



# Modeling World Englishes

*Assessing the interplay of emancipation  
and globalization of ESL varieties*

EDITED BY

*Sandra C. Deshors*

John Benjamins Publishing Company

# Modeling World Englishes

# *Varieties of English Around the World (VEAW)*

ISSN 0172-7362

A monograph series devoted to sociolinguistic research, surveys and annotated text collections. The VEAW series is divided into two parts: a text series contains carefully selected specimens of Englishes documenting the coexistence of regional, social, stylistic and diachronic varieties in a particular region; and a general series which contains outstanding studies in the field, collections of papers devoted to one region or written by one scholar, bibliographies and other reference works.

For an overview of all books published in this series, please see  
<http://benjamins.com/catalog/veaw>

## **Editor**

Stephanie Hackert  
University of Munich (LMU)

## **Editorial Board**

Manfred Görlach  
Cologne

Rajend Mesthrie  
University of Cape Town

Peter L. Patrick  
University of Essex

Edgar W. Schneider  
University of Regensburg

Peter Trudgill  
University of Fribourg

Walt Wolfram  
North Carolina State University

## **Volume G61**

Modeling World Englishes  
Assessing the interplay of emancipation and globalization of ESL varieties  
Edited by Sandra C. Deshors

# Modeling World Englishes

Assessing the interplay of emancipation  
and globalization of ESL varieties

*Edited by*

Sandra C. Deshors

Michigan State University

John Benjamins Publishing Company

Amsterdam / Philadelphia





The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/veaw.g61

**Cataloging-in-Publication Data available from Library of Congress.**

ISBN 978 90 272 0123 2 (HB)

ISBN 978 90 272 6370 4 (E-BOOK)

© 2018 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Company · <https://benjamins.com>

This volume is dedicated to the memory and scholarship  
of Alexander Kautzsch, who sadly passed away  
before its publication



# Table of contents

Alexander Kautzsch: In memoriam	IX
Modeling World Englishes in the 21st century: A thematic introduction <i>Sandra C. Deshors</i>	1
From colonial dynamism to current transnationalism: A unified view on postcolonial and non-postcolonial Englishes <i>Sarah Buschfeld, Alexander Kautzsch and Edgar W. Schneider</i>	15
Stabilising domains of English-language use in Germany: Global English in a non-colonial languagescape <i>Christian Mair</i>	45
Hybridity, globalisation and models of Englishes: English in South African multilingual digital repertoires <i>Bertus van Rooy and Haidee Kruger</i>	77
Placing ELF among the varieties of English: Observations from typological profiling <i>Mikko Laitinen</i>	109
Modeling World Englishes from a cross-linguistic perspective <i>Peter Siemund</i>	133
“I’m an Anglophile, but ...”: A corpus-assisted discourse study of language ideologies in the Netherlands <i>Alison Edwards</i>	163
American and/or British influence on L2 Englishes – Does context tip the scale(s)? <i>Gaëtanelle Gilquin</i>	187
It is time that this ( <i>should</i> ) <i>be studied</i> across a broader range of Englishes: A global trip around mandative subjunctives <i>Marianne Hundt</i>	217
A corpus-linguistic account of the history of the genitive alternation in Singapore English <i>Stefan Th. Gries, Tobias Bernaisch and Benedikt Heller</i>	245

Modeling World Englishes in the 21st century: New reflections on model-making	281
<i>Sandra C. Deshors and Gaëtanelle Gilquin</i>	
Index	295

# Alexander Kautzsch

## In memoriam

This volume commemorates the life and work of Alexander Kautzsch, dedicated scholar, committed teacher, and dear friend, who suddenly and unexpectedly passed away during the finishing stages of its production. Without Alex's recent work on modeling current realities of English around the world, the present volume would be missing one of its centerpieces; without his presence around, the lives of his colleagues, friends, and loved ones have grown much poorer.

Alex studied English and German at the University of Regensburg, with a year abroad at the University of Wolverhampton, in preparation for a career as a high school teacher. It never came to that, because his gift for linguistic research immediately landed him a job at Regensburg's English Department, one of the hotbeds of the study of varieties of English in Germany. Alex's interests spanned an unusually broad range of topics, methods, and data. In his Ph.D. dissertation, he examined the history of African American Vernacular English as reflected in a number of written sources; the resulting book publication (*The Historical Evolution of Earlier African American English*, Mouton de Gruyter, 2002) is among the works regularly quoted in relation to the acrimoniously debated question of the origins of the variety, in which it stands out for its sober reasoning and cool-headedness. In his post-doctoral thesis (*The Attainment of an English Accent: British and American Features in Advanced German Learners*, Lang, 2017), Alex explored in depth and detail the extra- and intra-linguistic factors influencing accent variation among German-speaking users of English. In the World Englishes community, Alex is probably best known for the model of Extra- and Intra-territorial Forces (EIF), which he developed together with Sarah Buschfeld and which integrates non-postcolonial varieties of English as used, for example, in Germany, Cyprus, or Namibia, in a unified framework comprising both foreign- and second-language English ("Towards an integrated approach to postcolonial and non-postcolonial Englishes," *World Englishes* 13:1, 2017). For all his theoretical prowess, Alex also did not shy away from the nitty-gritty work of corpus compilation, as evidenced in his enthusiasm for ICE Namibia, the Namibian subcomponent of the International Corpus of English.

Alex was a gifted academic teacher. He was dedicated to his students, and they loved his matter-of-fact, straightforward manner. Because of his heavy teaching load, innumerable students profited from his courses; most of them enjoyed even “dry” subjects such as English syntax if he taught them. In his long time at Regensburg, he also covered just about every function available at a large university department, from student advisor and IT representative all the way to member of the university council and senate. In all of these functions, Alex maintained his positive, fact-oriented spirit, which made collaborating with him a pleasure even when mere formalities were at stake.

Alex was liked, appreciated, and admired by all of his colleagues and friends, at Regensburg and elsewhere. He will be sorely missed for his loyal and caring way, his generosity and unpretentiousness, and his inspiring passion for linguistics.

# Modeling World Englishes in the 21st century

## A thematic introduction

Sandra C. Deshors  
Michigan State University

### 1. Background and introduction<sup>1</sup>

This book is a collection of research articles whose main focus lies at the interface of the notions of emancipation and globalization of world Englishes. At a time when globalization and the advent of the internet have accelerated the spread and diversification of English varieties worldwide and at a time when “English serves as a tool and symbol of modernization, globalization and economic prosperity” (Schneider 2014: 28), the main purpose of this book is to assess what theoretical models best account for those developments as well as the emancipation of world Englishes in the early 21st century. Over the past twenty years, the English language has spread rapidly around the world as a language of broad communication. Back in 1997, Graddol already predicted the spread of English as a global communication tool, and he anticipated that it would “continue to exert pressure towards global uniformity, requiring mutual intelligibility and common standards” (Graddol 1997: 56). In the same spirit, also almost twenty years ago, Yano (2001: 125) raised the question whether a single world standard of English would develop over and above varieties of first language, second language and foreign-language uses. Interestingly, although Graddol’s prediction holds true today in that English has become a means of global communication, rather than maintaining uniformity, its globalization has accelerated the development of a plurality of distinct Englishes around the world. Today, English comes in an unprecedented range of forms, dialects and varieties, and mutual intelligibility between speakers of different (native) varieties is not always straightforward (Schneider 2017). As varieties of English flourish around the

---

1. Throughout the volume, the term ‘World Englishes’ is used broadly as the different chapters span the full range of native English varieties, Englishes as a Second Language (ESL) and Englishes as a Foreign Language (EFL).



world, it is widely acknowledged that the development and the use of new technology, access to international radio and television through satellite and the use of the internet have contributed heavily to this rapid expansion of Englishes worldwide. As suggested in Buschfeld & Kautzsch (2017), the internet and other modern communication devices provide rich communication platforms that are fertile ground for new and diverse language contact situations. In these situations, speakers of various native linguistic backgrounds are able to enrich their linguistic repertoire through their (online) interaction with speakers of different linguistic backgrounds (van Rooy & Kruger this volume). Ultimately, new varieties of Englishes are gradually able to emerge (Buschfeld et al. 2014; van Rooy & Kruger this volume). As a result, within the general process of globalization, the use of technology therefore brings closer together the phenomena of geographical expansion and linguistic change. Put differently, the use of technology and the process of globalization at large have reached beyond geographical expansion and have started to affect directly the linguistic structure of English and, more generally, the way English is being used today. In his recent paper on ‘The World System of Englishes’, Mair (2013: 255) clearly explains this correlation between geographical and linguistic expansion: “the more English spreads globally, the more heterogeneous it becomes internally [and] [t]he farther the language extends its geographical domains, the more it is affected by the multilingual settings in which it is being used”.

Importantly, beyond the geographical expansion of the English language and its ongoing linguistic developments within and across varieties, the reality in which English is evolving today is of a nature and complexity that urge us to pause and reassess our theoretical models of world Englishes and their validity in the 21st century. As the relevant literature shows, investigating the evolution of Englishes is a multifaceted phenomenon that requires researchers to account for not only linguistic factors but also historical and sociological ones. While those factors are themselves, constantly evolving with time, likewise, our theoretical models should also demonstrate ongoing development by accounting for the ways that Englishes are used as a result of their global ongoing spread. This is an important point, as Onysko (2016) observes that

[t]he global spread of the English language, its multifarious uses, and its international role constitute highly complex processes that call for [...] descriptions and have actually given rise to explanations that categorize (or ‘catalogue’ in McArthur’s terms) the plurality of Englishes in the world. (Onysko 2016: 196)

Further, given the new geopolitical situation of English around the globe and its status as a world language, a number of questions arise about what aspects of its new “reality” should be considered informative and important components of theory formation in WEs. So in this context, and at a time when the field of world

Englishes is starting to reorient itself theoretically (see Filppula et al. 2017 for a number of studies on the theme of ‘world Englishes and linguistic theory’), taking a close look at the complex new reality associated with the global spread of English is of paramount importance to ensure that our conceptual representation of world Englishes is as aligned as it can be with their various current statuses and roles.

Against this background, pinning down the forces that drive the development of Englishes today and weighing their impact on varieties of English around the world is a crucial first step in the development of adequate theoretical models (e.g. the new language contact situations alluded to earlier in the paper, speakers’ attitudes towards the infiltration of English into other languages, their position on the notion of linguistic norm and how their perception of this norm is affecting their own uses of the language; see van Rooy & Kruger (this volume) for an in-depth discussion on the recent conceptual and methodological challenges of modeling Englishes today). By its very nature the theorizing process requires us to carve out the essential mechanisms that underlie a situation, system, or process in order to achieve global comprehension of how Englishes develop (Onysko 2016). Therefore, the more accurate our understanding of the complex reality in which Englishes evolve, the more fitting our theoretical models will grow and the greater their explanatory power will be (Evans 2014: 596). Finally, making sure that our connections between the reality of English today and our theoretical models are fully in line with one another will help us formulate adequate theories of the forces that drive language development in our day and age as well as help us grow stronger as a discipline.

