicorp.ff.cuni.cz/wp-content/uploads/sites/144/2018/09/SlaviCorp2018_Book_of_Abstracts.pdf [accessed 2018-12-30].

Kachru, Braj B. 1992. World Englishes: approaches, issues and resources. *Language Teaching* 25(1). 1-14.

Kidd, Evan/ Kemp, Nenagh/ Kashima Emiko S./ Quinn, Sara. 2016. Language, Culture, and Group Membership: an Investigation Into the Social Effects of Colloquial Australian English. *Journal of Cross-Cultural Psychology* 47(5). 713-733.

Kidd, Evan/ Kemp, Nenagh/ Quinn, Sara. 2011. Did you have a choccie bickie this arvo? A quantitave look at Australian hypocoristics. *Language Sciences* 33(3). 359-368.

Leitner, Gerhard/ Sieloff, Inke. 1998. Aboriginal words and concepts in Australian English. *World Englishes* 17(2). 153-169.

Macalister, John. 2006. The Maori presence in the New Zealand English lexicon, 1850–2000. Evidence from a corpus-based study. *English World-Wide* 27(1). 1-24.

McArthur, Tom. 1992. *Oxford Companion to the English Language*. Oxford: Oxford University Press.

Mcleay, Colin. 1997. Inventing Australia: a Critique of Recent Cultural Policy Rhetoric. *Australian Geographical Studies* 35(1). 40-46.

Melchers, Gunnel/ Shaw, Philip. 2013. *World Englishes*. London and New York: Routledge.

Millar, Robert McColl. 2012. *English Historical Sociolinguistics*. Edinburgh: Edinburgh University Press.

Oxford Dictionaries. https://en.oxforddictionaries.com/ [accessed 2018-04.13].

Petersen, Walter. 1910. *Greek diminutives in -ion: A study in semantics*. Weimar, Germany: R. Wagner Sohn.

Politzer, Robert L. 1959. Review on Etudes sur la formation diminutive dans les langues romanes by Bengt Hasselrot. *Language* 35(2). 332-335.

Schneider, Edgar. 2011. *English around the World: An Introduction*. Cambridge: Cambridge University Press.

Schneider, Klaus. 2003. *Diminutives in English*. Tübingen: Max Niemeyer Verlag.

Siatkowska, Ewa. 1967. Deminutywa rzeczowników we współczesnych literackich językach zachodniosłowiańskich. *Studia z Filologii Polskiej i Słowiańskiej* 6(1). 157-170.

Smith, Adam. 2009. Light verbs in Australian, New Zealand and British English. In: Peters, Pam/ Collins, Peter/ Smith, Adam (eds.), *Comparative Studies in Australian and New Zealand English: Grammar and beyond*, 139-155. Amsterdam/Philadelphia: John Benjamins Publishing.

Strang, Barbara. M. H. 1968. *Modern English Structure*. London: Edward Arnold.

Sussex, Roland. 2004. Abstand, Ausbau, Creativity and Ludicity in Australian English. *Australian Journal of Linguistics* 24(1). 3-19.

Tagabileva, Mariya. *Sopostavitelny analiz modeley diminutivnogo slovoobrazovanya (na materiale sushchestvitelnyh zhenskogo roda v juzhnoslavyanskih yazykah)*. Moskva: Moskovsky gosudarstvenny universitet imeni Lomonosova.

Torrend, J. 1891. *A comparative grammar of the South-African Bantu languages*. London: Kegan Paul, Trench, Trübner & Co.

Trudgill, Peter et al. 2000. Determinism in New-Dialect Formation and the Genesis of New Zealand English. *Journal of Linguistics* 36(2). 299-318.

Trudgill, Peter. 2004. *New-dialect Formation: The Inevitability of Colonial Englishes*. New York: Oxford University Press.

Watson, Catherine I./ Harrington, Jonathan/ Evans, Zoe. 1998. An acoustic comparison between New Zealand and Australian English vowels. *Australian journal of linguistics* 18(2). 185-207.

Watts, Richard J./ Trudgill, Peter. 2002. *Alternative Histories of English*. London and New York: Routledge.

Wierzbicka, Anna. 1984. Diminutives and depreciatives: Semantic representation for derivational categories. *Quaderni di semantica* 5(1). 123-130.

Clausal Verb Complementation in Varieties of English

Laetitia Van Driessche (University of Zurich) and Hubert Cuyckens (University of Leuven)[1]
laetitia.vandriessche@es.uzh.ch and hubert.cuyckens@kuleuven.be

Abstract

This article looks at the finite/non-finite complementation alternation with expect *and* suggest *(in its suasive meaning) from a diachronic and a synchronic perspective. It investigates whether the diachronic shift from finite to non-finite complementation in British English is reflected by distributional differences in finite vs. non-finite complementation patterns in World Englishes positioned at different stages in Schneider's Dynamic Model (2007). It also examines the factors that determine the complementation alternation in these varieties. Data have been extracted from the CLMET, BNC and GloWbE corpora. Methodologically, frequency analyses, random forest analyses, logistic regression analyses and conditional inference trees are employed. The analyses show that* expect *largely corroborates the hypotheses, whereas* suggest *shows unexpected tendencies.*

1 Introduction

This article deals with the variation of finite vs. non-finite complement clauses (CCs) with the complement-taking predicates (CTPs) *expect* and *suggest*. In the present study, finite complementation involves *that*-clauses (1a, 2a), whereas non-finite complementation comprises gerundial (2b) and infinitival clauses (1b, 2c).

(1) a. *She had half expected that he would deflect the question.* (BNC)
b. *I […] cannot expect the bookseller to take them again.* (CLMET)

(2) a. *Security experts suggest that you steer clear of these bogus pages.* (GloWbE BrE)
b. *[H]e suggested bringing his mum.* (GloWbE BrE)
c. *I would suggest to bring Android apps out first.* (GloWbE BrE)

The study will be restricted to CCs occurring *after* the CTP. Fronted CCs will not be considered here as it has been observed that processing works differently in those contexts (see Hawkins 1990).

Two research questions are addressed in this article. The first is diachronic as well as synchronic in nature. Diachronically, it has been observed that an increasing share of finite clauses has been replaced by non-finite clauses in British English (see Rohdenburg 1995; Denison 1998; Los 2005; Cuyckens/D'Hoedt/Szmrecsanyi 2014). Considering that finite CCs can be viewed as more isomorphic-iconic than non-finite CCs because they tend to display a one-to-one correspondence between meaning and form (they overtly express subjects, tense, aspect and modality (cf. Noonan 2007: 59)), clausal complementation can be said to have become less iconic, or more economic, over time. Synchronically, it has been argued that iconicity facilitates second-language acquisition, and by extension the development of second language (L2)

[1] We would like to thank Benedikt Szmrecsanyi and the two anonymous reviewers for their valuable feedback on this paper.

varieties in Schneider's (2007) Dynamic Model (Steger/Schneider 2012: 158–159, 164). Accordingly, it can be hypothesised that varieties of English at a less advanced stage in the Dynamic Model are likely to contain more iconic structures, i.e. finite CCs, than varieties at a more advanced stage (Steger/Schneider 2012: 164). Combining the diachronic and synchronic perspectives, this paper will investigate whether the diachronic shift from finite to non-finite complementation is reflected by varieties positioned at increasingly more advanced stages in the Dynamic Model. For this purpose, a frequency study will be carried out, comparing distributions of finite vs. non-finite complements across diachronic stages of British English with those of the varieties under investigation.

The second goal of this paper is to identify the factors that significantly determine complementation alternation in World Englishes. Special attention will be given to the strength of the factor 'second language variety' in predicting finite vs. non-finite clausal complementation in that we will try to ascertain whether there is indeed an L2 learning effect across the World Englishes and whether this learning effect is related to the level of advancedness in the Dynamic Model.[2] This will be examined with random forest analyses, logistic regression analyses and conditional inference trees. It will become clear that there are important differences between *expect* and *suggest*: whereas *expect* seems to conform to the hypotheses, *suggest* tells quite a different story.

2 Data and Methodology

The article will focus on two CTPs: *expect* and *suggest* in its suasive meaning.[3] The patterns for *expect* consist of infinitival CCs (non-finite) and *that*-clauses (finite) in the indicative (often with *will/would*) or the subjunctive. For *suggest*, these consist of infinitival and gerundial CCs (non-finite) and *that*-clauses (finite) in the indicative, the subjunctive or with mandative *should*. Modals different from suasive (or mandative) *should* were not included as it could be argued that these add a nuance that cannot be expressed by the non-finite complements.[4] Infinitival clauses following *suggest* are the proscribed variant (cf. Carter et al. 2016, "Suggest"). However, looking into possible influence from prescriptivism in the different varieties under study lies beyond the scope of this paper.

Zero-*that*-clauses were not included for either verb as they pose problems for recall: unlike for the other clause types, they do not contain an easily retrievable element (e.g. the *that*-complementiser).[5]

The patterns were extracted from the CLMET3.0 (Late Modern English (LModE)) and the BNC (Present-day British English (PDE)) to trace their relative distribution over time, and from GloWbE (World Englishes) to examine their frequency distribution across L2 varieties. The BNC data were restricted to a sample from the fiction subsection to maintain genre conformity with the CLMET.

[2] Investing the link between World and Learner Englishes with regard to this second language acquisition effect falls outside the scope of this paper. A useful study on clausal complementation in Learner Englishes is Tizón-Couto (2014).

[3] In its non-suasive meaning, there is no alternation with the gerund.

[4] For instance, in (i) the speaker moderates the suasive nature of the suggestion by using the modal *could* instead of *should*.

(i) *I have suggested over the years that more co-operatives and friendly societies* could *be set up for employment.* (GloWbE JamE)

The non-finite correlate of (i) would lose the moderating nuance.

[5] For a description of the regular expressions used to extract the clauses, see Van Driessche (2018).

The five varieties selected to represent the different phases in Schneider's Dynamic Model are British (reference variety), Canadian (phase 5), Jamaican (phase 4), Malaysian (phase 3) and Bangladeshi English (phase 2).[6] There is no phase 1 variety as this consists of the settlement of a new colony (Schneider 2007: 33-36), which means that it cannot be investigated for PDE.

To examine which factors determine the variation of finite vs. non-finite CCs in World Englishes, data extracted from GloWbE were coded for a number of factors. Coded data were then entered into random forest and logistic regression analyses and into conditional inference trees. Random forest analyses look at *factor importance*: they measure the factors' "relative importance" in the choice between two alternatives (Tagliamonte/Baayen 2012: 159–160, 172). Logistic regression analyses look at *factor strength*: the effect that the presence of a factor has on the response variable (in this case finite vs. non-finite complementation) (Tagliamonte/Baayen 2012: 150; Speelman 2014: 501–502). Conditional inference trees show how factors interact with each other and can therefore be used to find probabilistic grammar differences between varieties (see Szmrecsanyi et al. 2016).

In the following paragraphs (2.1-2.4), we discuss the factors used in the statistical analyses. The labels used for the factors are provided between brackets.

2.1 Factors Relating to the Cognitive Complexity Principle

As we already mentioned, the choice between finite and non-finite complementation also entails a choice between more or less explicit grammatical options. In this regard, it has been claimed by Rohdenburg (1996: 151) that "the more explicit [option] will tend to be favored in cognitively more complex environments"; this is Rohdenburg's Complexity Principle. His principle is here operationalised by the following six factors.

A first type of a cognitively complex environment (Rohdenburg 1996: 160–161) involves the **Intervening material** between the CTP and the complementiser **(Intervening_material)**, measured in terms of the number of words between the CTP and the *that*-complementiser (one word in (3a)), the CTP and the gerund (two words in (3b)), the CTP and the infinitive marker *to* (three words in (3c)), or between the infinitive marker *to* and the infinitive itself (one word in (3d)). Not considered intervening material are raised objects preceding *to*-infinitives with the CTP *expect* (such as *it* in (4a)), because for their finite counterparts the object would be part of the *that*-clause (i.e. would not precede the *that*-complementiser) (4b).

(3) a. […] *I suggested* then *that the Government should approach a programme of publicizing* […] (GloWbE JamE)

b. *Bryant suggested* to Watson *writing a private letter* […] (GloWbE BrE)

c. *I have already suggested* to my son *to go there* […] (GloWbE BrE)

d. *We've suggested to them to* just *make the proposal.* (GloWbE CanE)

(4) a. *Don't expect* it *to be easy, but give it a try* […] (GloWbE BrE)

b. *Don't expect that* it *will be easy, but give it a try* […]

[6] The only phase 2 variety that Schneider (2007) mentions for his Dynamic Model (i.e. Fiji English) is not represented in GloWbE. For this reason, BanE was chosen to represent this phase after a perusal of literature on the status of this variety.

For the same reason, a negator preceding the gerund or the *to*-complementiser (5a), or positioned between the *to*-complementiser and the infinitive (5b), is not considered intervening material either.

(5) a. *So, do you honestly suggest me* not *to refer to the Daily Star to improve my English?* (GloWbE BanE)

b. *I suggest to* not *join* […] (GloWbE BanE)

As this is a continuous variable (as are Clause length and Formality, see below), the regression analysis calculates the likelihood that one of the complementation patterns will be chosen when the number of words between the CTP and the complementiser increases (cf. Levshina 2015: 261). Counting for the continuous variables was done automatically in R.

A second complexity environment is **Clause length** (**Clause_length**) (Rohdenburg 1996: 164), which is measured in terms of the number of words from the CTP until the end of the sentence (seven in (6)).

(6) *[A]sk the social worker's supervisor to review the situation or* suggest that a mediator be brought in. (GloWbE CanE)

A third and fourth type of complex environments are **Negation of the CTP** (**Negation_CTP**) (Rohdenburg 1996: 164) and **Negation in the subordinate clause** (**Negation_sub**) (Rohdenburg 2015). Markers of negation comprise *not* (5a-b, 7a), *n't* (7b), *nor* (7c) *neither* (7d) and *not that* (7e).

(7) a. *I am* not *suggesting that we stop having political debates.* (GloWbE MalE)

b. *I do*n't *suggest doing any serious physical activity* […] (GloWbE JamE)

c. […] *I wasn't disappointed,* nor *was I expecting to be.* (GloWbE JamE)

d. *It* neither *suggests that school gates are locked to prevent students leaving at lunchtime or that pupils are forced to eat specific foods.* (GloWbE BrE)

e. Not that *I really expected to escape so easily.* (GloWbE MalE)

Fifth, **Passive CTPs** (**Passive**) are considered to be cognitively complex relative to their active counterparts (Rohdenburg 1996: 169, 173). Only attestations including *expect* (8a-b) are coded for this factor, since the gerund (a complementation pattern for *suggest*) does not usually combine with a CTP in the passive voice.

(8) a. […] *it* is expected that *hats be worn during prayer.* (GloWbE CanE)

b. *Buddhists* are expected to *know the difference between right and wrong* […] (GloWbE CanE)

A final factor is **Coreferentiality** (**Coreferentiality**), which, like Passive, will only be coded for *expect*. The subjects of the main and the subordinate clause can be coreferential, as in (9a), or refer to two different entities, as in (9b). Non-coreferentiality is cognitively more complex (Steger/Schneider 2012: 169-171).

(9) a. *How* can you expect to have *an honest relationship if you do such stunts* […] (GloWbE BrE)

b. Don't expect it to be *easy, but give it a try* […] (GloWbE BrE)

2.2 Formality: number of contractions (Contractions)

The degree of **Formality** may also have an effect on the variation between the complementation types. Since it has been claimed that more formal registers prefer more explicit coding (cf. Rohdenburg 1996: 160), this would imply more finite complementation in formal contexts.

As a measure of the text's formality, AntConc's window size was set at 1000 characters[7] starting with *expect*/*suggest* (i.e. the right context), after which the number of contractions (following Quirk et al. 1985: 123) was counted in R.

2.3 Person and Tense of the CTP (Person_CTP; Tense_CTP)

Because we wanted to test the influence of deictic distance[8] on the complementation type, we also coded for **Person (Person_CTP) and Tense of the CTP** (**Tense_CTP**). The **Person of the CTP** is coded for the first, second and third person. This factor was added because it proved to be a significant factor in Cuyckens/D'Hoedt's (2015: 89-94) analysis of complementation patterns of the verb *admit*, although with conflicting results. When *that*-clauses and zero-*that*-clauses were contrasted with gerundial, *to*-gerundial and *to*-infinitival clauses, their conditional inference tree showed that the third person predicted non-finite complementation in PDE when *admit* had the sense of *wrongdoing*. However, when *that*-clauses were contrasted with *to*-infinitival clauses and zero-*that*-clauses, the third person favoured *that*-clauses.[9]

As the future tense is not frequent, coding for **Tense of the CTP** is restricted to a 'past' and a 'non-past' level. While 'past' obviously comprises situations occurring in a past time period (including situations encoded by modals with past time reference, as in (10a)), 'non-past' comprises CTPs in the present and future tense as well as modals expressing a hypothetical situation such as *can* in (10b) or *could* in (10c).

[7] There were minimal variations in the window size that AntConc provided.

[8] The third person and the past tense are more deictically removed from the speaker's "zero-point […] of the deictic context" (Lyons 1977: 683).

[9] There are two different combinations of clausal complementation in this article since non-coreferentiality between the subject of the main and the subordinate clause does not allow for all complementation patterns of *admit* (Cuyckens/D'Hoedt 2015: 78, 87).

(10) a. *That was very common in England just 100 years ago. Women* could *not necessarily* expect *to survive childbirth.* (GloWbE MalE)

b. *By the end of 2011, we* can expect *the figure to reach RM450bn*. (GloWbE MalE)

c. *With its mass release in FTV sets, you* could expect *to see its price fall by quite a bit once it is released* […] (GloWbE MalE)

2.4 Variety (Variety)

As the present study probes the effect of the factor 'second language variety' on finite vs. non-finite clausal complementation, each of the attestations will be coded for **Variety** (as mentioned above, these are British (BrE), Canadian (CanE), Jamaican (JamE), Malaysian (MalE) and Bangladeshi English (BanE)).

3 Results

In what follows, the results for *expect* (3.1) and *suggest* (3.2) will be described. These results will then be discussed in section 4.

3.1 *Expect*

The first section (3.1.1) provides the frequency distributions for *expect*; the second section (3.1.2) presents the results of the statistical analyses.

3.1.1 Frequency distributions

Table 1 presents the frequency distributions for finite vs. non-finite complementation in the CLMET (1780-1850) and the BNC (1980-1993). The results from the CLMET are confined to a sample from the second subperiod of the corpus (i.e. the middle subperiod) as there were sufficient attestations for analysis in this subperiod (as opposed to *suggest*, which was less frequent).

Table 1: Frequencies and percentages for *expect* in the CLMET (1780-1850) and BNC (1980-1993); normalised per 100,000 words.

Corpus	Complement type	Raw frequencies	Percentages	Normalised frequencies
CLMET	Finite	54	16.5%	1.8
	Non-finite	274	83.5%	9.4
BNC	Finite	11	2.8%	0.4
	Non-finite	385	97.2%	12.4

Table 1 indicates that there is a significant rise in non-finite complementation between the periods 1780-1850 and 1980-1993 (a chi-square test indicates that p = 1.43263e-10). In fact, the BNC consists almost exclusively of non-finite clauses (97.2%).

Table 2 presents the frequencies for *expect* in a sample of the GloWbE corpus (General section).

Table 2: Frequencies and percentages for *expect* in the GloWbE corpus; normalised per 100,000 words.

Variety	Complement type	Raw frequencies	Percentages	Normalised frequencies
British English	Finite	50	5.7%	0.8
	Non-finite	831	94.3%	13.6
Canadian English	Finite	56	7.2%	1.0
	Non-finite	719	92.8%	12.8
Jamaican English	Finite	53	9.7%	1.3
	Non-finite	495	90.3%	12.5
Malaysian English	Finite	52	5.1%	0.7
	Non-finite	960	94.9%	13.0
Bangladeshi English	Finite	50	11.1%	1.2
	Non-finite	400	88.9%	9.4

Table 2 shows that non-finite complementation is highly frequent in all varieties. The highest percentage occurs in MalE (94.9%), followed by BrE (94.3%), CanE (92.8%), JamE (90.3%) and BanE (88.9%).

Chi-square analyses show that apart from JamE (p = 0.456986305), all varieties differ significantly from BanE. The difference between JamE and BrE (0.004508678) and JamE and MalE (p = 0.000646521) is also significant. The differences between the other varieties are not significant, however.

3.1.2 Statistical analyses

The random forest analysis[10] (Figure 1) shows the relative importance of each factor in explaining the variation between finite and non-finite complementation. Coreferentiality explains the alternation best, followed by the factors Variety, Passive CTP, Person of the CTP and Clause length. Intervening material and Tense of the CTP are not important factors. The index of concordance for this analysis is C= 0.8499125, which means that it has a good fit (Tagliamonte/Baayen 2012: 156).

[10] The default settings were used for the random forest analysis.

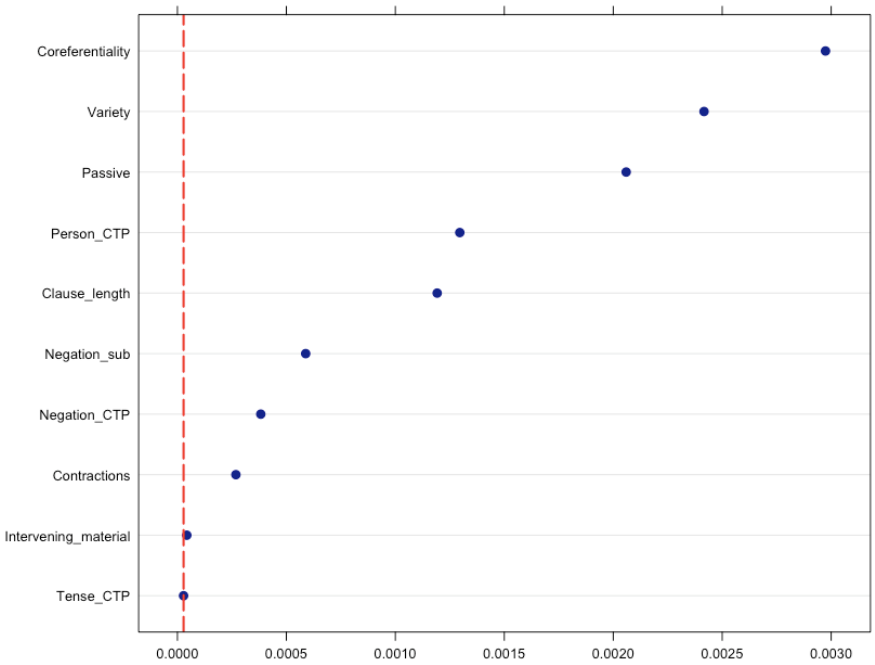


Figure 1: Random forest analysis for *expect*. C = 0.8499125.

The logistic regression analysis in Figure 2 provides information about the strength of each factor in predicting the alternation. The predicted output is for non-finite complementation. A negative value for 'Estimate' means that the factor level (e.g. BanE for the factor Variety) disfavours non-finite complementation; a positive value indicates a favouring effect. The effect is always measured against a reference level, e.g. BrE for Variety. The reference level is not visible in the model (Tagliamonte/Baayen 2012: 148; Speelman 2014: 521-522).