## 2. The dynamic process of theoretical modeling in world Englishes research

Several chapters throughout the book already offer in-depth presentations of a number of existing traditional models (see Buschfeld et al. and van Rooy & Kruger, this volume). Thus, the current section is not set up to provide an exhaustive overview of theoretical models of world Englishes but rather to cover selected models that together illustrate the dynamic nature of theoretical modeling in world Englishes research and the general trend, over the past twenty years or so, to move away from nation-bound theoretical frameworks. Although to date, earlier models of WEs such as Kachru’s (1982, 1989) or Schneider’s (2007) models still remain prevalent in the field, they have been openly criticized for “fail[ing] to reflect the vigorous spread of English and changes of its status in many regions over the last few decades” (Schneider 2011: 32) and generally for their strong foothold in a colonial framework, assigning varietal status based on political-historical considerations (Bruthiaux 2003; Deshors 2014; Edwards 2014; Edwards & Laporte 2015; Li & Mahboob 2012). In the specific case of Kachru’s model, Onysko (2016: 199) notes

how the model is a representation of English prior to the rise of the new media. As such, traditional models struggle to account for the fact that English has now become the dominant language globally and that “what dominates the linguistic ecology of the world today is not one standard language but the whole English language complex” (Mair 2013: 275).

To be more specific, Kachru’s (1982) Three Circles Model broadly refers to three concentric circles, the Inner, Outer and Expanding Circles, that represent patterns of acquisition, functional domains in which English is used across cultures and languages as well as types of spread (Kachru 1985). The Inner Circle comprises Englishes used as a mother tongue (e.g. British English, American English, Australian English; ENLs) and the Outer Circle is composed of Englishes used in former British and American colonies (ESLs) and which are acquired in a relatively naturalistic environment. In contrast, the Expanding Circle includes foreign varieties of English (i.e. EFLs) which are primarily learnt as a lingua franca in classroom settings. Kachru’s model assumes that EFL and ESL differ in that EFLs are intrinsically norm-dependent and ESLs are norm-developing. This means that ESLs “have a potential to develop their own norms and standards which are generally accepted as being characteristic features of a ‘new’ English variety” (Mukherjee 2010: 219). EFLs, however, are norm-dependent in that “foreign learners are bound to orient themselves towards exonormative standards set by speakers outside their own speech community” (Mukherjee 2010: 238). Today, despite the general effort of scholars to broaden their conception of the categorization of Englishes, Kachru’s tripartite classification remains relatively dominant in world Englishes research, despite the fact that, “static/categorical models such as Kachru’s (1985) Three Circles are no longer sufficient to capture the dynamics of English around the world” (Edwards 2016: 194; see also Kirpatrick 2007) and that, as noted in Onysko (2016), recent research has questioned the empirical validity of differentiating between ESL and EFL. Overall, the globalization of Englishes raises important questions for today’s analysts about the validity of varieties-based research and any suggestion that there exist bounded, discrete varieties at all or that there is any value in looking at national varieties or their purported features.

Taking an evolutionary/developmental perspective of world Englishes, Schneider’s (2007) Dynamic Model was received for many years as an improvement on Kachru’s classification. In a nutshell, the model assumes a five-stage developmental process including (i) *Foundation* (i.e. the first contact), (ii) *Exonormative Stabilization* (usually describing the colonial phase when the colonizers still ruled the country and British English was still considered the linguistic norm of the speech community in question), (iii) *Nativization* (when innovative forms and structures are starting to emerge), (iv) *Endonormative Stabilization* (where the new variety is starting to be codified, its speakers start to identify with their new

variety and one starts talking about X-English instead of English in X) and (v) *Differentiation* (where different social and regional varieties across the country start to emerge). Despite its long-lived popularity, however, similarly to Kachru's model, the main focus – and limitation – of the model remains almost exclusively on colonization as the driving force behind English (Edwards 2016).

In an attempt to bridge the disconnect between the development of world Englishes and adequate theoretical frameworks, a number of new theories and categorizations of world Englishes have started to emerge in recent years. These new models are based on the recognition that nation-bound varieties such as ENLs, ESLs and EFLs have become to some degree outdated (Seargant & Tagg 2011) and that “a colonial background does not necessarily and equally lead to the development of fully-fledged, prototypical second-language varieties” (Buschfeld, Kautzsch & Schneider this volume). For instance, a small number of scholars such as Mair (2003), Sand (2005) and Szmrecsanyi & Kortmann (2009) have started to move away from the prevalent theoretical models by exploring ‘angloversals’, namely linguistic features “that tend to recur in vernacular varieties of a specific language” (such as adverbs that tend to have the same morphological form as adjectives, the absence of plural marking after measure nouns, the lack of inversion in main clause yes/no question). In the same spirit, Mair (2013) and Schneider (2007, 2014) developed new theoretical frameworks that pay due attention to the forces of globalization, namely the World System of Englishes model, in the case of Mair (2013), and the Dynamic Model and the Transnational Attraction model, in the case of Schneider (2007 and 2014, respectively). In sharp contrast with prevalent traditional approaches, Mair (2013) assumes, based on de Swaan's (2010) language system theory, that a hub variety (specifically, American standard English) is a potential factor in the development of all other varieties. Mair (2013) generally argues that, at a time of cultural globalization, reaching beyond territorial approaches to English varieties is a crucial aspect of model-making. According to Mair (2013: 262), his model breaks new ground “in shedding light on the differential power of non-standard varieties of English, particularly in post-colonial world, in currents of migration and in the global mediasphere”.

In contrast, Schneider's (2007, 2014) Dynamic and Transnational Attraction models focus on Expanding varieties and offer a framework to capture the dynamism, hybridity and creativity of Expanding Circle varieties (e.g. China, Korean, Japan Englishes). Both of Schneider's models stand out from earlier frameworks such as Kachru's Three Circles in that they take a diachronic perspective and assume an underlying evolutionary path common to all postcolonial varieties. However, according to Buschfeld & Kautzsch (2017), despite this improvement “the Dynamic Model shows one major disadvantage in that it [still] does not fully grasp the complex realities of today's global status and spread of the English language” (Buschfeld

& Kautzsch 2017: 104). Despite their underlying commonality of attempting to bridge the gap between realities of English worldwide and adequate theoretical frameworks, and despite being promising, Mair's and Schneider's models are still in their infancy and they are yet to be put to the test. Further, it remains to be shown whether, and if so how, other theoretical models such as Moag (1982; described in detail in Gries et al, this volume), Mesthrie (2006) or Bao (2015), can also help us address the changes currently to be met in world Englishes research.

Building on Schneider's (2007, 2014) models, Buschfeld & Kautzsch (2017) step further in the direction of encapsulating as faithfully as possible the complex set-up of English varieties around the world and conceptualizing this reality into a theoretical model of Extra- and Intra-territorial Forces (EIF; described in detail in Buschfeld, Kautzsch & Schneider, this volume). Broadly, the EIF model reflects "the [current] need for a new, integrative framework which [...] helps describe the diverse forms of English worldwide and relates them to each other, not only in terms of their development but also with respect to their current status and linguistic forms" (Buschfeld & Kautzsch 2017: 104). In sharp contrast with earlier models, Buschfeld & Kautzsch (2017) aim for a pivotal joint approach to postcolonial and non-postcolonial settings involving a variety of aspects such as language policies, language attitudes, globalization and acceptance of globalization, foreign policies, sociodemographics, etc. More specifically,

[w]ith respect to globalization, the extra-territorial side mainly finds expression in, for example, linguistic and also cultural influences coming from the Internet, US popular culture, and modern media as well as trading relations between countries (i.e. forces includes any factor entering the country from the outside). However, this also has an intra-territorial side (i.e. forces are such that mainly operate on a local, that is, national or regional, level and therefore influence the cultural and linguistic development from within) since territories differ with respect to whether and to what extent they accept or even admit these facets of globalization.

(Buschfeld & Kautzsch 2017: 114)

Buschfeld & Kautzsch (2017: 113) argue for five major subcategories across both extra- and intra-territorial forces, including: (attitudes towards) colonization, language policies, (acceptance of) globalization, foreign policies and the sociodemographic background of a country. In light of these latest research efforts to align world Englishes with the reality they are currently part of, Mair's (2013), Schneider's (2014) and Buschfeld & Kautzsch's (2017) models represent a conceptual starting point for this volume which ultimately aims towards model-making in general and to open up theoretical discussions related to the developments and current challenges in world Englishes research. In that respect, the present volume is squarely in line with Onysko's (2016) view that

[o]verall, the discussion of the major models shows the complexities and the many factors that are involved in understanding the intricate picture of world Englishes [and] future research could gain from describing the interaction of the basic forces that shape the development of a variety. (Onysko 2016: 201)

### 3. First steps towards updating theoretical models

As a first step towards re-thinking our theoretical models, this volume brings together contributors who were allocated the task to test the reliability, usefulness, and adequacy of existing models to assess 21st century world Englishes, by evaluating empirically the benefits and limitations of those models. As “the field of world Englishes research is faced with new challenges in the categorization of the many different existing types of English” (Buschfeld & Kautzsch 2017: 104), this volume begins to address these challenges by focusing on the modeling process itself and what it should entail in the context of world Englishes. So at a time when research in world Englishes is still dominated by theoretical frameworks no longer in line with the contemporary dynamics of English use, it is hoped that this volume will make an important contribution to this research area by providing constructive assessments of new theoretical models of World Englishes by experts in the field and by contributing to a timely and necessary scholarly conversation of what constitutes adequate theoretical models of world Englishes in the 21st century. Further, as the first volume of its kind to test newly-developed theoretical models, this book showcases a collection of studies designed to serve as a springboard for future research on the globalization of Englishes and to contribute to a new and upcoming research trend. Collectively, the contributions featured in this book offer an in-depth exploration of the notions of linguistic emancipation and globalization based on (i) a wide range of ESLs, EFLs and ELF, (ii) synchronic as well as diachronic data, (iii) a variety of methodological approaches (e.g. corpus-based, sociolinguistic, ethnographic), and (iv) different data resources (e.g. social media, multiplayer online games, journalistic data, GloWbE, Corpus of Historical Singapore English, thematic blogs). Finally, this book provides an opportunity for scholars involved in the project to begin to develop their own theoretical models to address possible limitations identified in existing models (e.g. van Rooy & Kruger’s contribution). Ultimately, it is my hope that, together, the papers included in this volume will reflect the dynamic and multifaceted process that is the theoretical modeling of world Englishes in our day and age.

#### 4. The contributions to this volume

To assess and discuss the modeling process of world Englishes, the contributions in this volume adopt a wide range of theoretical and empirical approaches as well as foci such as language contact, typological comparisons, speakers' attitudes towards globalized English, linguistic norms and standards and the use of large-scale diachronic data. In the first chapter, **Buschfeld, Kautzsch & Schneider** explore the intrinsic relationship between what is happening in the real world and how linguists attempt to account for this reality by considering major theoretical aspects of modeling world Englishes in the 21st century. In the first part of their chapter, Buschfeld et al. define, discuss and compare the scope and mutual relationships of older and more recent approaches to the theorizing of WEs and, in the second part of their chapter, they relate these approaches to some examples and applications from post-postcolonial nations and contexts, including Namibia, Cyprus, Greece, Germany, Bosnia-Herzegovina, South-East Asia, social media, multiplayer online games, and others. Tying in their observations with the general postulates of Schneider's (2007) Dynamic Model and his notion of "Transnational Attraction" (Schneider 2014) as well as the different forces identified in Buschfeld & Kautzsch's (2017) EIF model, they argue that only models geared towards an integrated analysis of postcolonial and non-postcolonial Englishes can adequately capture the current linguistic realities worldwide and show how colonialism is not the only decisive force behind such developments and is in no way mandatory.