The final two columns show the significance of the (dis)favouring effect. The boundary for significance is p = 0.05 (Speelman 2014: 507-508). The degree of significance is represented more graphically in the final column: if there are no stars in the final column, the effect is not significant.

```
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)          2.883351   0.233417  12.353  < 2e-16 ***
VarietyBanE         -0.823640   0.219640  -3.750 0.000177 ***
VarietyCanE         -0.349547   0.208419  -1.677 0.093516 .
VarietyJamE         -0.538444   0.213778  -2.519 0.011779 *
VarietyMalE          0.134871   0.210490   0.641 0.521684
PassiveYes           0.540264   0.163897   3.296 0.000979 ***
Person_CTP2nd        0.484121   0.261260   1.853 0.063879 .
Person_CTP3rd        0.224733   0.177212   1.268 0.204740
Negation_CTPYes      0.090566   0.221266   0.409 0.682313
Negation_subYes     -2.269328   0.422248  -5.374 7.68e-08 ***
Tense_CTPPast       -0.211283   0.169810  -1.244 0.213414
CoreferentialitySame  2.431745   0.369332   6.584 4.57e-11 ***
Clause_length       -0.037779   0.005667  -6.666 2.62e-11 ***
Intervening_material -0.074183   0.063514  -1.168 0.242817
Contractions         0.145561   0.078453   1.855 0.063541 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 2: Logistic regression analysis for *expect*. The predicted output is for non-finite complementation.

The logistic regression analysis confirms the frequency count: BrE has a slightly stronger tendency to favour non-finite complementation, except in the case of MalE (although this favouring effect is not significant). The analysis also shows that the passive voice, the second person, coreferentiality and informality have a favouring effect on non-finite complementation, whereas negation in the subordinate clause and longer CCs have a disfavouring effect.

The conditional inference tree in Figure 3 chooses the factor that "has the strongest association with the response" (i.e. response variable), after which it "makes a binary split in this variable, dividing the dataset into two subsets" (Levshina 2015: 291). These steps are then repeated "until there are no variables that are associated with the outcome at the pre-defined level of statistical significance" (Levshina 2015: 291). The tree indicates that Clause Length is the most predictive factor (Node 1): CCs of less than or equal to 16 words strongly predict non-finite complementation (Nodes 4-6). In the absence of coreferentiality, longer clauses predict non-finite complementation slightly less than shorter clauses (Nodes 4 and 5). When Clause Length is higher than 16 words, there is a high predictive effect for non-finite complementation when the two subjects are coreferential (Node 19). When there is no coreferentiality, Negation in the Subordinate Clause is a predictive factor (Node 8), with negation predicting finite complementation (Node 18). Variety appears at a low branching level (Node 9), which indicates that it is only a reliable factor when there is no negation in the subordinate clause, when there is no coreferentiality, and when the clause is longer than 16 words. The varieties pattern differently: for BanE, CanE and JamE, Passive CTPs (Node 10), and, in the case of an active CTP, Negation of the CTP (Node 11) are predictive factors. For BrE and MalE, Intervening Material is a predictive factor (Node 15): there is a higher predictive effect for non-finite complementation in contexts with no intervening material (Nodes 16 and 17).

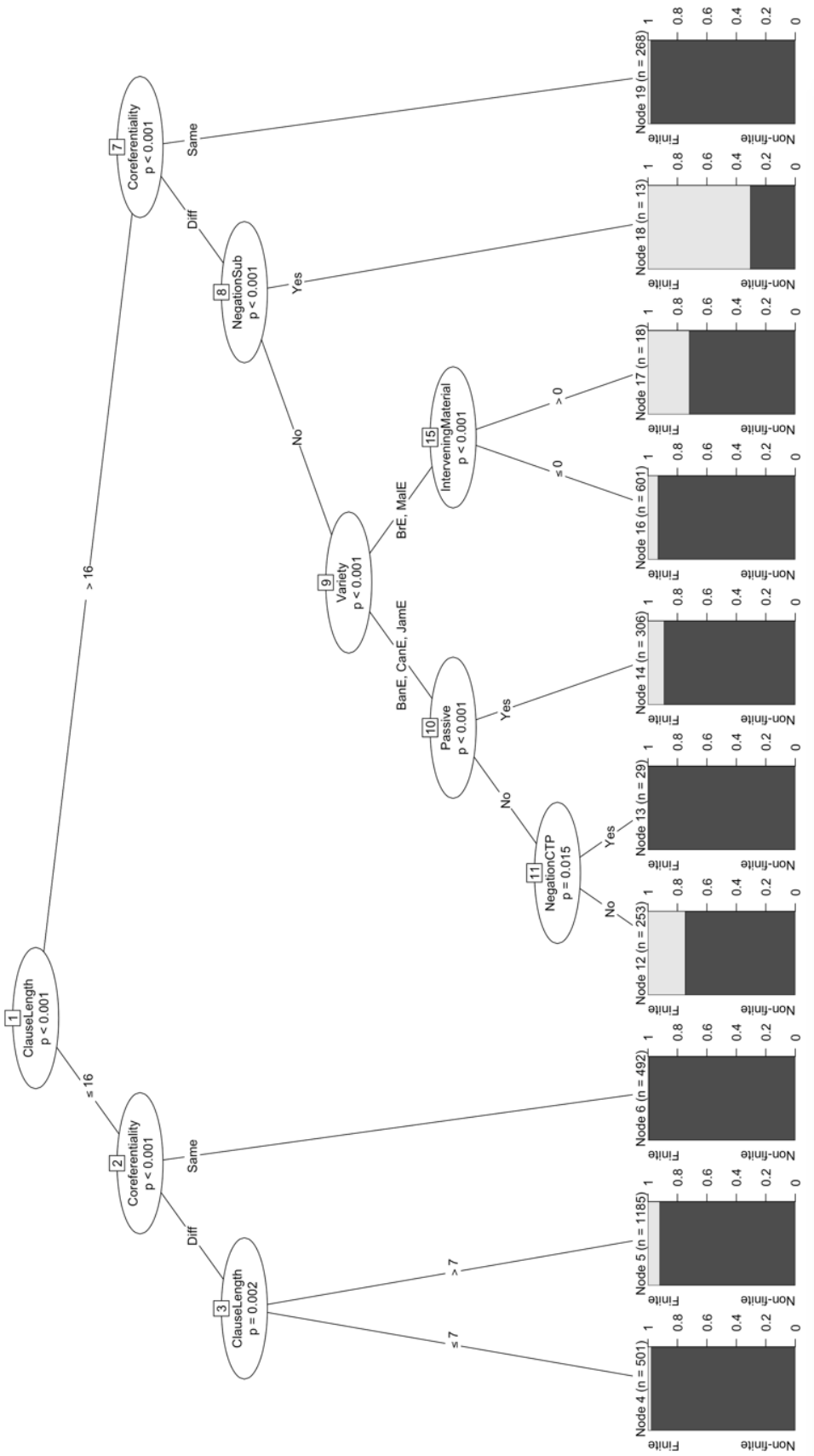


Figure 3. Conditional inference tree for *expect*.

3.2 *Suggest*

The following sections provide the results of the frequency distributions (3.2.1) and the statistical analyses (3.2.2) for *suggest* in its suasive meaning.

3.2.1 Frequency distributions

Table 3 shows the frequency distributions for suasive *suggest* in the CLMET and the BNC. Due to the low frequencies for this verb, all subperiods of the CLMET are included (covering the period 1710-1920).

Table 3. Frequencies and percentages for *suggest* in the CLMET and BNC; Normalised per 1,000,000 words.

Corpus	Complement type	Raw frequencies	Percentages	Normalised frequencies
1710-1780 (CLMET)	Finite	0	0%	0
	Non-finite	0	0%	0
1780-1850 (CLMET)	Finite	16	76.2%	1.4
	Non-finite	5	23.8%	0.4
1850-1920 (CLMET)	Finite	101	78.3%	7.9
	Non-finite	28	21.7%	2.2
1980-1993 (BNC)	Finite	90	75%	10.5
	Non-finite	30	25%	3.5

Table 3 shows that there is a slight increase in non-finite complementation when the periods 1780-1850 and 1980-1993 and the periods 1850-1920 and 1980-1993 are compared. However, the difference between these periods is not significant ($p = 0.907252532$ and $p = 0.538855245$ respectively). In fact, none of the periods differ significantly from each other.[11]

The frequencies for *suggest* in a sample of the GloWbE corpus (General section) are displayed in Table 4.

[11] The p-value for the periods 1780-1850 and 1850-1920 is $p = 0.829099237$.

Table 4. Frequencies and percentages for *suggest* in the GloWbE corpus; Normalised per 1,000,000 words.

Variety	Complement type	Raw frequencies	Percentages	Normalised frequencies
British English	Finite	203	62.7%	13.6
	Non-finite	121	37.3%	8.1
Canadian English	Finite	244	54.3%	14.5
	Non-finite	205	45.7%	12.2
Jamaican English	Finite	233	80.1%	18.0
	Non-finite	58	19.9%	4.5
Malaysian English	Finite	170	50%	12.8
	Non-finite	170	50%	12.8
Bangladeshi English	Finite	163	48.4%	12.1
	Non-finite	174	51.6%	12.9

The results in Table 4 show that MalE and BanE have the highest share of non-finite clauses, with an even distribution for MalE and more non-finite than finite complementation for BanE.[12] The other three varieties have a higher proportion of finite complementation. Chi-square tests indicate that the differences between CanE and MalE ($p = 0.226398397$), between CanE and BanE ($p = 0.097095452$) and between BanE and MalE are not significant ($p = 0.671056858$). In all other cases, the varieties differ significantly from each other.

3.2.2 Statistical analyses

The random forest analysis for *suggest* (Figure 4) shows that Intervening Material is the most important factor in the choice between finite and non-finite complementation, followed by Variety, Clause length and Formality. Tense of the CTP, Person of the CTP and Negation are less important factors in the alternation. The index of concordance is 0.7796408, which is just below the threshold of good performance (i.e. $C \geq 0.8$, Tagliamonte/Baayen 2012: 156).

[12] For the distributions of gerundial and infinitival clauses complementing *suggest* in the different varieties, see the Appendix (Table a).

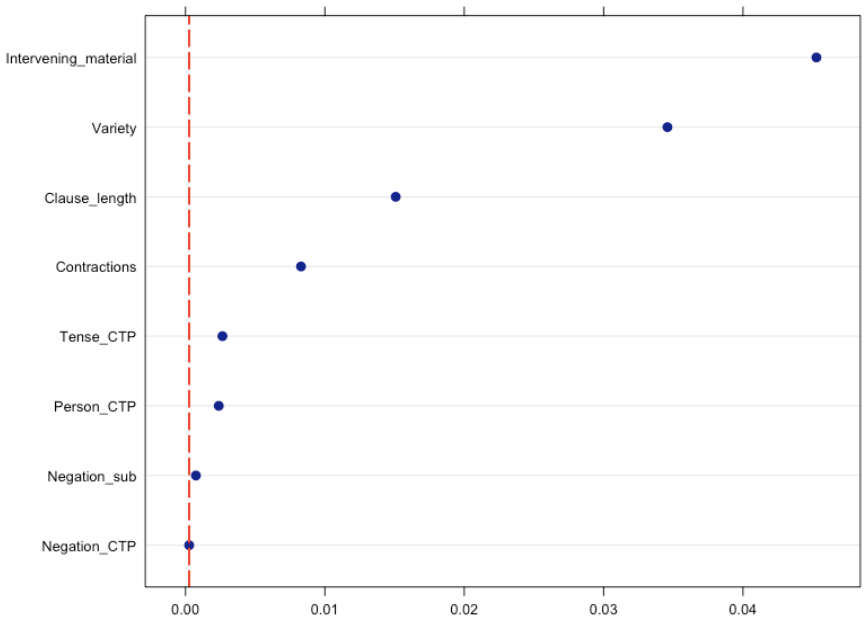


Figure 4: Random forest analysis for *suggest*. C = 0.7796408.

Figure 5 provides the logistic regression analysis for *suggest*. All levels are significant.

```
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)          0.184284   0.175610   1.049 0.293998
VarietyBanE          0.719062   0.167569   4.291 1.78e-05 ***
VarietyCanE          0.356119   0.155280   2.293 0.021825 *
VarietyJamE         -0.863874   0.194124  -4.450 8.58e-06 ***
VarietyMalE          0.409504   0.165986   2.467 0.013621 *
Person_CTP2nd       -1.147320   0.313950  -3.654 0.000258 ***
Person_CTP3rd       -0.386626   0.123534  -3.130 0.001750 **
Negation_CTPYes     -0.838021   0.349285  -2.399 0.016429 *
Negation_subYes     -0.639835   0.281584  -2.272 0.023070 *
Tense_CTPPast       -0.275851   0.129966  -2.122 0.033796 *
Clause_length       -0.029435   0.005258  -5.598 2.17e-08 ***
Intervening_material  0.161906   0.038474   4.208 2.57e-05 ***
Contractions          0.228863   0.052389   4.369 1.25e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 5: Logistic regression analysis for *suggest*.

The regression analysis indicates that JamE is the only variety that disfavours non-finite complementation more than BrE. The first person, more intervening material and more informal texts favour non-finite complementation. Negation, the past tense and longer CCs disfavour this complementation type.

The conditional inference tree (Figure 6) indicates that Variety is the most important predictive variable (Node 1) and that the varieties pattern differently. For JamE, there is relatively little internal variability in the alternation, as Clause Length is the only factor mentioned on the right side of the tree: longer CCs (more than 12 words) predict finite clauses more than shorter clauses in this variety (Nodes 16 and 17). For the other varieties, the tree mentions a wider range of predictive factors. If Clause Length is less than or equal to 16 words, Person of the CTP is a reliable factor (Node 3), with the first person predicting non-finite complementation more than the second and third person (Nodes 4 and 5). If Clause Length is higher than 16 words,

Formality is a predictive factor (Node 6). In the presence of contractions, negation in the subordinate clause has a higher predictive effect for finite complementation than subordinate clauses without negation (Nodes 13 and 14). In more formal texts (no contractions), Intervening material is a predictive factor (Node 7). In the context of intervening material, Variety comes into play again (Node 9): only BanE and MalE have a predictive effect for non-finite complementation when there is intervening material, whereas BrE and CanE predict finites in this context. This indicates that the favouring effect of the logistic regression analysis (Figure 5) for non-finites in the context of intervening material is restricted to BanE and MalE.

4 Discussion

Our hypothesis that there would be a diachronic rise in non-finite complementation in BrE has only been confirmed for *expect*. This means that *expect* has indeed become less iconic-isomorphic and more economic over time. In GloWbE, non-finite complementation is by far the preferred variant for all varieties. At the same time, the varieties (except for MalE) can be positioned along a cline of increasingly higher shares of non-finite complementation in accordance with their position in the Dynamic Model: BanE (phase 2) has the lowest share of non-finite complementation (88.9%), whereas CanE (phase 5) has the highest share (92.8%).[13] The CTP *expect* thus shows a second language acquisition effect in that the less advanced varieties have higher proportions of the more iconic variant (which is easier to acquire).

The CTP *suggest*, however, does not display a significant rise in non-finite complementation over time. Our hypothesis that an increasing share of non-finite complements in the various Englishes under study corresponds with an increasingly higher phase in Schneider's Dynamic Model is not confirmed either: the least advanced varieties – MalE and BanE – have the highest proportion of the non-iconic complementation pattern. The low frequency for non-finite complementation in JamE (19.9%) is also striking.

[13] Note that BrE, the reference variety, has a slightly higher percentage (94.3%). However, this is not significant ($p = 0.198377442$).

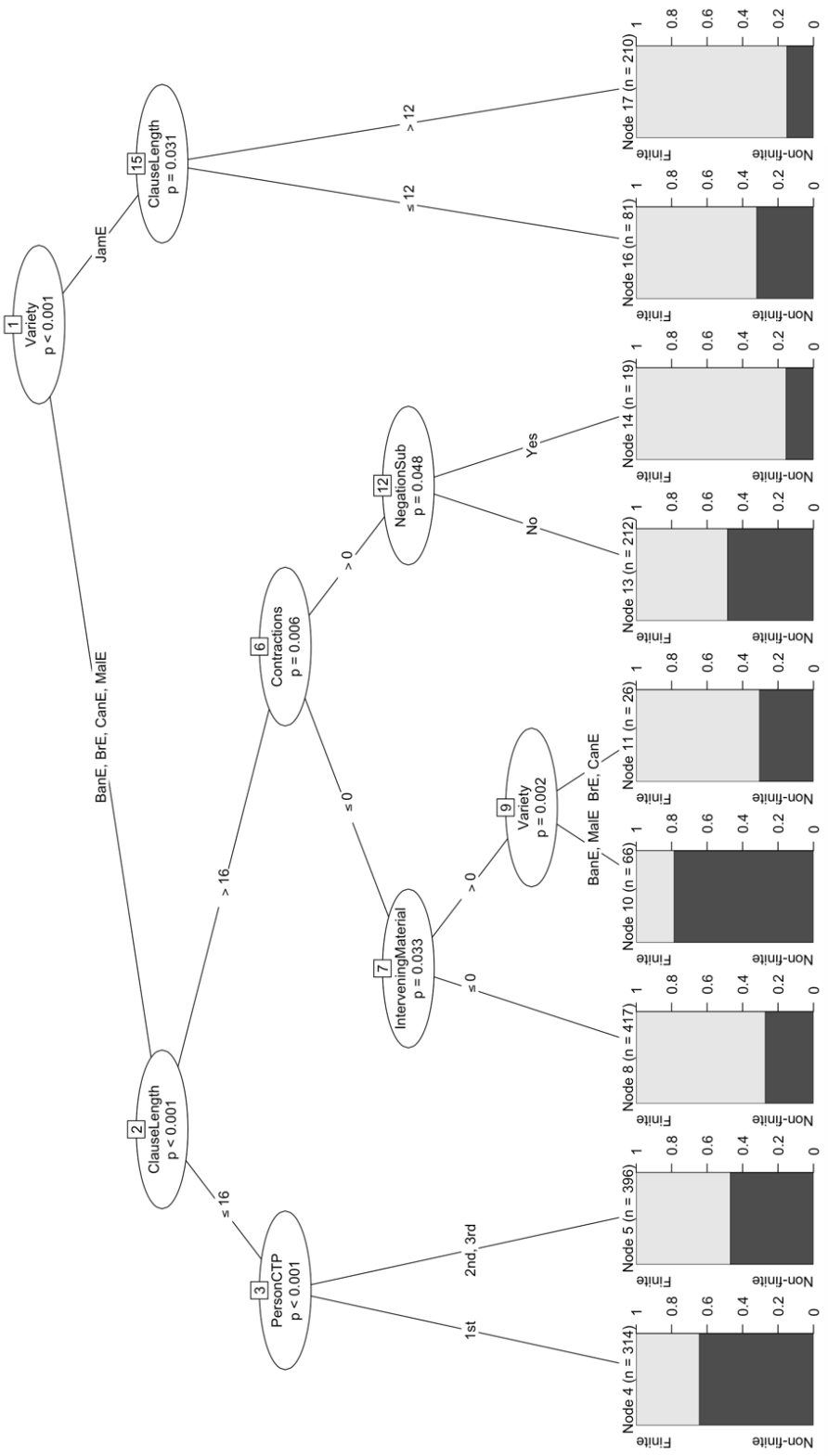


Figure 6: Conditional inference tree for *suggest*.

A possible explanation of these unexpected results for *suggest* and *expect* is substrate influence. In this respect, Patrick (2004: 423) states that Jamaican creole does not use "gerund forms with -*in(g)*". This could explain why the gerund is so infrequent for *suggest* in the phase 4 variety. The high number of gerundial and infinitival clauses in BanE for *suggest* could be explained by the rich non-finite system in Bangla, which includes both an infinitive and a "*gerund-participle*" with "both nominal and verbal properties" (David/Conners/Chacón 2015: 229, 235). Finally, the high number of infinitival clauses in MalE for *expect* could be a transfer from Malay, which "does not have a finite/non-finite distinction. The base form of the Malay verb, which is used extensively especially in spoken/informal language, is […] unmarked for tense" (Svalberg/Chuchu 1998: 33). However, more research is necessary to rule out other possible factors.

The statistical analyses in this article have indicated which factors play a role in the alternation between finite and non-finite alternation, which effect they have and how they interact. The random forest analyses give an overview of the most important factors in the decision between finite and non-finite complementation: for *expect*, Coreferentiality, Variety and Passive CTPs are the most important factors; for *suggest*, these are Intervening Material, Variety and Clause Length.

An overview of the hypotheses and results of the logistic regression analyses is given in Table 5.

Table 5: Hypotheses and results from the logistic regression analyses.

Factor	Hypothesis: preference for which complementation type?	Confirmed for *expect*?	Confirmed for *suggest*?
(More) intervening material	Finite	Not significant	No
Longer CCs	Finite	Yes	Yes
Negation CTP	Finite	Not significant	Yes
Negation subordinate clause	Finite	Yes	Yes
Passive CTPs	Finite	No	
Coreferentiality	Non-finite	Yes	
Formality	Finite	Yes (borderline significance)	Yes
Person CTP: 3rd	Finite	Not significant	Yes
Tense CTP: past	Finite	Not significant	Yes
Variety: BrE	Non-finite	Yes Exception: MalE	No

Table 5 indicates that the only factors related to the Complexity Principle that do not predict the hypothesized results are Intervening material (for *suggest*) and Passive CTPs (for *expect*): they favour non-finite complementation even though they constitute cognitively complex environments. However, two of the factors that measure cognitive complexity are not significant for *expect*. For *suggest*, the third person and the past tense significantly favour finite complementation.

However, the conditional inference trees indicate that there are differences between the varieties and between the two verbs concerning the predictive patterns. An important example is the factor Intervening Material: this only predicts non-finite complementation for the varieties MalE and BanE in the case of *suggest*. For CanE and BrE, more intervening material predicts finite complementation. This means that the favouring effect in the regression analysis is mostly influenced by the MalE and BanE data. For *expect*, however, the tree indicates that in MalE, intervening material predicts finite complementation slightly more than non-finite complementation. The conditional inference tree for *expect* also indicates that Variety is only a reliable factor for longer complement clauses without negation whose subject is not coreferential with the main clause subject.

5 Conclusion

In this article we have looked at the alternation between finite and non-finite complementation from a diachronic and a synchronic perspective. More specifically, we have investigated whether the observed rise of non-finite complementation in the history of BrE is reflected by distributional differences across World Englishes, whereby higher shares of non-finite CCs would be attested in varieties at a higher stage in Schneider's Dynamic Model. Underlying this research question is the hypothesis that just like clausal complementation may become less transparent over time, World Englishes will increasingly use less transparent complementation patterns as they move along the Dynamic Model. This is linked to a learner effect: less transparent complementation patterns are more difficult to acquire.

Our hypothesis is only borne out for *expect*, which shows a significant rise in non-finite complementation. Complementation patterns with *expect* thus become less iconic and more economic over time. It also shows a second language acquisition effect since, with the exception of MalE, the frequency of non-finite clauses in the varieties under study correlates with their level of advancedness in the Dynamic Model. For *suggest*, however, there is neither a significant rise of non-finite complementation over time nor a second language acquisition effect.

The random forest analyses indicate that for *suggest*, the most influential factors are Intervening material, Variety and Clause length; for *expect*, these are Coreferentiality, Variety and Passive CTPs. For both verbs, Variety thus appears to play a significant role.

The logistic regression analysis confirms most of our hypotheses. Non-coreferentiality, the third person, the past tense, formal texts, negation and longer CCs all favour finite complementation. Contrary to our hypotheses, however, the passive voice (*expect*) and more intervening material (*suggest*) favour non-finite complementation despite their status as cognitively complex environments. However, the conditional inference tree for *suggest* indicates that the latter favouring effect is variety-specific: only MalE and BanE favour non-finite complementation when there is intervening material following the CTP *suggest*.

Future research should examine our unexpected results in more detail, taking into account possible substrate influence and influence from prescriptivism, or could also include zero-*that*-clauses.

Corpora

CLMET 3.0. A corpus of Late Modern English Texts. See https://perswww.kuleuven.be/~u0044428/clmet3_0.htm

BNC. British National Corpus. See http://www.natcorp.ox.ac.uk

GloWbE. Corpus of Global Web-Based English. See https://corpus.byu.edu/glowbe/.

References

Carter, Ronald et al. 2016. Suggest. *Cambridge Dictionary Online (English Grammar Today. An A-Z of Spoken and Written Grammar)*. Cambridge: Cambridge University Press. https://dictionary.cambridge.org/grammar/british-grammar/common-verbs/suggest [accessed 2018-6-5].

Cuyckens, Hubert/ D'Hoedt Frauke. 2015. Variability in Clausal Verb Complementation: the Case of *Admit*. In: Höglund, Mikko et al. (eds.), *Perspectives on Complementation: Structure, Variation and Boundaries*, 77–100. London: Palgrave Macmillan.