Starting from the vantage point of language contact, Mair's and van Rooy & Kruger's chapters both clearly demonstrate how fertile a breeding ground language contact situations are for the expansion of the English language and how, in Schneider's (2000: 204) own words, "[t]he application of language contact to our understanding of world Englishes bears important implications for reconceptualizing some ingrained notions in the field" (see also Siemund & Davydova 2017). While this type of approach is of course not new (see Filppula 1999; Mesthrie 2006, 2010; Mesthrie & Bhatt 2008; Ansaldo 2010; Bao 2015; Biewer 2015; Mufwene 2015; and the contributions in Schreier & Hundt 2013), Mair's and van Rooy & Kruger's chapters offer fresh, insightful and unique perspectives by focusing on new types of data. In the case of Mair's chapter, the study deals with English uses within the domain of urban youth cultures (specifically, German popular music) and communication between African immigrants, representatives of German institutions and the resident populations. More specifically, through a qualitative analysis within the theoretical frame of his 2003 World System of Englishes model, Mair differentiates between elite (business, academia) and non-elite (pop culture, migrants) domains as regards the use of English in Germany and convincingly argues that language mixing is largely restricted to the non-elite domains. Further, Mair shows



how through songs and lyrics hip hop communities develop a languagescape that ultimately transcends national borders and language boundaries. Ultimately, the chapter provides empirical evidence of how strongly integrated English has become to the multilingual languagescape of contemporary Germany. In their chapter, **van Rooy & Kruger** focus on the case of South African English and the usefulness of multilingual digital repertoires to expand on current theoretical models of WEs. After a close examination of the current empirical challenges triggered by the globalization of Englishes such as multilingualism, hybrid varieties, online communication and complex identities, van Rooy & Kruger adopt a corpus-linguistic methodological approach to analyze a corpus of online data consisting of interactive user comments that accompany daily summaries of the content of the most popular South African television soap operas. Based on those data, van Rooy & Kruger observe how the core of online interactions consists of a shared pool of English resources as well as global and local nonstandard English forms complemented by forms from South African languages. Van Rooy & Kruger conclude that adequate theoretical models of Englishes should account for a diverse resource pool and the diversity of text types that emerge from online interactions.

Bridging the gap between language contact and typology, Onysko (2016) recently observed that

if language contact is taken as a lens for looking at the multitude of scholarly reified Englishes and their dynamic uses, a new typology of Englishes emerges, which can be taken as an access point to studying certain facets of world Englishes and Englishes in the world. (Onysko 2016: 197)

In this context, Laitinen's and Siemund's chapters follow a recent trend in WEs research to adopt typological perspectives to explore how Englishes develop world-wide (Filppula et al. 2017). More specifically, **Laitinen's** chapter profiles ELF varieties vis-à-vis other varieties of English. Drawing on corpus data, the contribution specifically uses an aggregate data analysis method inspired by work in quantitative morphological (dialect) typology to establish the extent to which ELF differs from other varieties in terms of three parameters: analyticity, syntheticity, and grammaticity. Laitinen's study yields empirical results pointing towards structural differences between EFL and ESL varieties. In contrast, **Siemund's** chapter adopts a cross-linguistic typological perspective to model world Englishes. More specifically, the author assesses the distinctions and commonalities between language-internal and typological approaches to language variation. In this endeavor, particular attention is paid to language universals. While world Englishes unveil typological variation, they also reveal linguistic features that are rarely observed in other languages and that should inform theoretical models of Englishes world-wide.



Compared to the above chapters, **Edwards'** contribution takes on a more socio-linguistic perspective and focuses primarily on the attitudes of non-native English speakers towards the infiltration of the English language into other languages (Dutch, specifically), and to what extent those attitudes have an effect on the spread of English within local communities. In this regard, research has shown that

[a] speaker's attitude towards language behavior is the decisive criterion that regulates the occurrence of language contact. A speaker's choice of applying or perpetuating contact features in their language use is seen as a contextually-bound reaction that reflects the speaker's communicative intentions. (Onysko 2016: 211)

In this context, Edwards' chapter is situated within the literature on world Englishes and that on Transnational Attraction. More specifically, Edwards is primarily concerned with perceptions of English in the Netherlands through recurrent discourse patterns in a corpus of language-attitudinal commentary of 724 Dutch informants, analyzed using computer-assisted corpus linguistics and charting perceptions of English in the Netherlands. She concludes that English is more than a foreign language in the Netherlands, a mere tool for international communication; rather, it is an additional local language for creative self-expression and identity performance. Connecting her findings with Schneider's (2014) Transnational Attraction model, Edwards concludes that the new role of English transcends solely economic motivations.

**Gilquin's** and **Hundt's** chapters explore the modeling process of Englishes through the notions of norm, standards and nativeness. As Yano (2001) notes, even though the distinction between standard and nonstandard use of English remains prevalent in the field today, the notion of native norm has become, more than ever, questionable as a plurality of norms have started to emerge: speakers of English, native and non-native alike, are increasingly aware of the existence of different norms and "English for global use should be dissociated from the norm of any English-speaking society" (Yano 2001: 129). In this context, Gilquin's chapter explores the notion of norm by assessing to what extent Europe-based EFL varieties, traditionally assessed against the British English yardstick, show influences of American English. Her investigation is based on the analysis of twenty pairs of items distinctive of American and British English as extracted from the Global Web-based English Corpus and the EF-Cambridge Open Language Database. Overall, Gilquin observes a global influence of American English, as predicted by Mair's model. Thus, like ESL varieties, EFL varieties may be subject to the forces of globalization and the resulting influence of American English. Unlike ESL varieties, however, in EFL these forces may not lead to nativization processes through which speakers would express their local identity. That said, for Gilquin, EFL varieties appear to be worthy of inclusion in the World System of Englishes, although their position within the hierarchy would need to be defined, as would the real impact

of Americanization on EFLs. Continuing with the notion of norm, **Hundt's** contribution complements Gilquin's discussion in that she identifies possible traces of Americanization in the uses of mandative constructions worldwide. More specifically, Hundt focuses her attention on the alternation between mandative subjunctives and modal periphrastic constructions with *should* across ENL, EFL and ESL varieties. Overall, her findings are not squarely in line with any individual model of world Englishes such as Kachru's (1982, 1989) tripartite distinction or Strevens' (1992) genetic model according to which varieties are grouped based on whether they derive from British and American Englishes. Ultimately, Hundt's study points towards the need to approach theoretical models as "network[s] of local centers which speakers might be relating to", thus urging us to reconsider norm orientation as a central aspect of contemporary model-making.

**Gries, Bernaisch & Heller's** chapter addresses the issue of modeling Englishes from the perspective of synchronic vs. diachronic data. At a time when the great majority of large-scale corpus-based research in WEs is based on synchronic data to infer patterns of development of English varieties, Gries et al.'s chapter provides the first sound analysis of processes of structural nativization based on real-time corpus data. Focusing on the history of the genitive alternation (*of-* vs. *s-*genitive) in Singapore English based on corpus data covering both British English (as the historical input variety) and Singapore English, the study examines nearly 7000 instances of the two constructions annotated for phonetic, semantic, syntactic, and pragmatic variables. Throughout their contribution, the authors demonstrate the methodological limitations of the typical apparent-time approach taken by most studies of structural nativisation. Methodologically, Gries et al.'s study stands out as an extension of the regression-based method of Multifactorial Prediction and Deviation Analysis with Regressions/Random Forests (MuPDAR) method to deal with historical data. With this approach, the authors are not only able to assess how genitive choices in Singapore English differ from those in British English but also how genitive choices changed over time in Singapore English. Relating their findings to Moag's (1982) and Schneider's (2007) models, Gries et al. discuss how diachronic MuPDAR analyses can have significant theoretical implications for the modeling of structural nativization in WEs.

In the final section of the book, **Deshors and Gilquin** take stock of the approaches adopted throughout the volume as well as the findings that those approaches have helped uncover. By bringing together converging lines of argumentation and findings across contributions, Deshors & Gilquin make suggestions as to what contemporary theoretical models of WEs should look like in order to truly capture the diversity of uses of globalized Englishes. In doing so, Deshors & Gilquin discuss possible avenues for future research in the modeling process of WEs.

Considering all the above contributions, I hope that the present volume will stimulate scholarly conversations and exchanges in our collective effort to understand the many forces at play behind the globalization of English and how those forces leave their mark on the diverse uses and structure of the English language. Finally, I would like to thank Stephanie Hackert, the series editor, and Kees Vaes for supporting the idea of this volume right from the start and for providing valuable feedback and editorial support throughout the preparation of the volume. I am also very grateful to all contributors who showed much interest in participating in this project and who were a pleasure to work with. I wish to thank all the reviewers who generously contributed their time and expertise to this volume. Last but not least, thank you also to Alison Edwards and Tobias Bernaisch who provided me with helpful feedback on this introductory chapter.

## References

- Ansaldò, U. 2010. Contact and Asian varieties of English. In *Handbook of Language Contact*, R. Hickey (ed.), 498–517. Malden, MA: Wiley-Blackwell.  
<https://doi.org/10.1002/9781444318159.ch24>
- Bao, Z. 2015. *The Making of Vernacular Singapore English: System, Transfer and Filter*. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139135375>
- Biewer, C. 2015. *South Pacific Englishes: A Sociolinguistic and Morphosyntactic Profile of Fiji English, Samoan English and Cook Islands English* [Varieties of English Around the World G52]. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g52>
- Bruthiaux, P. 2003. Squaring the circles: Issues in modeling English worldwide. *International Journal of Applied Linguistics* 13(2): 159–178. <https://doi.org/10.1111/1473-4192.00042>
- Buschfeld, S. & Kautzsch, A. 2017. Towards an integrated approach to postcolonial and non-postcolonial Englishes. *World Englishes* 36(1): 104–126. <https://doi.org/10.1111/weng.12203>
- Buschfeld, S., Hoffmann, T., Huber, M. & Kautzsch A. 2014. Introduction. In *The Evolution of Englishes: The Dynamic Model and Beyond* [Varieties of English Around the World G49], S. Buschfeld, T. Hoffmann, M. Huber, & A. Kautzsch (eds), 1–17. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g49.01int>
- de Swaan, A. 2010. Language systems. In *The Handbook of Language and Globalization*, N. Coupland (ed.), 56–76. Malden, MA: Wiley-Blackwell. <https://doi.org/10.1002/9781444324068.ch2>
- Deshors, S. C. 2014. A case for a unified treatment of EFL and ESL: A multifactorial approach. *English World-Wide* 35(3): 279–307. <https://doi.org/10.1075/eww.35.3.02des>
- Edwards, A. & Laporte S. 2015. Outer and Expanding Circle Englishes: The competing roles of norm orientation and proficiency levels. *English World-Wide* 36(2):135–169.  
<https://doi.org/10.1075/eww.36.2.01edw>
- Edwards, A. 2014. The progressive aspect in the Netherlands and the ESL/EFL continuum. *World Englishes* 33(2): 173–194. <https://doi.org/10.1111/weng.12080>
- Edwards, A. 2016. *English in the Netherlands: Functions, Forms and Attitudes* [Varieties of English Around the World G56]. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g56>