Cuyckens, Hubert/ D'hoedt, Frauke/ Szmrecsanyi, Benedikt. 2014. Variability in verb complementation in Late Modern English: finite vs. non-finite patterns. In: Hundt, Marianne (ed.), *Late Modern English Syntax*, 182–203. Cambridge: Cambridge University Press.

David, Anne B./ Conners, Thomas J./ Chaćon, Dustin A. (eds.). 2015. *Descriptive grammar of Bangla* (Mouton-CASL Grammar Series 2). Berlin/Boston: De Gruyter Mouton.

Denison, David. 1998. Syntax. In: Romaine, Suzanne (ed.), *The Cambridge History of the English Language. 1776-1997*, Vol. IV, 92-329. Cambridge: Cambridge University Press.

Hawkins, John A. 1990. A Parsing Theory of Word Order Universals. *Linguistic Inquiry* 21(2). 223-261.

Levshina, Natalia. 2015. *How to do Linguistics with R: Data exploration and statistical analysis.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Los, Bettelou. 2005. *The Rise of the* To-*Infinitive.* Oxford: Oxford University Press.

Lyons, John. 1977. *Semantics*. Vol. 2. Cambridge: Cambridge University Press.

Noonan, Michael. 2007 [1985]. Complementation. In: Shopen, Timothy (ed.), *Language Typology and Syntactic Description. Volume II: Complex Constructions*, Vol. 2, 2nd edn., 52-150. Cambridge: Cambridge University Press.

Patrick, Peter L. 2004. Jamaican Creole: morphology and syntax. In: Kortmann, Bernd/ Schneider, Edgar/ Burridge, Kate/ Mesthrie, Rajend/ Upton, Clive (eds.), *A Handbook of Varieties of English. Morphology & Syntax: A Multimedia Reference Tool*, Vol. 2, 407-438. Berlin: Mouton de Gruyter.

Quirk, Randolph et al. 1985. *A Comprehensive Grammar of the English Language*. London/New York: Longman.

Rohdenburg, Günter. 1995. On the Replacement of Finite Complement Clauses by Infinitives in English. *English Studies: A Journal of English Language and Literature* 76(4). 367-388.

Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7(2). 149–182.

Rohdenburg, Günter. 2015. The Embedded Negation Constraint and the Choice between More or Less Explicit Clausal Structures in English. In: Höglund, Mikko et al. (eds.), *Perspectives on Complementation: Structure, Variation and Boundaries*, 101–127. London: Palgrave Macmillan.

Schneider, Edgar. 2007. *Postcolonial English: Varieties around the World.* Cambridge: Cambridge University Press.

Speelman, Dirk. 2014. Logistic regression: A confirmatory technique for comparisons in corpus linguistics. In: Glynn, Dylan/ Robinson, Justyna A. (eds.), *Corpus Methods for Semantics: Quantative studies in polysemy and synonymy*, 487–533. Amsterdam: John Benjamins Publishing Company.

Steger, Maria/ Schneider, Edgar. 2012. Complexity as a function of iconicity: The case of complement clause constructions in New Englishes. In: Kortmann, Bernd/ Szmrecsanyi, Benedikt (eds.), *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*, 156-191. Berlin/Boston: De Gruyter.

Svalberg, Agneta M-L./ Chuchu, Hjh Fatimah B. H. A. 1998. Are English and Malay worlds apart? Typological distance and the learning of tense and aspect concepts. *International Journal of Applied Linguistics* 8(1). 27–60.

Szmrecsanyi, Benedikt et al. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37(2). 109-137.

Tagliamonte, Sali A./ Baayen, R. Harald. 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.

Tizón-Couto, Beatriz. 2014. *Clausal Complements in Native and Learner Spoken English: A Corpus-Based Study with LINDSEI and VICOLSE*. Bern: Peter Lang AG.

Van Driessche, Laetitia. 2018. *Clausal Verb Complementation in Varieties of English: A Corpus-Driven Approach.* Leuven: KU Leuven MA thesis.

Appendix

Table a: Distributions for the clause types complementing *suggest* for the varieties in GloWbE. Normalised per 1,000,000 words.

Variety	Clause type	Raw frequencies	Normalised frequencies
British English	That-clause	203	13.6
	Gerundial clause	103	6.9
	Infinitival clause	18	1.2
Canadian English	That-clause	244	14.5
	Gerundial clause	184	10.9
	Infinitival clause	21	1.2
Jamaican English	That-clause	233	18.0
	Gerundial clause	43	3.3
	Infinitival clause	15	1.2
Malaysian English	That-clause	170	12.8
	Gerundial clause	72	5.4
	Infinitival clause	98	7.4
Bangladeshi English	That-clause	163	12.1
	Gerundial clause	95	7.0
	Infinitival clause	79	5.8

Derhoticisation in Scotland - fine-grained variation and phonemic stability

Monika Pukli (University of Strasbourg)
mpukli@unistra.fr

Abstract

Sociophonetic patterning of rhotics in Scotland has been at the centre of linguistic research in the past decades and a gradual loss or weakening of rhoticity appears to be linked to age, social class, gender and identity (in different combinations). There is undoubtedly a wide range of different phonetic realisations of /r/ in general, and of coda-r in particular, available to Scottish speakers. However, before we can envisage current phonetic variation as part of a more global phonological change resulting in a non-rhotic system, it is essential to make an explicit link between variation in speech and the stability of the phonological system. Results presented in this paper suggest that non-rhotic speakers are not necessarily young, and it is argued that considering the interplay of various social factors today, phonological change is not underway.

Keywords: *complex systems, Scottish English, lenition, rhoticity*

1 Introduction

Is structured variation a prequel or a sequel to language change? Is change inevitable? Variation observed at any given moment of time reflects the usage of elements that are potentially undergoing change. The weakening of consonantal coda /r/ in Scottish English today is believed to represent gradual change that will eventually lead to a phonemic re-categorization of the system without coda-r. English accents are typically divided into rhotic vs. non-rhotic varieties according to the phonotactic constraints of /r/: rhotic accents, such as General American, pronounce <r> in every position (*rate, cart, car, car park, car insurance*), while non-rhotic accents, such as RP, only pronounce <r> when it is followed by a vowel sound (*rate, car insurance*). The distribution of the phoneme also has an indirect impact on the vowel inventory. The weakening of consonantal coda /r/ in Scottish English refers to the non-prevocalic context being pronounced without /r/ as in *cart, car, car park*.[1]

The original /r/ phoneme used by the different speech communities in Scotland in the past is usually described as an alveolar trill. Accounts from the beginning of the twentieth century (Williams 1909, Grant 1914) identify three major variants: trills, taps and approximants. The explicit or implicit prescriptivist ideology that often accompanies linguistic description well into the fifties and sixties, however, makes it difficult to appraise any real patterning of use in the population of that time. Trills are certainly weakened to a one-stage tap realisation; today, all studies agree on the sporadic presence of trills in different regions in Scotland, but already in 1938 only three in ten students surveyed used them and approximants were recommended in 'polite speech' (McAllister 1938). More recent descriptions, however, show that not only central approximants had been used in Scotland for over a century, but that they were adopted more typically than other phonetic realisations more than 40 years ago, at least in some varieties of Scottish English (Romaine 1978 and Speitel/ Johnston 1983 for Edinburgh, Macafee 1983 for Glasgow). It turns out that while an approximant is phonetically closer to vocoids, it is this variant that resists further lenition the most in contemporary varieties. It is also worth mentioning that while the focus on sociophonetic factors is prevalent in present-day reports, earlier Wells (1982) suggested a *phonological* distribution between alveolar taps occurring in

[1] The author wishes to thank the two anonymous reviewers for their valuable feedback and insightful comments on the paper.

intervocalic position (as in *sorry*, *agree*) and post-alveolar or retroflex fricatives and approximants occurring word-finally (as in *word*, *care*), with no preference for either group of forms in initial position. While evidence of non-rhotic and derhoticised forms of coda-r is abundant today (Dickson/ Hall-Lew 2017, Jauriberry 2016, Stuart-Smith/ Lawson/ Scobbie 2014, Watt/ Llamas/ Johnson 2014, to mention only the most recent reports from different parts of Scotland), it is not certain that Scottish English will become non-rhotic.

The outstanding potential of any phonological system to resist change has been underlined for a long time. Variationist and sociophonetic research usually endeavours to unveil patterns of usage in an attempt to highlight the actuation and diffusion of innovating forms, but as Labov (1994) says "the absence of change may be even more difficult to account for than its presence. The long-term stability of many components of the linguistic system is even more striking than the rapid transformation of others" (1994:42). This long-standing question of relative stability was already discussed extensively decades earlier in Weinreich/ Labov/ Herzog (1968) also drawing attention to "the opposite problem – of explaining why language fails to change" (1968:112).

Current approaches to language in complex systems theory[2] (Larsen-Freeman/ Cameron 2008, Beckner et al. 2009, Kretzschmar 2015, Mufwene/ Coupé/ Pellegrino 2017, for example) can tackle this problem with new insights and methods, but most importantly complexity theory offers a theoretical model of speech in which variation is the key to stability. Constant, repeated everyday interactions among language users in a speech community need a stable communication system that never breaks down and the robustness of the system is achieved by its architecture (e.g. Ferrer i Cancho/ Solé 2001, Corominas-Murtra/ Valverde/ Solé 2009, Arbesman/ Strogatz/ Vitevitch 2010, Steels et al. 2010, Pierrehumbert/ Stonedahl/ Daland 2014, Goldstein/ Vitevitch 2014, Burridge 2017, Jiang/ Yu/ Liu 2019, for example) and this architecture emerges from large-scale variation.

The aim of this paper is first to provide new findings of weakened coda-r realisations from Ayrshire speakers, and then to further discuss questions relating to socio-phonetic patterning and the supposed weakening of rhoticity in Scotland in general.

2 Derhoticisation in Scotland

Over the last twenty years or so, more and more findings confirmed what seems to be a growing derhoticization of coda-r[3] in Scottish English. Only a very brief overview of the fine-grained variation in phonetic realisations across different social groups and geographical areas is provided here and the reader is referred to Stuart-Smith/ Lawson/ Scobbie (2014) and Dickson/ Hall-Lew (2017) for a comprehensive and more detailed account.

Scottish English is traditionally described as a rhotic accent of English and while actual realisations of /r/ changed considerably over time, rhoticity in the standard variety of Scottish English remains unaffected. Weakening rhotic instances of coda-r in Glasgow were reported apparently more than hundred years ago (cf. Watt/ Llamas/ Johnson's account tracing back derhoticised variants to 1901 and 1913 (2014:85) and Stuart-Smith et al. (2015:4)). Derhoticisation in Edinburgh has also been reported (starting with Romaine's study on children

[2] Complex systems in linguistics are most often called complex adaptive systems (CAS); this paper uses the more general term that the reader will find in different fields from mathematics, chemistry and physics to biology, economics, computer science, and social sciences.

[3] The term *non-prevocalic* /r/ is preferred in many papers on the subject because it unequivocally identifies the segmental environment of the rhotic under study, i.e. one that is not followed by a vowel sound, thus excluding positions of linking-r that, for obvious reasons, are less likely to undergo weakening. This paper refers to this same environment by the shorter although less exact term *coda-r*.

(1978) and followed by Speitel/ Johnston's survey of 91 adults (1983)), but accounts become scarce as we move away from the Central Belt of Scotland.[4]

Researchers typically combine articulatory, auditory and acoustic measurements in order to distinguish coarticulatory effects from socially relevant allophones that are independent from the phonological context (Lawson/ Scobbie/ Stuart-Smith 2011, Watt/ Llamas/ Johnson 2014, for example). Ultrasound tongue imaging uncovers the articulatory mechanisms, and acoustic analyses reveal the interaction between perception of rhoticity and formant trajectories. Beyond the traditionally described post-alveolar approximant and the alveolar tap, phonetic realisations include pharyngealisation that may or may not be indicative of the presence of /r/, fricative noise in non-aspirated positions, approximants with different places of articulation and tongue shape, and finally a continuum of possibilities between a clear absence of /r/ and different vocalised realisations and degrees of colouring of the preceding vowel. Watt/ Llamas/ Johnson inventoried more than 40 variants in their 55,000 tokens from 160 speakers (2014:87), and even if Northern English variants account for some of these, the array of actual realisations is spectacular.