- Evans, S. 2014. The evolutionary dynamics of postcolonial Englishes: A Hong Kong case study. *Journal of Sociolinguistics* 18(5): 571–603. <https://doi.org/10.1111/josl.12104>
- Filppula, M. 1999. *The Grammar of Irish English*. London: Routledge.
- Filppula, M., Klemola, J. & Sharma, D. (eds). 2017. *The Oxford Handbook of World Englishes*. Oxford: OUP.
- Graddol, D. 1997. *The Future of English*. London: The British Council.
- Kachru, B. B. 1982. Models for non-native Englishes. In *The Other Tongue: English across Cultures*, B. B. Kachru (ed.), 48–74. Urbana IL: University of Illinois Press.
- Kachru, B. B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the World. Teaching and Learning the Language and Literatures*, R. Quirk & H. G. Widdowson (eds), 11–30. Cambridge: CUP.
- Kachru, B. B. 1989. World Englishes and applied linguistics. In *Languages & Standards: Issues, Attitudes, Case Studies* [Anthology Series 26], M. L. Tickoo (ed.), 178–205. Singapore: SEAMEO Regional Language Centre.
- Kirpatrick, A. 2007. *World Englishes. Implications for International Communication and English Language Teaching*. Cambridge: CUP.
- Li, E. & Mahboob, A. 2012. *English Today: Forms, Functions, and Uses*. Hong Kong: Pearson Education.
- Mair, C. 2003. Kreolismen und verbales Identitätsmanagement im geschriebenen jamaikanischen Englisch. In *Zwischen Ausgrenzung und Hybridisierung*, E. Vogel, A. Napp, & W. Lutterer (eds), 79–96. Würzburg: Ergon.
- Mair, C. 2013. The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34(3): 253–278. <https://doi.org/10.1075/eww.34.3.01mai>
- Mesthrie, R. & Bhatt, R. M. 2008. *World Englishes: The Study of New Linguistic Varieties*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511791321>
- Mesthrie, R. 2006. Contact linguistics and world Englishes. In *The Handbook of World Englishes*, B. B. Kachru, Y. Kachru, & C. Nelson (eds), 273–288. Maldon, MA: Wiley-Blackwell.
- Mesthrie, R. 2010. Contact and African Englishes. In *Handbook of Language Contact*, R. Hickey (ed.), 518–537. Maldon, MA: Wiley-Blackwell. <https://doi.org/10.1002/9781444318159.ch25>
- Moag, R. 1982. The life cycle of non-native Englishes: A case study. In *The Other Tongue: English across Cultures*, B. B. Kachru (ed.), 233–252. Urbana IL: University of Illinois Press.
- Mufwene, S. 2015. Colonization, indigenization, and the differential evolution of English: Some ecological perspectives. *World Englishes* 34(1): 6–21. <https://doi.org/10.1111/weng.12129>
- Mukherjee, J. 2010. Corpus-based insights into verb-complementational innovations in Indian English: Cases of nativised semantico-structural analogy. In *Grammar between Norm and Variation*, A. N. Lenz & A. Plewnia (eds), 219–241. Frankfurt: Peter Lang.
- Onysko, A. 2016. Modeling world Englishes from the perspective of language contact. *World Englishes* 35(2): 196–220. <https://doi.org/10.1111/weng.12191>
- Sand, A. 2005. The effects of language contact on the morpho-syntax of English. In *Anglistentag 2004 Aachen, Proceedings*, L. Moessner (ed.), 449–460. Trier: Wissenschaftlicher Verlag Trier.
- Schneider, E. 2000. Feature diffusion vs. contact effects in the evolution of New Englishes: A typological case study of negation patterns. *English World-Wide* 21(2): 201–230. <https://doi.org/10.1075/eww.21.2.03sch>
- Schneider, E. W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511618901>

- Schneider, E. W. 2011. *English around the World. An Introduction*. Cambridge: CUP.  
<https://doi.org/10.1017/CBO9780511618901>
- Schneider, E. W. 2014. New reflections on the evolutionary dynamics of world Englishes. *World Englishes* 33(1): 9–32. <https://doi.org/10.1111/weng.12069>
- Schneider, E. 2017. Models of English in the World. In *The Oxford Handbook of World Englishes*, M. Filppula, J. Klemola & D. Sharma (eds), 35–60. Oxford: OUP.
- Schreier, D. & Hundt, M. (eds). 2013. *English as a Contact Language*. Cambridge: CUP.
- Seargeant, P. & Tagg, C. 2011. English on the internet and a ‘post-varieties’ approach to language. *World Englishes* 30(4): 496–514. <https://doi.org/10.1111/j.1467-971X.2011.01730.x>
- Siemund, P. & Davydova, J. 2017. World Englishes and the study of typology and universals. In *The Oxford Handbook of World Englishes*, M. Filppula, J.Klemola & D. Sharma (eds). Oxford: OUP.
- Stevens, P. 1992. English as an international language: directions in the 1990s. In *The Other Tongue: English across Cultures*, B. B. Kachru (ed), 19–26. Urbana: University of Illinois Press.
- Szmrecsanyi, B. & Kortmann, B. 2009. Vernacular universals and angloversals in a typological perspective. In *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*, M. Filppula, J. Klemola & H. Paulasto (eds), 33–53. London: Routledge.
- Yano, Y. 2001. World Englishes in 2000 and beyond. *World Englishes* 20(2): 119–131.  
<https://doi.org/10.1111/1467-971X.00204>

# From colonial dynamism to current transnationalism

## A unified view on postcolonial and non-postcolonial Englishes

Sarah Buschfeld, Alexander Kautzsch and Edgar W. Schneider  
University of Regensburg

The ever-increasing dynamism of the diffusion of English calls for an integrated approach to postcolonial and non-postcolonial Englishes and to new contexts such as transnational cyberspace, new media, or “grassroots” usage. We focus on major theoretical approaches to modelling World Englishes in the 21st century by briefly defining, discussing, and comparing older and more recent approaches to the theorizing of World Englishes, their scope and their mutual relationships, notably the “Dynamic Model”, “Transnational Attraction”, and the “Extra- and Intra-territorial Forces Model”. We discuss these approaches in relation to some examples and applications from especially post-postcolonial nations and contexts, including South-East Asia, Namibia and Germany, Cyprus and Greece, Bosnia-Herzegovina, social media (with examples from Facebook), fanfiction writing, and multiplayer online games.

**Keywords:** English in Bosnia-Herzegovina, Cyprus English, Dynamic Model, Extra- and Intra-territorial Forces Model, Facebook, fanfiction, German English, globalization, Greek English, multiplayer online games, Namibia, postcolonialism, South-East Asia, theories of World Englishes, Transnational Attraction

### 1. Introduction

The starting point of World Englishes as a coherent, self-aware scholarly discipline can be dated back to the early 1980s, with the publication of two influential collective volumes which explicitly noted (and in the case of Kachru’s book promoted) the existence of new and distinct national varieties of English (Bailey & Görlach 1982; Kachru 1982/92) and the foundation of two journals devoted to the subject

(*English World-Wide*, since 1980; and *World Englishes*, since 1982). It seems hard to believe that this is barely a little more than three decades ago, given the growth, productivity and development of the discipline since then. We have seen an immensely rich array of feature documentations and studies, based on fieldwork data, large-scale electronic corpora and other sources, of sociolinguistic and historical descriptions, of political debates and pedagogical applications, of analyses of cultural adaptations, and, last but not least, of explanations, generalizations, modeling and theory formation. The momentum of the discipline has been a direct consequence of the vibrancy of linguistic developments, with English continuously moving, expanding and growing into new regions, functions and application domains.

The 21st century in particular has been experiencing an ever-increasing dynamism of the diffusion of English into new territories and contexts. While once a fairly clear distinction between second-language and foreign-language varieties could be postulated, today this distinction is best seen as a continuum (e.g. Buschfeld 2013) and is increasingly getting blurred (see e.g. Edwards 2016), calling for an integrated approach to postcolonial and non-postcolonial Englishes (e.g. Buschfeld & Kautzsch 2017; Buschfeld & Schneider 2018). In addition, English has been encroaching into new usage contexts such as transnational cyberspace, new media, and creative “translanguaging” (employed in Canagarajah 2013), or “grassroots” usage (Schneider 2016b) and is used there in innovative ways; some of these will be looked at in greater detail in this paper. Usage conditions and forms vary along a range of parameters, including proficiency levels, exposure to different varieties and forms, regional and social settings, formality of the situation, number and hierarchical relations between interactants, and mode (spoken, written, electronic, ...), and others. The early models of World Englishes tended to abstract from real-life details quite strongly and to categorize varieties along fairly generic lines, disregarding such parameters and largely ignoring internal variability. The recent dynamism and increasing fragmentation of linguistic developments as well as the desire to reach more finely-graded levels of granularity in accounting for variability within global uses of English have motivated several innovative proposals, such as a “post-variety approach” called for by Seargeant and Tagg (2011). Their approach has been influenced by poststructuralist thinking in the social sciences and, more specifically, sociocultural and ethnographic approaches in sociolinguistics represented in notions such as “transcultural flows” (Pennycook 2007), “translingual practice” (Canagarajah 2013) or “sociolinguistics of globalization” (Blommaert 2010). Motivated by such considerations, some newer models, partly further developments of earlier ones, have been proposed in the recent past. It is the purpose of this contribution to discuss and apply, and possibly test, these innovative models in the following sections. We will survey theory developments in the first part of this paper, moving from the older models via a closer look at motivations for new



ones to discussions of the innovative frameworks; and will then, in a second part, briefly describe some interesting new usage contexts of English and apply the varying frameworks comparatively.

## 2. Theorizing World Englishes

### 2.1 Established approaches

During the 1970s and 1980s, the growing rejection of a monolithic view of the English language (cf., e.g., McArthur 1987: 9) led to the development of several models (e.g. Kachru 1985; McArthur 1987; Görlach 1990[1988]) accounting for the formation of different varieties of English around the world. The section at hand surveys the three most influential approaches, i.e. the ENL-ESL-EFL distinction (English as a Native Language, English as a Second Language, English as a Foreign Language; cf. Quirk et al. 1972: 3–4), Kachru’s (1985) Three Circles of World Englishes, and Schneider’s (2003, 2007) Dynamic Model. For a discussion of further models like McArthur’s (1987) Circle of World English, Görlach’s (1990 [1988]) Circle of International English, Gupta’s (1997) classification of “output types”, or Mesthrie & Bhatt’s (2008) classification of The English Language Complex, see Kautzsch & Buschfeld (forthcoming).

One of the earliest but still most widely used categorizations of World Englishes is the ENL-ESL-EFL distinction (see, among many others, Görlach 1990[1988]; cf. Schneider 2007: 12, 2011: 30). Building on Strang’s (1970) classification of speakers of English around the world as A, B, and C speakers, Quirk et al. (1972: 3–4, see also 1985: 3–4) devised their own terminology of distinguishing between English as a “native language”, “second language”, and “foreign language”. In ENL countries (e.g. USA, Great Britain, and Australia), English is spoken as a native language by most speakers. In ESL countries (e.g. India, Singapore, and Nigeria) English has prominent intranational functions in many domains (e.g., education, media, politics, or law), is used for interethnic communication and coexists with the indigenous languages of the local population, while in EFL countries English is mostly restricted to international communication and is primarily learned through formal education (cf. Quirk et al. 1985: 4–5; Schneider 2007: 12).

The second widely adopted model is Kachru’s (1985) Three Circles of World Englishes, which, based on the ENL-ESL-EFL distinction, classifies the English-speaking world into Inner Circle, Outer Circle, and Expanding Circle countries, respectively (cf. Schneider 2011: 31). In the Inner Circle countries, i.e. the United Kingdom, the United States, Australia, Canada and New Zealand English, English is the *de facto* or *de jure* official (Bruthiaux 2003: 160) and primary language (Kachru



1985: 12), as well as the native language for most inhabitants. Inner Circle Englishes are thus typically endonormative in orientation and provide norms for the Expanding and partly the Outer Circle (Kachru 1985: 16, 1992b: 5). Spoken non-natively in countries like India, Kenya, or Singapore, Outer Circle Englishes are the by-product of extended (British) colonization, with political, sociocultural, and linguistic changes facilitating the nativization and institutionalization of English. In such typically bi- or multilingual countries, English frequently serves as one of the *de jure* official languages (Kachru 1985: 12–13), is used in a wide range of domains (e.g., education, administration, literature) and across ethnic groups, and often develops widely accepted local norms (cf. Kachru 1985: 17). By contrast, Expanding Circle countries, like China, Indonesia, Greece, Japan, or Saudi Arabia, were not subjected to British (or American) colonial rule. As a consequence, English is traditionally considered a foreign language there, is taught through formal education, mainly serves as a lingua franca mostly for international communication, and is exonormatively oriented towards the Inner Circle, mostly British and American English (Bruthiaux 2003: 160). Consequently, these Englishes are often referred to as “norm-dependent varieties” (Kachru 1985: 17).