Different /r/ realisations are reported to be linked to social class, gender and age, and within derhoticised variants, different *forms* of vocalisation distinguish different groups of speakers (Dickson/ Hall-Lew 2017). However, when we consider sociophonetic variation from a more distant and perhaps necessarily simplified phonological angle, the question arises as to where the boundary should be drawn in a weakening process between a more consonantal vs. a more vocalic variant, and the criteria on which the decision is based are crucial. When does a weak approximant become a mere colouring of the vowel sound? Are r-coloured vowels rhotic segments when their articulation remains purely vocalic? Is there a difference between derhotics and non-rhotics? As it will be explained in section 3, all vocalic forms of /r/ will be considered non-realised in the results presented here.

Working-class speakers are reported to be less rhotic than middle-class speakers in general, at least as far as Glasgow and Edinburgh are concerned. Further social stratification can be observed in the frequency and the quality of approximant realisations between broadly termed working-class and middle-class speakers. For example, Lawson/ Scobbie/ Stuart-Smith (2014) show that a lower rate of approximant realisation plus a different *quality* of realisation within the approximant category is characteristic of working-class speakers. As they explain, such weakly rhotic approximants are probably due to the position of the tongue-tip, while the difference in the articulation of the approximant gives the overall rhotic quality of middle-class speech. At the same time, diastratic variation can sometimes be observed across speaker groups, for example young female speakers in Edinburgh favour post-alveolar approximants in formal style overall irrespective of social-economic background (Speitel/ Johnston 1983). Also, vocalised /r/, especially the colouring of the reduced vowel /ə/, is found typically in female speakers, both working and middle-class in Edinburg and Livingston (Lawson/ Scobbie/ Stuart-Smith 2011) while non-rhotic /r/ is typical in working-class men.

Younger and older speakers are often shown to use different /r/ variants. Alveolar taps are typical in older speakers, approximants are more widespread among younger speakers, and non-rhotic realisations are reported to be on the increase in younger age groups, especially working-class males (Jauriberry et al. 2012, Stuart-Smith/ Lawson/ Scobbie 2014). Scarcity of data and constraints on longitudinal surveys explain why evidence of change in real time is more difficult to find. Stuart-Smith et al. (2015) report on indirect real-time change comparing older male speakers recorded at different times (1970s-80s-90s and 2000s) and suggest that a higher third

[4] It is also important to note that most Scots are bilectal and code-switch between (one of the local dialects of) Scots and Scottish English. The continuum of varieties used in Scotland today may be affected to different degrees across regions and social groups.

formant observed in speakers born in the 1910s and 1920s may be a sign of /r/ weakening (Stuart-Smith et al. 2015:4). Lastly, some accounts reveal inverse tendencies in apparent time - younger speakers are more rhotic than older speakers - for example, in the border town of Eyemouth (Watt/ Llamas/ Johnson 2014), and in Edinburgh (Schützler 2010). Both papers seem to offer a logical explanation: stronger identification with a typical Scottish sounding /r/ in Eyemouth, and a side-effect of middle-class professional life in Edinburgh may result in stronger rhoticity in younger versus weaker rhoticity in older speakers, respectively.

Overall, if approximant realisations are innovations with regard to traditional Scottish trills from long ago, then clearly this innovation is not leading to change because approximant usage maintains perceptually strong rhoticity. On the contrary, speaker groups who did *not* adopt approximants at the beginning are now innovators with a new zero-realisation. Moreover, social patterning is not restricted to the coda, it also appears in the onset and intervocalically. Conscious or unconscious identification with a specific social group (e.g. Glasgow working class vs. middle-class) or a general social group (e.g. Scottish vs. English) can also have a strong impact on rhoticity.

In the following sections, the present study examines the overall rate of rhoticity in speakers from Ayr, a town which is situated outside the Central Belt of Edinburg and Glasgow. Section 4 presents results as far as correlations between rhoticity and sociophonetic factors (age, gender, socio-economic background) and phonological aspects (lexical stress and syllabic position) are concerned, and section 5 discusses these findings in the light of previous research on derhoticisation returning to the questions raised above relating to the diversity of phonetic forms.

3 Method

The dataset used in the study is part of a collection of corpora of contemporary English (the PAC corpora, from the French abbreviation of ***P****honologie de l'****A****nglais* ***C****ontemporain*; see Pukli 2006, Durand/ Przewozny 2015). Participants were recruited in a loosely interconnected local network of Ayrshire speakers living in and around the city of Ayr, Scotland. The corpus, recorded between 2002 and 2004, consists of read and spontaneous speech from ten female and eight male speakers from different socio-economic backgrounds (cf. Table 1 below). All participants were born and bred in Ayrshire, they have had infrequent contact with Anglo-English speakers both at the time of the recording and during their childhood. There was no pre-selection according to socio-economic criteria, this variable was established based on interview data.

Table 1 - Ayrshire speakers: gender, socio-economic background (working-class (WC) vs. lower-middle-class (LMC)) and age

FEMALE SPEAKERS			MALE SPEAKERS		
F1	LMC	19	M1	LMC	18
F2	LMC	25	M2	WC	19
F3	LMC	28	M3	WC	21
F4	LMC	45	M4	LMC	23
F5	LMC	48	M5	WC	50
F6	WC	50	M6	LMC	51
F7	--	60	M7	LMC	64
F8	LMC	71	M8	LMC	65
F9	WC	73			
F10	--	82			

Data used for the analysis of coda-r realisations come from informal unscripted conversations that took place in the participants' home. The casual, non-directed setting with self-selected pairs cannot guarantee authentic usage but it is reasonable to suppose that self-monitoring rapidly decreases soon after the onset of the recording. In some instances, fieldworker interviews had to be used (these were recorded in a slightly more formal context since the interviewers were outsiders).[5]

A preliminary perceptual analysis was carried out to determine the absence vs. presence of /r/ followed by an acoustic analysis on the tokens for which the preliminary analysis was non-conclusive. All vocalic forms displaying continuous flat formants throughout the rhyme were counted as non-rhotic. Reliable visual cues indicate the presence of diverse forms of /r/ articulation such as approximants characterised by a drop in F3 or taps showing a short interruption in spectral energy but there always remain ambiguous realisations. These were either discarded or categorised with rhotic realisations. This decision was motivated by a wish to compensate for the possible loss of articulated but non-perceptible consonantal /r/, namely, possible occurrences of lingual gestures taking place after phonation has ceased (Lawson/ Stuart-Smith/ Scobbie 2008).

A final total of 566 occurrences of coda /r/ after filtering (an average of 30 tokens per speaker) were analysed using Praat (Boersma/ Weenink 2013). Although /r/ is a very frequent phoneme, the greater number of words containing coda-r were function words (e.g. pronouns, prepositions, etc.) that were excluded from the sample because of their overall reduction in an unstressed position that affects every phoneme not just /r/. Thus, data comprise all occurrences of coda-r from the conversational corpus with the exception of 1) function words 2) words directly followed by a vowel sound (linking-r sites), and 3) unclear realisations due to the following segment or noise (e.g. following word with initial /r/, coarticulation and elision of the entire syllable in very rapid speech, etc).

[5] For more information on the participants, as well as on the entire recording protocol, see Pukli (2006).

4 Results

Figure 1 below shows overall results for 18 speakers, the top part of each bar representing absent /r/ realisations and the bottom part of each bar displaying pronounced /r/ realisations. The overall average of non-consonantal coda /r/ is 42 percent (individual percentages are given in Figure 1). When comparing groups of speakers, a simple chi-squared test of independence shows strong correlation between absence of /r/ and *age* (re-coded in 3 categories, p-value = 0.003793) and *word stress* (p-value = 0.01575). Correlations between the non-realisation of /r/ and gender (p-value = 0.06642), socio-economic background (p-value = 0.1084) and phonological position[6] (p-value = 0.5111) are non-significant.

Figure 1. Overall absence vs. presence of coda-r

The speakers' age ranged from 18 to 82 years, and when we compare the distribution of absent versus present r-realisations across three groups - around 20 years of age, around 50 years of age, and 60 or older - we can see in Figure 2 below that rhotic realisations (represented by a darker shade in each bar) are more prevalent in the older age range than they are in younger speakers. Non-rhotic realisations (top, light-shade area within each bar) are present in all three age groups.

[6] This variable compared r-realisations in three positions: word-medial (*ward*) vs. word-final followed by a consonant (*war*# C) vs. word-final followed by a pause (*war*##).

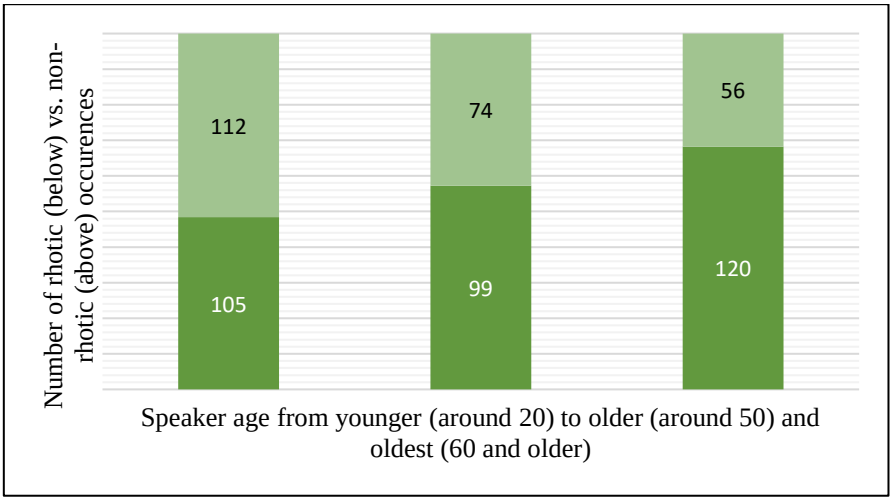


Figure 2. Realization of /r/ by age: around 20 years of age (18,19,21,23,25,28), around 50 (45,48,50,51), and 60 years of age and older (60,64,65,71,73,82).

Rhoticity is known to be sensitive to stress, Figure 3 below shows a clear difference between stressed and unstressed syllables: unstressed syllables (displayed in the second bar) are just as often non-rhotic as rhotic, whereas stressed syllables account for a higher proportion of rhoticity overall (shown in the first bar with 227 occurrences of consonantal /r/ as opposed to 138 realisations of non-consonantal /r/).

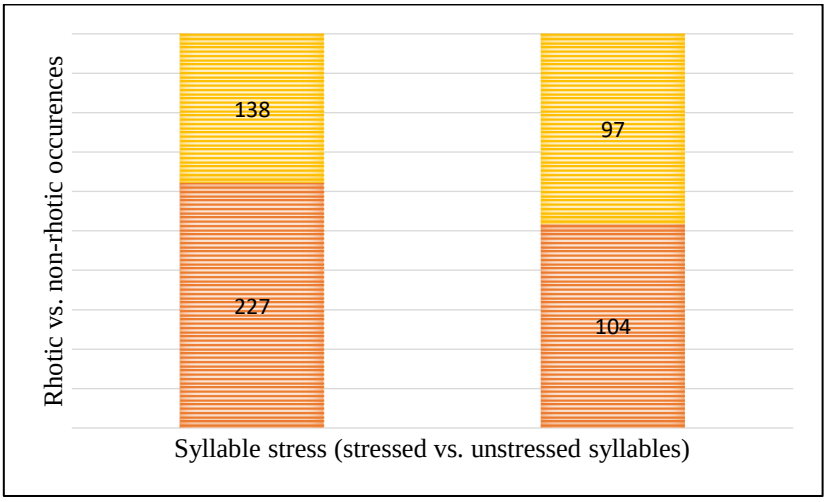


Figure 3. Realization of /r/ in stressed syllables (to the left) vs. unstressed syllables (to the right): the total of rhotic occurrences are in red (bottom) and the total of non-rhotic occurrences in yellow (top)

Since the observations originate from only eighteen speakers and hence are not independent in themselves, generalised linear mixed models (GLMM) were fitted on the full data frame with *word* and *speaker* as random effects using the *lme4* package (Bates et al. 2014) in R (R Core Team 2018). The only significant correlation found was between absence of coda-r and word-level stress (p-value = 0.03516; see Appendix for details).