What has been identified as problematic about both the ENL-ESL-EFL distinction and Kachru’s circles is that they are too static, imprecise, and superficial, i.e. they neither capture the heterogeneity of speech groups (e.g. both ESL and ENL being spoken in one country), nor transitions from one type of English to another (e.g. from EFL to ESL), nor diachronic developments.

This is where Schneider’s (2003/2007) Dynamic Model comes in, in which, on top of a flexible set of classificatory parameters, the diachronic perspective takes center stage. Schneider (2007: 21) suggests “an underlying uniform process” in the evolution of post-colonial Englishes around the globe, which involves five major stages: (1) foundation, (2) exonormative stabilization, (3) (structural) nativization, (4) endonormative stabilization, and (5) differentiation. In other words, the English language is relocated to a new territory where new social and linguistic realities pave the way for linguistic nativization and stabilization, and ultimately potential internal differentiation (Schneider 2007: 29–30). In each phase, four parameters influence the setting in which the English language undergoes some change: The historical and political conditions in a particular territory (“extralinguistic factors”) shape new “characteristic identity constructions”; these have an effect on the language contact situation as well as on language attitudes and use (“sociolinguistic determinants of the contact setting”) and ultimately lead to the emergence of lexical, phonological, and grammatical characteristics (“structural effects”) (Schneider 2007: 30–35).

The model is centered around the conceptualization and realignments of identity constructions and “their symbolic linguistic expressions” (Schneider 2007: 28) in the settler (STL) and indigenous (IDG) strands in a colonized territory. Moving

through the phases, the experiences of these two groups increasingly converge, involving a gradual assimilation of identity constructions and linguistic accommodation, and a single speech community with many shared linguistic features and norms emerges (Schneider 2007: 32; for more details on the different phases, see Schneider 2007: 33–35, or Buschfeld et al. 2014, among many others, for a summary).

Despite some suggestions for minor modifications based on case studies (e.g. Buschfeld 2013, 2014; Buschfeld & Kautzsch 2017; Evans 2009, 2014; Huber 2014; Mukherjee 2007; Weston 2011) and some more fundamental criticism as regards, for example, the inclusion of both dominion Englishes and ESLs, the linear progression from phase to phase, an underrepresentation of class or status in identity formation, or its strong focus on identity constructions (Mesthrie & Bhatt 2008: 35; see also Trudgill 2004 on the latter point), the Dynamic Model has received wide acclaim and application (cf. Schneider 2014 for a detailed survey), not the least due to its flexible, modular conception and its aim to describe the emergence of a variety in its entirety.

## 2.2 Changing linguistic realities

The “established approaches” outlined in the previous section served their purposes well – Kachru’s Three Circles model captured a setting which was perceived as innovative in the 1980s and triggered the independent recognition of non-native varieties; and Schneider’s Dynamic Model broadly accounted for the evolutionary processes that produced new varieties in colonial history and the early postcolonial phase. The recent decades have seen an unprecedented dynamism, however, which has transformed the roles and forms of Englishes used globally. Boundaries and categories which just a few decades ago were established convincingly and appeared clear-cut are increasingly getting blurred – like the ones between ESL and EFL (or Outer vs. Expanding Circle) settings, as was shown convincingly for Cyprus (Buschfeld 2013) or the Netherlands (Edwards 2016) and some other cases.

The main driving force behind this has been globalization, with English as its main tool – Edwards (2016) postulates “foundation through globalization”, i.e. that globalization rather than colonization can trigger phase 1 in non-postcolonial countries. In practice, this has taken on a wide range of forms, manifestations, and facets, all of which keep disseminating English in one way or another, very often in a lingua franca function (see Seidlhofer 2011).

Formal contacts and exchanges, across national boundaries and continents, both spoken in personal encounters and written via e-mail etc., occur on a regular basis these days – in international business, in technological cooperation, in academic exchanges, in international politics and diplomacy, the arts and media, and

so on. Informal domains of life practiced transnationally as well as tourism have also adopted English as the default language; Schneider (2013) discusses the example of scuba diving. Popular movies and TV series make English, often American English, accessible in all corners of the earth. So do YouTube clips (cf. Schneider 2016c), pop music, and popular culture in general, including advertising (Lee & Moody 2012). Many of these formats (and Hip Hop music in particular) disseminate nonstandard forms of the language and all kinds of varieties, and they also produce hybrid art forms in which, for instance, elements of African American Vernacular English are being appropriated for and built into local art products (cf. Pennycook 2007; Lee 2011). In general, English has come to be available and to be used in whatever forms, as “broken” or with fraudulent intentions (Blommaert 2010), in bits and pieces, in interactions between speakers with widely varying proficiency ranges (cf. Meierkord 2012). Increasingly it can be found growing in “grassroots” forms, picked up by speakers and in contexts with very little support and resources but very high motivation (cf. Schneider 2016b; Arnaut-Karović 2016). It is used and adjusted creatively in contact situations, in code-switching and code-mixing, yielding innovative hybrid forms (see Schneider 2016a for a survey) or phenomena such as “translanguaging” (Canagarajah 2013). It is used in the new media and cyberspace (see Section 3.2 below; Mair 2013; Deumert 2014).

Seargant & Tagg (2011) have addressed many of these issues, looking at hybrid language forms in cyberspace, and argue that the concept of nation-bound varieties, which the traditional models rest on, has been outlived to some extent – it is not to be abandoned altogether but needs to be supplemented by a “post-varieties approach”. Hence, today’s changing linguistic realities are calling for new approaches, theories, and models – an observation which stands behind the present volume and this paper. This is what we now turn to.

### 2.3 Theorizing World Englishes II: Recent approaches

All these changes are multidimensional, and constitute a challenge for scholars working to understand the properties of World Englishes and the processes that have produced them. “Understanding” means theory formation, which essentially means the development of models or conceptual frameworks which highlight shared or distinguishing properties and overall processes. It is thus not a surprise that the vibrant changes noted have triggered a new round of theory development.

Like the established models, it is also possible to categorize innovative ideas into two branches, the relatively more static, classifying ones as opposed to those focusing on evolution and dynamism. Mair’s World System (2013) essentially represents the static type, reminiscent of the older “wheel”-type models by McArthur and

Görlach (see the discussion in Buschfeld & Schneider 2018). It proposes a hierarchy amongst varieties, arguing that nowadays American English, and American English only, is the “hub” and “hyper-central variety”, the most important and globally influential form of English, with other varieties, both standard and non-standard, being ranked as “super-central”, “central”, or “peripheral”, respectively. This is clearly relevant and important, but it provides a synchronic snapshot, disregarding the potential for status changes or developments (although Mair’s paper considers diffusion patterns via cyberspace). In a similar vein, Werner (2016: 131–134) offers a few overlapping circles as visual representations of relationships between different kinds of variety types, highlighting the (uncontroversial) fact that all kinds of varieties (national varieties; variety types induced by varying degrees of contact; and also varieties at different developmental phases) share a “common core” of (structural) properties. However, we believe that truly innovative theorizing should zoom in on ongoing developmental processes, and in particular it should seek explanations for the recent blurring between major variety types such as ESL and EFL, mentioned earlier.

When aiming at a joint approach to postcolonial and non-postcolonial settings, a range of forces and factors are important to consider which operate in many of the varieties, to different degrees and sometimes at different times. Many of these have already been addressed and described for PCEs in the World Englishes literature, to varying extents and mostly for individual varieties, for example issues such as language policies, language in education, attitudes to English, English and identity, and language in use. From a purely linguistic perspective structural properties and features have been identified and interpreted as effects of such forces. The very same factors often determine the spread and depth of entrenchment of the English language in non-postcolonial territories, too. As is true for the set of linguistic features found in these varieties (i.e. the structural similarities shared between postcolonial and non-postcolonial Englishes as argued and illustrated in Section 3.1), the sociopolitical forces operating in postcolonial and non-postcolonial territories as such appear to be very similar in nature, too, the only real difference being the historical factor of colonization (for a similar line of reasoning, see Buschfeld & Schneider 2018). This is, indeed, a strong predictor for second-language variety status, but not all countries with a colonial background have developed second language varieties, and especially not to the same degree of entrenchment and local restructuring. It has been shown that a colonial background does not necessarily lead to the development of fully-fledged, prototypical second-language varieties (e.g. Schneider’s case study of Tanzania [2007: 197–199] and the cases of Tswana English [Gilquin & Granger 2011] and English in Cyprus [Buschfeld 2013]). And conversely, there are countries which lack any colonial background but nevertheless have developed properties and usage frames of English which are very close to

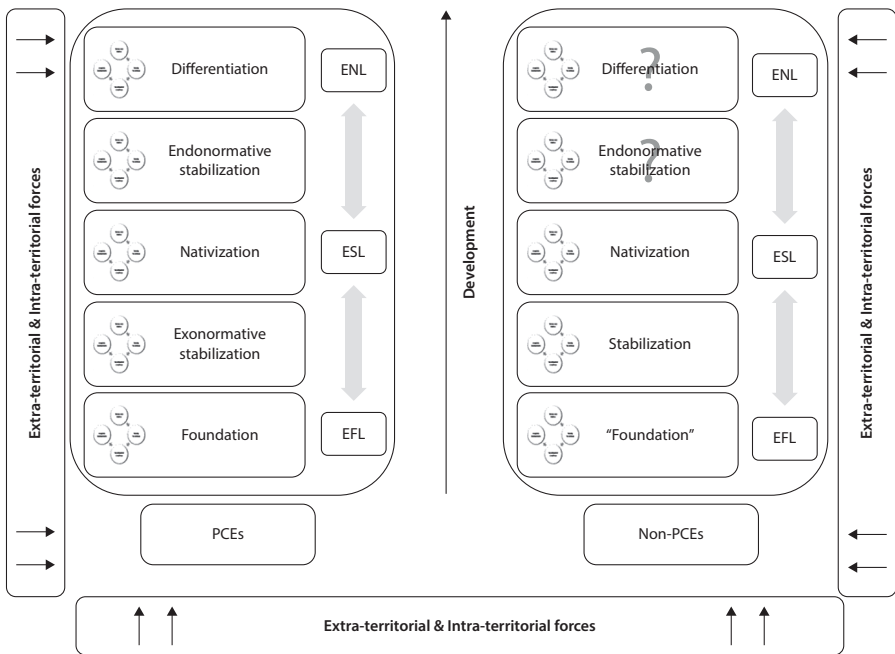
those of second-language settings, e.g. the Netherlands (Edwards 2016). This once more justifies the attempt to integrate PCEs and non-PCEs in a joint theoretical framework.