The influence of the relative prominence of the syllable containing /r/ on the realisation of the sound has been shown repeatedly (Scobbie/ Stuart-Smith/ Lawson 2008; Lawson/ Stuart-Smith/ Scobbie 2008, Schützler 2010). The most remarkable finding here is the absence of correlation with age, gender or social class. The fact that Ayr is removed from the Central Belt and its

densely populated urban areas might mean that sociolinguistic patterns related to gender and socio-economic background are different. Also, social groups were not represented in equal proportions in our sample, thus perhaps the group was simply too homogeneous.

Figure 4. The percentage of rhotic realization of /r/ per speaker by year of birth

The representation of the percentage of rhotic realisations in coda position per speaker, shown in Figure 4 above plotted against the age of the speakers, does not display a clear tendency. The non-random sampling and the smallness of the sample may have led to skewed data on which the results are based. There can also be an abrupt adoption of non-rhotic forms that is taking place among speakers born in the 1990s who were too young at the time to participate in the investigation that targeted adult speakers.

Nevertheless, non-rhoticity is undoubtedly present in older speakers, which is principally why a clear tendency cannot be seen here. For a meaningful discussion, we would need to look into realisations of speakers born before 1940 to see whether our eldest speaker is a simple outlier, in which case the majority of the older age group would be fairly strongly rhotic. Interestingly, Jauriberry (2016)[7] working on a larger corpus comparing Kinross and Dundee found no significant correlation between non-rhoticity of coda-r and the age of the speakers either, although this might be due to methodological decisions in an attempt to hierarchise /r/ realisations in a single score that will be discussed in the next section. Schützler (2010), as mentioned earlier, also found just as many non-rhotic speakers in older age groups as in younger ones (even more) in middle-class Edinburgh speakers.

Another question, whether lexical items themselves could have an influence on rhoticity - some of them being consistently pronounced with a coda-r, others being systematically pronounced without coda-r - could be raised.[8] However, word sets across speakers were quite different in the part of the corpus that was analysed here due to the unscripted context of conversational data. Despite the diversity of the tokens, frequency effects were tested in Pukli (2018) but no significant correlation emerged between non-rhotic realisations and lexical frequency.

[7] In his study of 29 speakers, significant factors include word-level stress (stressed syllables are more rhotic than unstressed syllables), geographical origin (Dundee speakers are more rhotic than Kinross speakers) and phonological position (the word-final position is more rhotic than the word-medial pre-consonantal position) (Jauriberry 2016:294).

[8] I thank one of the anonymous reviewers for this suggestion that I was going to ignore.

5 Discussion

The results indicate that non-rhotic realisations of coda-r are more frequent in unstressed syllables than they are in stressed syllables. Speakers are variably rhotic in the corpus but there is no clear correlation between rhoticity and sociophonetic factors such as age, gender and socio-economic background.

5.1 Identity and prestigious forms

First, considering the overall low rate of rhoticity of the 18 speakers one can wonder to what extent this can be linked to identity and prestige. While the social stratification related to rhoticity was repeatedly shown (Lawson/ Scobbie/ Stuart-Smith 2011, Watt/ Llamas/ Johnson 2014, Lawson/ Scobbie/ Stuart-Smith 2014, Dickson/ Hall-Lew 2017) it is not clear whether the use of certain phonetic variants is conscious or not. Furthermore, speaker identities and behaviour in different parts of Scotland are not necessarily similar. On the one hand, rhotic /r/ realisations are clearly part of the repertoire of speakers affirming their Scottish identity in Eyemouth (Watt/ Llamas/ Johnson 2014), on the other hand, Ràcz (2013) argues the non-salience of /r/ variants and claims that Scottish speakers are on the whole unaware of derhoticization.

It is also interesting to compare standard Scottish English and General American in terms of their influence on regional variation. In Boston, speakers are variably rhotic, and phonological change (if there is one) is in the other direction: the previously non-rhotic New England variety becomes rhotic to converge with the rhotic norm of the United States. Navarro (2013), for example, reports an increase of rhotic coda-r realisation from 54% in unstressed syllables to 70% in foot initial syllables and to 76% in syllables with nuclear stress.

Clearly, on the scale of the present study, one cannot decide whether speaker behaviour favouring absence vs. presence of coda-r is conscious or not. In the context of code-switching between local and more standard speech forms, rhoticity is generally considered to be an attribute of the latter variety, but it was either simply not available to the speakers of the corpus, or no attempt was made to use it in casual conversations.

5.2 The diversity of phonetic realisations - articulation and perception

This leads to the next question that is crucial in determining the nature of structured variation in linguistic behaviour pertaining to the use of consonantal coda-r. Phonetically speaking it is habitual to represent rhoticity in categories deriving from articulatory action and acoustic markers, but also in terms of perceived strength. Thus, variants are often ordered based on their consonantal strength, for example from the strongest to the weakest: trill, tap, approximant, rhotacised vowel, absence (in Jauriberry 2016, Jauriberry/ Pukli 2011, see Figure 5 below) or, similarly from the strongest to the weakest: trill, tap, rhotacised schwa, approximant, derhoticised vowel, absence (in Dickson/ Hall-Lew (2017) and Lawson/ Scobbie/ Stuart-Smith (2014)). Rhotacised schwa realisations, as pointed out by Dickson/ Hall-Lew (2017), are particularly difficult to rank on such a scale because they are simultaneously very rhotic (in their perception) and non-rhotic (in their absence of any consonantal material).

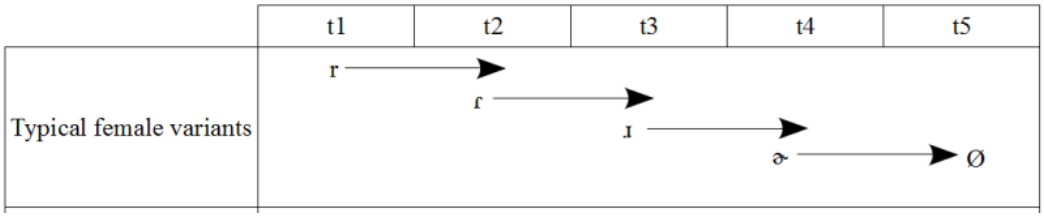


Figure 5. Possible /r/ realisations in female speakers from Jauriberry/ Pukli (2011:10, adapted from Figure 9)

Crucially, what is wrong with this approach is that it not only conflates articulatory and auditory descriptors but also creates an artificial aggregate of rhoticity expressed in terms of strength. The representation creates a false sense of continuum as put forward and argued convincingly by Dickson/ Hall-Lew (2017). It also makes the comparison of results across studies very difficult, especially when statistical tests indicate a different ranking of parameters in terms of significance depending on which phonetic categories are considered absent vs present realisations. As suggested in the following section, there are probably different trajectories in lenition from full, non-variable to weakening, variable rhoticity and there is no unique continuum that fits the different speaker groups.

5.3 The overall preference of phonetic variants

When one widens the scope of enquiry to include pre-vocalic environments as well, it is probable that non-rhoticity is in fact related to a global preference in speakers for one type of /r/ realisations over another in all syllable positions (Jauriberry/ Pukli 2011). According to a purely phonological reasoning, onset and coda /r/ realisations could be different for all speakers: taps in the onset vs. approximants in the coda, and the two could be entirely predictable contextual allophones. All findings from sociophonetic studies contradict this hypothetical distribution in the coda, but very little is known about the onset. As mentioned earlier, Wells (1982) suggested a complementary distribution-like preference for approximants word-finally and taps intervocalically, leaving the word-initial position for both. When looking at a subset of 8 speakers from the present corpus, Jauriberry/Pukli (2011) found a significant correlation between the dominant type of /r/ and the overall rate of rhoticity. A logical consequence of r-variants being socially indexed in all positions is that the sensitive site of the non-prevocalic context has divergent variants across speaker groups to start with *before* derhoticisation can take place.

Interestingly, when looking at the least frequent type of realisations, trills and taps, correlations appear: older speakers use significantly more trills and taps than approximants in the onset in Jauriberry (2016), and middle-aged working class men use a significantly greater proportion of trills and taps in the coda according to Dickson/ Hall-Lew (2017) while they remain overall less rhotic than other cohorts. This adds up because if working class male speakers have a stronger tendency to use taps in general, then they are more likely to be affected by a potential articulatory weakness word-finally, resulting in inaudible segments and a massive impression of missing r's.

To take another example, middle-class female speakers, who adopted approximants as their dominant type of r-realisation, tend to transform coda-r into a strong secondary articulation on the preceding vowel. Stuart-Smith/ Lawson/ Scobbie (2014) found approximants and perceptually strong vocalised variants typically in middle-class speakers (who remain rhotic) and taps characteristically in working class speakers (who are much less rhotic). This adoption and preference of approximant use that took place between 50-100 years ago, and that might

have seemed as a weakening at the time, or that might appear on a consonantal scale weaker than a trill today, is actually 'protecting' these users from complete derhoticization. Traditional tap users, on the other hand, are at the same time the least rhotic working-class males.

As Dickson/ Hall-Lew (2017) point it out, there is a mismatch between sociophonetic patterning, i.e. which group of speakers favour which variant, and the way in which the variants are construed as a continuum of realisations in linguistic experiments. Thus, rhotic middle-class speakers whose most frequent variant is an approximant /r/ that appears to be stable, figure as less rhotic on a phonetic scale that starts with trills and taps, the most consonantal realisations:

> "Rhoticity is interesting sociophonetically because what can be framed as continuous variation on a phonetic scale does not correspond to similarly structured variation at a socioindexical level. For example, while high rates of rhotic use appear to index middle-class and female identities, the most fortis variants (taps and trills) do not have the same social correlates." (Dickson/ Hall-Lew 2017:245)

The sociolinguistically motivated preferences make sense phonetically because the articulation of these variants is very different: globally vowel-like, on the one hand, i.e. approximants and vowels with secondary articulation (derhotics or rhotacised schwas), and consonantal on the other, i.e. taps supposedly on their way diachronically from trills to nothing. The contradiction arises from the forced alignment of the variables on a scale that conflates different dimensions. Thus, it is because of the articulatory difference between a tap and an approximant that this finding is recurrent and *not despite* the phonetic differences between the derhoticised variants and the strongly rhotic taps and trills.

5.4 The diversity of phonetic realisations on the phonological level

Phonological variation is socially salient for some features because as Eckert/ Labov (2017) succinctly put it, they are "most readily adapted to convey social meaning by their frequency, flexibility and freedom from referential functions" (2017:467). Thus, new weakened variants of /r/ are not necessarily signs of change from rhoticity to non-rhoticity but, on the contrary, they may signal stability in the system. That sounds are available for social indexing because they are unrelated to meaningful units, because of their inherent elasticity, and because of their frequency means that frequent sounds should show the widest range of variation. There will sometimes be a correlation between lexical frequency and a given phonetic variant, but the causality of this correlation is open to interpretation because frequent does not mean first to change, but simply first to vary.

Unfortunately, the conceptual enquiry that relates variationist studies to phonological theory, and what should and should not be part of the latter, lies outside the scope of the present paper but it is a fundamental question that must occur to anyone interested in sociophonetic variation. The current turn in usage-based grammar that views speech and, perhaps more arguably, language as a complex dynamical system (Kretzschmar 2015, Pierrehumbert 2016, Mufwene/ Coupé/ Pellegrino 2017, Steels/ Van Eecke/ Beuls 2018, for example) places variation back at the centre of its conceptual models: the absence of variation would make the system unviable. The phonology of a given language as a system of sounds and sound patterns is extremely flexible *and* very stable. This is not contradictory because stableness does not mean homogeneity, variation is not only present in speech, it defines speech. One is sometimes tempted to view phonology at a given moment of time as an ideal, settled system: the result of and ready to change, but functioning, and hence stable and static. Speech (and the phonological

model thereof) is certainly functioning and stable, but never static. Phonological systems are happy with variation, variation can remain present for a very long time before, after, and *without* any final change.