Even if Schneider's (2003, 2007) Dynamic Model is explicitly geared towards PCEs only, it accounts for many of the forces which need to be covered for a description of PCEs and non-PCEs alike, with the important issue of identity (re)writings as one of its central components. However, several pertinent processes which have entered the scene in postcolonial times and are thus not immediately connected with British colonization are missing. In particular, this concerns the effect of globalization and its related aspects (e.g. computer mediated and other means of transnational communication, language contact via the internet, mass tourism, trade etc.), and related aspects such as foreign policies and domestic political decisions on trading relationships. Attempts have been made to apply the Dynamic Model to non-postcolonial contexts as well (e.g. Buschfeld & Kautzsch 2017; Edwards 2016; see also Schneider 2014), clearly paying tribute to its scientific pertinence. However, the "colonial trappings of the model" (Edwards 2016: 187) have turned out to be insurmountable in its original version. They manifest themselves in three fundamental problems in applying the model: (1) English was transplanted to non-postcolonial regions in a completely different way than to postcolonial territories (affecting the applicability of phase 1, Foundation); (2) non-postcolonial societies lack both a settler strand and an external colonizing power exerting political, social, and linguistic influence on the colony from the outside (mainly affecting the applicability of phase 2, Exonormative Stabilization); (3) as a consequence of the missing settler strand, the type(s) of language contact and the development of identity constructions and consequently linguistic accommodation between the two strands as observed in postcolonial societies do not emerge in non-postcolonial scenarios (relevant for all phases of the model; for further details on these aspects, see Buschfeld & Kautzsch 2017). Therefore, Schneider concludes that "[i]n essence, the Dynamic Model is not really, or only to a rather limited extent, a suitable framework to describe this new kind of dynamism of global Englishes" (2014: 27–28) and instead suggests another conceptualization to account for the recent developments of global Englishes, viz. the notion of "Transnational Attraction" (Schneider 2014). It attempts to address the vibrant changes and new sociolinguistic realities identified earlier in this article (cf. Section 2.2), viewing English as an "attractor" which transcends national boundaries in orientation and impact. This concept has not been worked out in any greater detail yet. Essentially, it can be seen as an overarching notion to capture the boundless spread of English today, the facts that very many speakers and nations are keenly motivated and spend a lot of resources to acquire some proficiency in English, that English is therefore diffusing into new contexts and settings very energetically, and that this is happening largely outside of national boundaries or

perspectives and also without norms of correctness in mind (“grassroots” diffusion is a case in point; cf. Schneider 2016b). Although this framework captures a crucial factor guiding the ever-increasing global spread and local entrenchment, uses, and structural reshaping of the English language (cf. the notion of “glocalization”, which is often used to describe this janus-faced development), it only works on a rather abstract level, accounting for the general, global reasons behind its spread rather than addressing and describing specific manifestations and forces at work in individual countries. It has thus strong explanatory power but lacks the amount of granularity needed for analyses and comparisons of individual case studies.

Assuming “Transnational Attraction” as a prerequisite for the current development of further second-language varieties (especially in non-postcolonial territories) and building on Schneider’s Dynamic Model, Buschfeld and Kautzsch (2017) have developed the “Extra- and Intra-territorial Forces Model” (EIF Model). This model spells out details of the notion of “Transnational Attraction” and identifies various forces which operate within and beyond national confines and affect and strengthen the role and status of English in specific frameworks. Several of these forces have been listed, such as language policies and language attitudes, globalization and ‘acceptance’ of globalization, foreign policies, and the effect of the sociodemographic background of a country, and of course colonization and attitudes towards the colonizing power (seen in this context as one out of several forces but not the main or exclusive one). Further forces will certainly – and hopefully – be worked out and illustrated in application to different non-postcolonial but also postcolonial case studies (cf. Buschfeld & Kautzsch in prep.). In principle, Buschfeld and Kautzsch postulate that non-PCEs, too, evolve along the lines of a uniform process, when compared to each other but also when compared to PCEs (cf. Schneider’s [2007: 21] general assumption for the development of PCEs). The EIF Model assumes that various extra- and intra-territorial forces (identified above) operate on the development of different types of English at all times, i.e. from their very initial stages down to their current developmental status. Of course, differences in the manifestations, i.e. occurrence, strength, and impact, of the individual forces show between the different cases; the fundamental set of forces, however, seems largely identical. Figures 1 and 2 illustrate how the model works in theory.

The model consists of five major components, some of them new, some of them taken over from earlier conceptual suggestions and notions which have proven highly successful and influential in categorizing World Englishes: (1) It implements a joint description of PCEs and non-PCEs and introduces the notion of “extra- and intra-territorial forces” (EIF; some of these will be listed later); (2) it assumes that Transnational Attraction is the major driving force behind today’s further entrenchment of the English language and its spread beyond postcolonial contexts and even national boundaries; (3) its diachronic conception builds on Schneider’s Dynamic

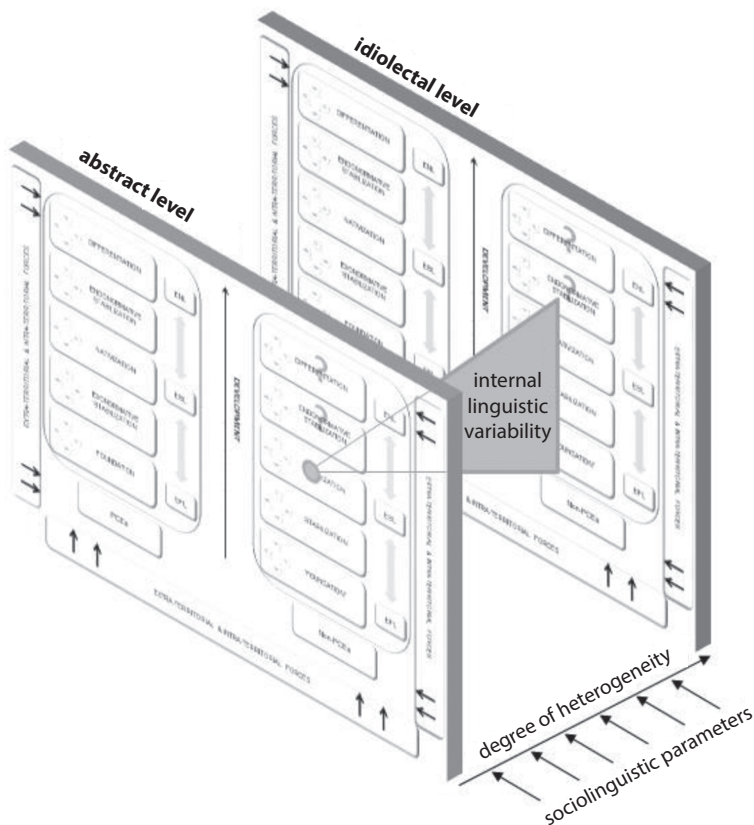


**Figure 1.** The extra- and intra-territorial forces model (adapted from Buschfeld & Kautzsch 2017)

Model, viz. the development of varieties along five phases, shaped by four parameters; (4) it draws on the long-established categorization of English into EFL, ESL, and ENL. Broadly it identifies a match between the five developmental phases and these major variety types, given that language acquisition leads to EFL speech in the initial (foundation) phase, may proceed to intensified internal functions (characteristic of ESL varieties) in the central phases and may ultimately lead to new native speakers and hence ENL varieties in the final stage (as in today's Singapore). However, the EIF Model does not picture the three categories as clear-cut and distinct from each other; instead, transitions from one category to the other are possible at all times, and developments are not to be viewed as exclusively monodirectional, as some cases show a reverse development (e.g. English in Cyprus from ESL to EFL; Buschfeld 2013). This is indicated by the bidirectional arrows between the variety types and is seen as an alternative developmental route to the otherwise linear progression of stages.<sup>1</sup> (5) The third dimension illustrated in Figure 2 accounts for the variety-internal

1. It has to be noted, however, that this is meant to depict the "prototypical" development only and as a yardstick against which the development of actual cases can be compared; not only does the model leave room for reverse developments, it also potentially allows for stages to be skipped or to be taken in some other order.





**Figure 2.** Depicting internal linguistic variability in the EIF Model

heterogeneity found in almost every regionally defined type of English to a greater or lesser extent, determined by sociolinguistic and situational factors such as status, formality of a situation, speakers' proficiency levels, etc. Locating varieties at one point of the continuum, as is done in our case studies in Section 3.1, always involves a high degree of abstraction and can only be seen as a very rough approximation of the status and forms of English in a specific region. To fully capture the sociolinguistic realities as well as the status, functions, and uses of English requires a higher level of granularity, i.e. zooming in to possible differences between speaker groups and – in its most detailed form – into the idiolects of individual speakers. Such heterogeneity can be motivated by sociolinguistic variables such as age (e.g. the case of Cyprus), ethnicity (e.g. the differences between White, Black, Indian, etc., South African Englishes), social status (e.g. the often observed differences between white collar and blue collar occupations), gender (as has often been shown in sociolinguistic research), etc. The branching-off axes in Figure 2 attempt to visualize this



heterogeneity, with the starting node constituting the highest level of abstraction and the vertical plane in the back representing the highest level of detail, depicting language use of the individual. Within the triangle, the different sub-varieties can be located closer towards the node or closer towards the outside plane, depending on the level of granularity one aims at. The more heterogeneous a variety, the more widely spread the external points of the fan are at the level of the individual.

### 3. Innovative approaches applied

#### 3.1 Bridging the postcolonial/non-postcolonial divide: Applied perspectives from Europe, Namibia, and South-East Asia

As illustrated in Section 2.2, it has repeatedly been shown that the static ESL-EFL distinction can no longer be maintained. The conventional mapping of older categorizations (ESL, EFL and Outer and Expanding Circle alike) onto a classification of countries into postcolonial and non-postcolonial territories is too short-sighted. We begin with an overview of some relevant sociolinguistic parameters in South-East Asia, followed by a comparison of two countries traditionally assigned to the same category (EFL, as in the cases of Germany and Namibia) yet being very different in their sociolinguistic developments and current linguistic setups. In a next step, we look into two cases which, according to traditional practices, would fall within different categories (ESL for the postcolonial case of Cyprus and EFL in the case of Greece) but turn out surprisingly similar in some respects. Ultimately we report on findings from a recent investigation of the use and status of English in Bosnia and Herzegovina, which adds yet another perspective to the various ways of entrenchment the English language has taken. In the end, Section 3.1.5. will point out aspects in which recent models overcome shortcomings of earlier ones.

##### 3.1.1 *South-East Asia*

As noted elsewhere (Schneider 2014: 22–23), the adoption of English as its “sole working language” (Chapter 34 of the association’s 2009 Charter) by ASEAN, the Association of South-East Asian Nations, constitutes a remarkable example of the attractiveness of English and the forces which contribute to its diffusion. Malay, spoken in four out of the ten member states, would have been an obvious alternative (Kirkpatrick 2010: 12–14), but it was English that was selected. In all ASEAN countries English is now the first foreign language in the education system and “a compulsory subject in primary school in each country except Indonesia” (Kirkpatrick 2010: xi). According to statements by “key ASEAN figures”, the decision to adopt English came “automatically”, without any explicit regulation, and was simply

taken “for granted”, “quietly accepted by the founding member states” (Kirkpatrick 2010: 9). The reasons for this choice include the role of English as “the language of modernization and advancement [...] and [...] of democratically supported power” (Kirkpatrick 2010: 12). The advantages of adopting English in this role, spelled out by “ASEAN bureaucrats”, sound fairly pragmatic: “It saves enormously on costs and labour, [...] it allows easy dialogue internationally and it facilitates technology and knowledge transfer” (Kirkpatrick 2010: 14). Motivations for language teaching and learning across South-East Asia are “entirely instrumental” without any “integrative or humanistic motivations” (Kirkpatrick 2010: 18). In addition to the political importance of English, the growing tourism industry across the region is another factor that promotes English, since English is the default language in tourism interactions and knowledge of it promises well-paid jobs to many locals.