6 Conclusion

The results presented in this paper confirmed the role of lexical stress in stronger rhoticity in otherwise massively non-rhotic speakers with high proportion of occurrences of absent coda-r in unstressed syllables. Due to the unexpected non-correlation between weakened r-realisations and socio-economic background – gender and age were not statistically significantly constraining rhoticity in the model – it was argued that derhoticisation may not always be linked to social factors, and that conflating different dimensions in the aggregate of consonantal strength can confound the picture. It was also suggested that when sociolinguistic factors are examined, the analysis should not be restricted to coda-r, as they could explain the overall preference of /r/ realisations in the speaker irrespective of phonological position or speech style, which, in turn, can have an impact on the rate of weakened rhoticity. It was further argued that although there is undeniable variation in /r/ realisations in Scotland today, this does not necessarily imply phonological change from rhoticity to non-rhoticity.

References

Arbesman, Samuel/ Strogatz, Steven H./ Vitevitch, Michael S. 2010. The Structure of Phonological Networks Across Multiple Languages *International Journal of Bifurcation and Chaos* 20(3). 679-685.

Bates, Douglas/ Maechler, Martin/ Bolker, Ben/ Walker, Steven. 2014. Linear mixed-effects models using Eigen and S4. R package. Version 1.1-7. http://CRAN.R-project.org/package=lme4 [accessed 2019-02-08].

Beckner, Clay/ Ellis, Nick C./ Blythe, Richard/ Holland, John/ Bybee, Joan/ Ke, Jinyun/ Christiansen, Morten H./ Larsen-Freeman, Diane/ Croft, W./ Schoenemann, Tom. 2009. Language Is a Complex Adaptive System: Position Paper. *Language Learning* 59. 1-26.

Boersma, Paul/ Weenink, David. 2013. Praat: Doing phonetics by computer. Version 5.4.04. http://www.praat.org/ [accessed 2019-02-08].

Burridge, James. 2017. Spatial Evolution of Human Dialects. *Physical Review X*. 7(3). 031008-1- 031008-27.

Corominas-Murtra, Bernat/ Valverde, Sergi/ Solé, Ricard V. 2009. The ontogeny of scale-free syntax networks through language acquisition. *Advances in Complex Systems* 12(3). 371-392.

Dickson, Victoria/ Hall-Lew, Lauren. 2017. Class, Gender, and Rhoticity: The Social Stratification of Non-Prevocalic /r/ in Edinburgh Speech. *Journal of English Linguistics* 45(3). 229-259.

Durand, Jacques/ Przewozny, Anne. 2015. La variation et le programme PAC. In: Brulard, Inès/ Carr, Phil/ Durand Jacques (eds.), *La Prononciation de l'anglais: variation et structure*, Toulouse: Presses Universitaires du Midi.

Eckert, Penelope/ Labov, William. 2017. Phonetics, phonology and social meaning. *Journal of Sociolinguistics* 21(4). 467–496.

Ferrer i Cancho, Ramon/ Solé, Ricard V. 2001. The small world of human language. *Proceedings. Biological sciences* 268(1482). 2261–2265.

Goldstein, Rutherford/ Vitevitch, Michael S. 2014. The influence of clustering coefficient on word-learning: how groups of similar sounding words facilitate acquisition. *Frontiers in Psychology* 5. 1307. doi: 10.3389/fpsyg.2014.01307

Grant, William. 1914. *The Pronunciation of English in Scotland*. Cambridge: CUP.

Jauriberry, Thomas. 2016. Rhotiques et rhoticité en Ecosse: une étude sociophonétique de l'anglais écossais standard. Thèse de doctorat. Université de Strasbourg. https://tel.archives-ouvertes.fr/tel-01532023[accessed 2019-02-08].

Jauriberry, Thomas/ Pukli, Monika. 2011. *Sociophonetic variation and phonological patterning of rhotics: the case of Ayrshire* [Unpublished manuscript].

Jauriberry, Thomas/ Sock, Rudolph/ Hamm, Albert/ Pukli, Monika. 2012. Rhoticité et dérhoticisation en anglais écossais d'Ayrshire. In: *Proceedings of the joint JEP-TALN-RECITAL conference*. 89–96.

Jiang Jingyang/ Yu Wuzhe/ Liu Haitao. 2019. Does Scale-Free Syntactic Network Emerge in Second Language Learning? *Frontiers in Psychology* 10. 925.

Kretzschmar, Jr. William A. 2015 *Language and Complex Systems*. Cambridge: Cambridge University Press.

Labov, William 1994. *Principles of Linguistic Change. Vol.1: Social Factors*. Oxford: Blackwell.

Larsen-Freeman, Diane/ Cameron, Lynne. 2008. *Complex systems and applied linguistics*. Oxford: Oxford University Press.

Lawson, Eleanor/ Scobbie, James M./ Stuart-Smith, Jane. 2011. The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics* 15(2). 256-268.

Lawson, Eleanor/ Scobbie, James M./ Stuart-Smith, Jane. 2014. A socioarticulatory study of Scottish rhoticity. In: Lawson, Robert (ed.), *Sociolinguistics in Scotland*, 53-78. Basingstoke: Palgrave Macmillan.

Lawson, Eleanor/ Stuart-Smith, Jane/ Scobbie, James M. 2008. Articulatory insights into language variation and change: Preliminary findings from an ultrasound study of derhoticization in Scottish English. *Selected Papers from NWAV 25 University of Pennsylvania Working Papers in Linguistics* 14. 102-109.

Macafee, Caroline. 1983. *Varieties of English around the world: Glasgow*. Amsterdam: John Benjamins.

McAllister, Anne H. 1938. *A year's course in speech training*. London: University of London Press.

Mufwene, Salikoko S./ Coupé Christophe/ Pellegrino, François (eds.). 2017. *Complexity in Language: Developmental and Evolutionary Perspectives*. Cambridge: Cambridge University Press.

Navarro, Sylvain. 2013. Rhoticité et 'r' de sandhi en anglais: du Lancashire à Boston. Thèse de doctorat. Université Toulouse II. https://tel.archives-ouvertes.fr/tel-00927763 [accessed 2019-02-08].

Pierrehumbert, Janet B. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2. 33-52.

Pierrehumbert, Janet B./ Stonedahl, Forrest/ Daland, Robert. 2014. A model of grassroots changes in linguistic systems. arXiv:1408.1985 [cs.CL]. https://arxiv.org/abs/1408.1985 [accessed 2019-02-08].

Pukli, Monika. 2006. Investigation sociophonétique de l'anglais en Ecosse: le cas de Ayr. Thèse de doctorat, Université Toulouse 2, https://hal.archives-ouvertes.fr/tel-01986064 [accessed 2019-02-08].

Pukli, Monika. 2018 L'effet de la fréquence lexicale sur les réalisations des rhotiques en Écosse. Actes des 32e Journées d'Etudes sur la Parole. 205-212. https://jep2018.sciencesconf.org/data/ActesJEP2018.pdf [accessed 2019-02-08].

Ràcz, Péter. 2013. *Salience in Sociolinguistics: A quantitative approach*. Berlin: Mouton de Gruyter.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/ [accessed 2019-02-08].

Romaine, Suzanne. 1978. Postvocalic /r/ in Scottish English: sound change in progress? In: Trudgill, Peter (ed.), *Sociolinguistic patterns in British English*, 144-157. London: Edward Arnold.

Schützler, Ole. 2010. Variable Scottish English consonants: The cases of /ʍ/ and non-prevocalic /r/. *Research in Language* 8. 5-21.

Scobbie, James/ Stuart-Smith, Jane/ Lawson, Eleanor. 2008. Looking variation and change in the mouth: Developing the sociolinguistic potential of ultrasound tongue imaging. *ESRC end of award report, RES-000-22-2032*. Swindon: Economic and Social Research Council.

Steels, Luc/ Corominas-Murtra, Bernat/ Valverde, Sergi/ Solé, Ricard V. 2010. Language Networks: Their Structure, Function, and Evolution. *Complexity* 15, 20-26.

Steels, Luc/ Van Eecke, Paul/ Beuls, Katrien. 2018. Usage-based learning of grammatical categories. *BNAIC 2018 Preproceedings*. 253-264.

Stuart-Smith, Jane/ Lawson, Eleanor/ Scobbie, James M. 2014. Derhoticisation in Scottish English: A sociophonetic journey. In: Celata, Chiara/ Calamai, Silvia (eds.), *Advances in Sociophonetics. Studies in Language Variation 15*, 59-96. Amsterdam, The Netherlands/Philadelphia, Pennsylvania: John Benjamins Publishing Company.

Stuart-Smith, Jane/ Lennon, Robert/ Macdonald, Rachel/ Robertson, Duncan/ Sóskuthy, Márton/ José, Brian/ Evers, Ludger. 2015. A dynamic acoustic view of real-time change in word-final liquids in spontaneous Glaswegian. In: The Scottish consortium for ICPhS 2015 (ed.), *Proceedings of the 18th international congress of phonetic sciences. Glasgow, UK: University of Glasgow*.
https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS1028.pdf [accessed 2019-02-18].

Speitel, Hans-Henning/ Johnston, Paul. 1983. A sociolinguistic investigation of Edinburgh speech. Unpublished *Economic and Social Research Council (ESRC) end of grant report*. Swindon, UK.

Watt, Dominic/ Llamas, Carmen/ Johnson, Daniel Ezra. 2014. Sociolinguistic variation on the Scottish-English border. In: Lawson, Robert (ed.), *Sociolinguistics in Scotland*, 79-102. Basingstoke: Palgrave Macmillan.

Weinreich, Uriel/ Labov, William/ Herzog, Marvin I. 1968. Empirical foundations for a theory of language change. In: Lehmann, Winfred P./ Malkiel, Yakov (eds.), *Directions for historical linguistics: a symposium*, 95-195. Austin: University of Texas Press.

Wells, John C. 1982. *Accents of English*. 3 volumes. Cambridge: Cambridge University Press.

Williams, Irene F. 1909. *Phonetics for Scottish students: the sounds of polite Scottish described and compared with those of polite English*. Glasgow: James Maclehose and Sons.

Appendix - GLMM in R

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']

Family: binomial (logit)

Formula: rful ~ gender + age + SEb + WS + position + (1 | speaker) + (1 | word)

Data: munich.data

Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))

AIC	BIC	logLik	deviance	df.resid
736.8	780.2	-358.4	716.8	556

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.4696	-0.8216	0.4921	0.6973	1.6235

Random effects:

Groups	Name	Variance	Std.Dev.
word	(Intercept)	0.4976	0.7054
speaker	(Intercept)	0.5404	0.7351

Number of obs: 566, groups: word, 389; speaker, 18

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.004139	1.022192	0.004	0.9968	
genderm	-0.405236	0.437783	-0.926	0.3546	
age	0.013605	0.011053	1.231	0.2184	
SEbnwc	0.022664	0.747869	0.030	0.9758	
SEbwc	0.420131	0.814472	0.516	0.6060	
WSu	-0.487020	0.231196	-2.107	0.0352	*
positionfp	0.302156	0.328198	0.921	0.3572	
positionm	0.025883	0.268942	0.096	0.9233	

Correlation of Fixed Effects:

	(Intr)	gendrm	age	SEbnwc	SEbwc	WSu	pstnfp
genderm	-0.121						
age	-0.769	0.150					
SEbnwc	-0.792	-0.197	0.400				
SEbwc	-0.707	-0.269	0.349	0.825			
WSu	-0.099	0.016	-0.047	-0.007	-0.012		
positionfp	-0.176	-0.015	0.047	0.033	0.018	-0.018	
positionm	-0.214	-0.005	0.001	0.020	-0.002	0.266	0.537

Analysis of Deviance Table (Type II Wald chisquare tests)

Response: rful

	Chisq	Df	Pr(>Chisq)	
gender	0.8568	1	0.35463	
age	1.5150	1	0.21837	
SEb	0.7539	2	0.68596	
WS	4.4374	1	0.03516	*
position	1.0705	2	0.58552	