### 3.1.2 *Namibia and Germany*

Traditionally, these two countries have been assigned Expanding Circle or EFL status, since neither has experienced British (or American) colonization. The linguistic realities there and the accompanying status of English and its use, however, could not be more different. With a population of about 80 million, Germany is the most populous country in the European Union; it is largely mono-ethnic<sup>2</sup> and its only official language is German (CIA 2017a). By contrast, Namibia has a population of about 2 million, is one of the least populated countries of the world (United Nations Department of Economic and Social Affairs 2015: 56) and displays a high level of ethnic diversity,<sup>3</sup> which also shows in the presence of at least 27 languages from the Bantu, Khoisan and Indo-European language families (CIA 2017b; Buschfeld & Kautzsch 2014: 122). Although Oshiwambo is spoken natively by a vast majority and Afrikaans was used as a lingua franca during South African rule (1915–1990), English was made the only official language with Namibia’s independence in 1990. As a consequence, English is now used in many domains other than the family (i.e., in the education system, the media, court, administration, parliament, literature) and is increasingly becoming a means for interethnic communication, especially in the younger generation (for details of attitudes and use surveys see Buschfeld & Kautzsch 2014 and Kautzsch & Schröder 2016). The use of English by Namibians of different ethnicities and thus first languages, who mostly also have full command of Afrikaans as a second language, leads to a plethora of contact scenarios, which in turn is likely to give rise to ethnic differences in the use of English (cf. Buschfeld

---

2. German: 91.5%, Turkish: 2.4%, Other: 6.1%, mostly Greek, Italian, Polish, Russian, Serbo-Croatian, and Spanish.

3. Ovambo 50%, Kavango 9%, Herero 7%, Damara 7%, Nama 5%, Caprivian 4%, Bushmen 3%, Basters 2%, Tswana 0.5%, White 6% (mostly Afrikaners; some Germans), Mixed 6.5%.

& Kautzsch 2014 for a tentative list of features on different linguistic levels; cf. Kautzsch & Schröder 2016 for vowel realization).

By contrast, English in Germany is mostly used for international business communication (Ammon 2006: 30; cf. Hilgendorf 2007) and is generally taught through formal instruction. In the education sector, however, a growing trend for bilingual education is clearly visible (cf. Kautzsch 2014): English teaching begins in primary school and many university courses have introduced English as a medium of instruction (Kautzsch 2014: 208–210; cf. Ammon 2006; Hilgendorf 2005; Hilgendorf & Ehrling 2006). In some other domains English is also making inroads: some courts allow cases to be negotiated in English, if one of the parties does not speak German (*Frankfurter Allgemeine Zeitung*, 2010-01-28); some newspapers partly offer content in English; advertising makes ample use of English, although most claims are not understood by the wider public (cf., e.g., Samland 2011); youth language employs many lexical items from American and British popular culture; people working in the service industries require some basic skills in English for communication with tourists. In this context, the question has been raised “whether the [English] language [in Germany] is not better characterized as a second or additional language instead of as a foreign code” (Hilgendorf 2005: 64). But this might not fully meet realities, since English in Germany is largely absent from the domains of politics, media (films and TV series are dubbed), administration, or literature and is thus much less prominent in daily life than, for example, in Namibia.

### 3.1.3 *Cyprus and Greece*

The need for remodelling the World Englishes paradigm – in particular rethinking the idea that a clear-cut distinction can be drawn between EFL and ESL – shows particularly nicely when comparing Englishes and the sociolinguistic ecologies of countries in which they have developed in similar language contact settings, i.e. with a shared L1, but which differ in terms of whether they have experienced British colonization or not. Unfortunately, such constellations are extremely rare, but Cyprus and Greece are two such cases in point (see Percillier 2016 and his analysis of English in Singapore, Malaysia, and Indonesia for another such constellation). In both the Greek part of Cyprus<sup>4</sup> and mainland Greece, Greek is spoken as the first language by the majority of the population<sup>5</sup> and English is an important additional

---

4. The study was limited to the Greek part of the island due to huge differences in the sociolinguistic realities between the two parts, which would have required two independent investigations.

5. Note that differences exist between the dialectal variants of Greek spoken in the two speech communities, viz. Standard Modern Greek and Cypriot Greek. These, however, are not fundamentally relevant when it comes to the question of feature nativization in the two countries and do not affect the comparison at hand.

language.<sup>6</sup> Cyprus is a former colony of the British Empire, which exerted authority over the country between 1878–1960, while Greece is a country without British colonial background. According to a survey of proficiency levels in English (European Commission 2012: 21, based on self-reports of people stating they can hold a conversation in English), language proficiency in English is more widespread in Cyprus (73%) than in Greece (51%). Buschfeld (2013, in prep.) has shown that, while in Cyprus English is widely used by all generations, in Greece it is mainly the young generation who is in good command of English. What is more, the situation in Cyprus is characterized by the existence of local and, at least to some extent, nativized characteristics of English on all levels of linguistic description, especially in the older generation (Buschfeld 2013). When looking into the linguistic characteristics of English in Greece (GrE), many similarities are shared between the two Englishes. GrE generally employs the same linguistic features as English in Cyprus (CyE), e.g. the use of zero subjects and objects, uses of definite and indefinite articles which diverge from the traditional standard varieties, intransitive usage of the verb *like*, and the time reference pattern *before X days/weeks/months/years*). However, an exemplary quantification of two of the high-frequency morphosyntactic features of CyE in the mainland Greek data, i.e. the use of *will*-future and simple present to express hypothetical context and the use of the intensifier *too + much* instead of *very + much*, reveals clear and partially even statistically significant differences: CyE makes greater use of the above local features than does GrE (for details on the methodological procedure and results, see Buschfeld *fc.*).

The sociolinguistic situation in both countries is different, yet not distinctive enough to justify an assignment to two completely different categories, such as ESL for the case of Cyprus and EFL for the case of Greece, as would be the traditional practice on the basis of historical criteria. Considering the significant historico-political difference between the countries, one would certainly expect greater disparity in their linguistic situation and use of English. Both countries embrace English as the language of globalization and modernization, as a means of communication with considerable cohorts of tourists, in the education sector, and in the new media. Cyprus is certainly still “ahead” in terms of intranational usage frequencies of English, mainly due to traces of the colonial legacy still to be found on the island. At the same time, however, the intranational use of English has increasingly declined since decolonization and the 1974 division of the island into a Greek and Turkish part (for details and reasons, see Buschfeld 2013).

---

6. The “additional” here is used as a neutral means of describing the fact that English is spoken in addition to Greek, not suggesting anything about its status.

### 3.1.4 *Bosnia and Herzegovina*

Bosnia and Herzegovina declared its independence in 1992 and was recognized as a nation in 1995 as an outcome of the Dayton Peace Accords which ended the war of succession of the former Socialist Federal Republic of Yugoslavia. The changing roles and status of English in this country, richly documented in Buckingham (2016), can be viewed as a model case of what has been happening to English in non-postcolonial countries in the recent past. By documenting various aspects of English usage and English teaching conditions Buckingham's book offers a comprehensive survey of a rapid linguistic and cultural transition.<sup>7</sup>

English is the first foreign language of all children today, taught obligatorily from grade 3 (age 8) onwards. Remarkably, however, this was not at all the case as recently as some twenty years ago – in the early 1990s competence in English was not widespread at all. In Yugoslavia, English was only marginally known, as a fun language of pop culture in the 1960s and a rare language of business and professional contacts in the 1980s. This changed drastically during the war years (1992–1995). International military forces (“United Nations Protection Forces”, deployed since 1992), non-governmental organizations (NGOs) and other international organizations came to bring humanitarian aid and establish peace, and all of them needed local support staff and interpreter services. Irrespective of their origins, English was the lingua franca among all international organizations, and all of a sudden there was a huge demand for local interpreters with English language skills. The salaries offered by these institutions were many times higher than the local ones (for teachers, for instance), so English-speaking skills guaranteed an income that sustained an entire family during the years of crisis and warfare, often to young people with no formal qualifications. The consequences of this situation were drastic and remarkable: under life circumstances where the provision of basic utilities was most difficult many individuals “demonstrated a remarkable capacity to teach themselves the language of survival”, largely without books or other teaching aids but driven by an immensely high motivation (Buckingham 2016: 2). The same applies to the conditions of teaching and acquiring English in the education system: with most qualified teachers having left their teaching posts for better-paid jobs as interpreters and, at the same time, the demand for the language growing intensely, the teaching of English continued and even gradually expanded under extremely difficult conditions, lacking trained staff (practically no formal qualifications were required for a role as English instructor), textbooks, and also suitable classrooms – but nevertheless some knowledge of English kept growing. This process

---

7. The following description draws heavily on Buckingham (2016), and notably also on Imamović & Džanić (2016).

has continued to the present day, with conditions slowly improving in the post-war years. Today, academically qualified staff is trained at several newly founded English departments at universities across the country. Bosnia-Herzegovina is an EU candidate state and is making progress towards admission to the NATO, and both of these goals of course entail a persistently high demand for proficiency in English and translation services involving English. Buckingham (2016) documents a wide range of facets and contexts which characterize and promote the use of English in the country, including language policy, teacher training, textbook production, publishing practices, translation and interpreting needs, language contact effects, and personal narratives of the importance of language usage.

### 3.1.5 *Discussion: Theoretical implications and applications*

The above case studies clearly support the general theoretical assumption that EFL and ESL cannot be strictly separated from each other but should be considered two poles on a continuum on which varieties are more or less ESL or EFL in nature (cf. Section 2.2). In addition to that, the individual analyses and comparisons reveal some related and more detailed insights, which all corroborate our overall argument:

1. The case of Namibia (as have others before; cf. Section 2.2) again shows that second-language varieties can emerge even in countries lacking a colonial background and where therefore the historical roots of English are comparatively weak. This clearly indicates that colonialism is not the only decisive force behind such developments and is in no way mandatory.
2. The cases of Namibia and ASEAN demonstrate how language-political decisions can drastically change and determine the status of English in a region, resulting in strong entrenchment and potentially the development of second-language varieties.
3. The case of Bosnia and Herzegovina illustrates how the presence of international organizations and military (UNPROFOR) can strongly influence the role and spread of English.
4. The comparison of Namibia and Germany shows that countries which would traditionally have been assigned to the same category (EFL or Expanding Circle) can be very different in their sociolinguistic setups and developments: Therefore, the EFL category has to be considered very heterogeneous in itself (cf. Buschfeld *fc.*; Gilquin & Granger 2011: 74).
5. In a similar way, the comparison of Cyprus and Greece has revealed that territories which would traditionally be assigned to different categories (ESL vs. EFL or Outer vs. Expanding Circle) can be more similar in their sociolinguistic setup than a strict differentiation between the two categories would suggest.

6. In line with finding 5, the cases of Cyprus and Greece corroborate earlier observations that learner Englishes and second-language varieties (or rather non-postcolonial and postcolonial Englishes, as this denomination is historically straightforward and can be assigned uncontroversially) often share the same linguistic features. This applies especially when their speakers have the same substrate language as L1 and thus acquire and use the English language in similar contact settings. Differences in feature use are often just of a quantitative nature.

On the basis of these observations, we argue that only a model geared towards an integrated analysis of postcolonial and non-postcolonial Englishes can adequately depict the current linguistic realities worldwide. We need an approach which captures the diachronic aspects of such developments, since often Englishes have experienced sudden changes in status and direction of development (e.g. from ESL to EFL in the case of Cyprus, from EFL to ESL in the case of Namibia), mostly on the basis of quasi unpredictable socio-political events (cf. Schneider's 2007 notion of 'Event X' in the development of PCEs). Last but not least, all case studies demonstrate that there are different external and internal factors at work in the spread and entrenchment of English in specific regions. These forces can be of universal nature, operating on many countries and Englishes in very similar ways. They come in the guise of what can be subsumed under "globalization" and include, for example, international communication in trade, computer-mediated communication, language contact scenarios provided through the internet (cf. Section 3.2), and also mass tourism, though this latter aspect slightly varies from country to country. On the other hand, some forces are very specific in nature and have operated on single countries (or a very restricted set) only, e.g. Namibia's decision to make English its sole official language in 1990, or the influence of international organizations and the military on the significance and spread of English in Bosnia and Herzegovina.

When turning towards the question how these findings are best captured in theoretical terms, it becomes clear that the synchronic models, be they of older (e.g. Kachru 1985) or more recent (Mair 2013) origin, cannot satisfy all the "requirements" identified above. Despite its diachronic orientation, Schneider's Dynamic Model cannot adequately accommodate non-PCEs either. What we find in all cases illustrated above is "Transnational Attraction" at work, but this is too broad an abstraction to account for the developments of individual cases. A whole set of other factors exists – global and individual, and extra- and intra-territorial in nature – that have influenced the extent of the spread and entrenchment as well as the specific forms the English language has taken in different parts of the world. Next to the forces illustrated above, and by way of some more concrete examples, in the context of ASEAN, the extra-territorial force of language policy, with modernization



and integration as its main goals, meeting the intra-territorial force of individual advancement, majorly promoted the entrenchment of the English language. For Cyprus, it was the extra-territorial force of the Turkish invasion which led to an ad hoc change in the sociolinguistic realities on the island (English lost its function as a neutral link language between Turkish and Greek Cypriots) and the intra-territorial force of resistance (the Greek population strongly adhered to their Cypriot Greek identity and language; for further details, see Buschfeld 2013) which hindered a strong and steady entrenchment of the English language.

We believe that the EIF Model can best capture the worldwide developments and shapes the English language is going through, since it takes account of post-colonial and non-postcolonial Englishes alike, operates on the basis of different extra-territorial and intra-territorial forces, employs a diachronic perspective, and envisages intravarietal heterogeneity. It has to be tested, however, whether the framework still holds if we take on board Englishes which develop beyond the nation state. We will make some first steps into that direction and discuss how in theory the EIF model can also integrate what Friedrich and Diniz de Figueiredo (2016) call “digital Englishes”, i.e. Englishes as used on the internet.

### 3.2 Digital Englishes

The internet and computer-mediated communication in its various ways and channels, including e-mail, chat groups, blogging, instant messaging, posting in social media, creative writing, etc., have significantly modified and promoted the global ways of using English (Crystal 2006). One of the older usage forms along these lines, the language of short messages or texting, has generated innovative abbreviations (due to space constraints) and other genre-specific conventions (Crystal 2008), as has Twitter. Ultimately, the introduction of powerful and affordable personal computers in the 1980s together with the commodification of the internet and the introduction of the World Wide Web around the 1990s made e-mail and other unprecedented means of effective and easy electronic communication available to individuals on a global scale. This process was further enhanced by the introduction of the smartphone, which reached mass accessibility around 2010 and made internet access available at nearly all times and places. The late 1990s and especially 2000s brought forward a number of widely used platforms and websites, used for communication, social networking, discussion, creative writing, and blogging activities (e.g. [www.FanFiction.net](http://www.FanFiction.net) in 1998, Skype in 2003, Facebook in 2004, Tumblr in 2007, WhatsApp in 2009, to mention just a few). All of these resources have reached global audiences and have been strongly associated with, and have promoted access to, English.



In these contexts, English is used by native and non-native speakers alike. In the latter case it often co-occurs and interacts with a user's native language, a process which has produced unprecedented forms of language contact and code-switching and mixing, still to be fully understood and described by linguists. Seargeant and Tagg (2011) document and discuss the creativity of computer-mediated discourse, involving liberal linguistic mixing and forms of English which are far from traditional standard norms and which often transcend national boundaries, and hence they advocate a "post-varieties approach". In the following, we briefly present some examples of such innovative language contact and usage situations, including the use of English by Germans and Singaporeans on Facebook, linguistic strategies and their consequences in the domain of fanfiction writing, and language choice and use in multiplayer online games. None of these examples will be analyzed in detail; they just serve as examples of the different shapes electronic Englishes can take and as a basis for our discussion of what all this implies for theorizing World Englishes (in Section 3.2.4 below).

### 3.2.1 *Facebook communication*

Social media constitute a wealth of platforms and virtual communities to exchange information, ideas, and opinions, with English as the language used by far most frequently. According to Web Technology Surveys (2017), 52.1% of all websites use English as their major language, with the second strongest language, Russian, representing only 6.5% of the web contents. Next to the very common case in which English is used as a lingua franca in conversations involving speakers who do not share a common language, both native and non-native speakers alike, it is used in many different ways and for a variety of different purposes.

Example (1) constitutes the reaction of a German 31-year-old female reacting to a friend's Facebook post about a sports achievement, which itself is in German:

- (1) Male, 28: Juchu! Ich hab meinen ersten Halbmarathon geschafft!  
 Female, 31: Heartly luckwish! Zis is ä real glossätschiefment, my holy dear singingclub!

Such examples are not uncommonly found and exhibit in playful ways local characteristics from their native languages (here orthographic representation of a German accent in English and a literal translation of a German idiom). Other examples involve both intra- as well as intersentential code-switching, as Examples (1) to (8) show, all produced by highly educated German speakers of English:

- (2) Female, 35: Triple love! The song is really one of the best songs ever (und das Video just super cool) :) :) And Happy Birthday to us, indeed, it's been a great year! :) Dann machen wir doch mal weiter und laden zu einer Runde "memorable songs" ein [...] ["Then let's carry on and invite another round of ..."]

- (3) Female, 35: Well, Nirvana halt, just kult!
- (4) Female, 35: [...] Und das Gitarrensolo-Thema finde ich suuuuper, let's keep that in mind for next week!
- (5) Female, 35 [about a posted song]: Ich leg einen nach ['I'll add one more'], gimme a minute.
- (6) Female, 35: Oh yesssss! (More comments later tonight, du hältst mich vom schweren Arbeiten ab ['you're keeping me off from hard work']).
- (7) Male, 28: Mensch, da kriegt man so viele schöne Geburtstagsgrüße [Oh man, you get so many nice birthday greetings'], but this knocks it out of the park [...]
- (8) Male, 47: Congrats, du Wahnsinniger ['you crazy guy']!!! WOW!!

Example (9), a conversation between two young Singaporean women, is another example of how informal local usage (here Singaporean English) finds expression in Facebook communication and illustrates another typical characteristic of on-line communication, viz. the use of text messages shorthands such as rebuses and grammograms:

- (9) Female, 36: nice? no frog porridge this time? next time must jio me
- Female, 38: Hahhaah u not in your hood mah. Yes. Ate up already 🍜
- Female, 36: Wah u 2 ate so much. U had the kung bao spicy one or spring onion?
- Female, 38: Just 1 frog kungpao

### 3.2.2 *Fanfiction writing*<sup>8</sup>

Another interesting, net-based use of the English language, clearly underresearched from a linguistic perspective, can be found in fanfiction writing. Fans transform their favorite stories by using their respective fictional universes (or simpler: characters) for their own writing which they then post to online fanfiction communities (e.g. FanFiction.net). Most of these texts are written in English, again by native and non-native speakers alike. The specific character of the “‘strange’ fannish jargon” (Herzog 2014: 145) has been summarized as employing “unclear acronyms and lots of punctuation” (Hellekson 2009: 113), which, for the outsider, is indeed strange and partly indecipherable, as Example (10) from a fanfiction on the Twilight mythology illustrates:

---

8. We would like to thank Alexandra Herzog for offering feedback on an earlier version of this section.

- (10) She's a single mom who lives & works w/her dad. It's safe, secure. But sometimes security isn't all that it's cracked up to be and you need to open your heart a little. He should probably do the same. Fluff/UST/AH Alternating ExBPOV Rated M. (22blue, "Dragonflies.")

Mostly based on English, fanfiction writing makes use of different strategies, thus rendering it partly inaccessible to an outside reader. According to Herzog (2014: 146), *fanspeak* "has an explicit double purpose that transcends the (rather) superficial task of communicating information about the stories to their readers" in that it obscures the information for outsiders, i.e. anyone who is not a fanfiction writer. On the other hand, it constitutes and symbolizes membership in their respective communities and is employed as a demonstration of authority by fan authors in showing expertise and distancing themselves from non-fans and the producers of the original texts. To achieve this double aim, *fanspeak* employs its own complex terminology to provide information on the setting and atmosphere of the story, major plotlines and structural characteristics, its characters as well as the relationships between them. For example, in the paratext to the story "Dragonflies" (see above), "Fluff/UST/AH Alternating ExBPOV Rated M" provides the information that the text constitutes

a rather lighthearted romance ("Fluff") that features both protagonists', i.e. Edward's and Bella's, points of view ("ExBPOV") and has any sexual tension between the characters remain unresolved, which also entails that the story does not contain any too graphic description of sexual acts ("UST"; "M"); moreover, all characters appear as humans ("AH") instead of being, as they would be according to Twilight mythology, vampires or werewolves. (Herzog 2014: 146)

In terms of which linguistic strategies fans employ to create their specific jargon, Herzog lists six mechanisms (in descending frequency):

1. Abbreviations in form of acronyms and initialisms, e.g. *E&B*, announcing the pairing of characters (here: Edward and Bella) or *OOC* indicating the characters as "out of character" when compared to their equivalents in the original text.
2. Blends, e.g. *Destiel* (to indicate that a story features a romantic relationship between Dean and Castiel, two characters in the series *Supernatural*), giving information about the *OTP* ("one true pairing"), i.e. the main characters of the story and their relationship towards each other.
3. Semantic change in which the fannish meaning has little or nothing in common with the original, non-fannish one, e.g. *lemon* indicating "extensive sex scenes described in graphic detail" (Herzog 2014: 153); interestingly *lemon* is derived from a Japanese slang term meaning 'sexy', so this example furthermore involves the linguistic process of borrowing and thus constitutes an example of language contact (cf. mechanism 5 below).

4. Neologisms (opaque and transparent ones, the latter being created according to regular processes of English word formation), e.g. *drabble* ('an exactly 100 words long story'), *ficlet* ('a short fanfiction'), *podfic* ('a fanfiction story one can listen to').
5. Borrowings (in their original meaning or with redefined ones), e.g. *shounen ai* ('boys' love'; Japanese slang).
6. Graphic writing (term coined by Herzog 2014: 154), e.g. the exclamation mark in *soulless!Sam*, with the ! having a fan-fiction specific meaning: the exclamation mark here signifies a specific construction of the character of *Supernatural's* Sam Winchester that informs the fannish reader about the writer's approach to the story. (all examples from Herzog 2014: 145–163)

### 3.2.3 *Multiplayer online games*

"Virtual worlds" (Crystal 2006: 12), multiplayer online games in which a fictitious adventure world is created and speakers participate via virtual identities, constitute a specific manifestation of internet language. In large, often transnational player communities the default language of choice is, not surprisingly, English. Game contexts impose both technological and social constraints. For instance, how are the players' utterances produced and perceived – on keyboard and screen or as sound and talk? Of course, they need to master game-specific concepts and vocabulary in order to cooperate (for example as members of a team or virtual army); they need to express themselves clearly and reasonably concisely; and there is clearly some room for context-specific accommodation, the emergence of shared linguistic habits and conventions within a player community. Linguistic indexicality may be deliberately employed, depending on the nature of the virtual scenario and its characters, settings and events created in a game. Game worlds with multiple players, often counting to hundreds or even thousands, produce unique "possibilities for creative, idiosyncratic, imaginative expression and the likelihood that this situation will produce a distinctive linguistic variety", "a fresh strand of linguistic creativity" (Crystal 2006: 186).

Weiß (2016) investigates linguistic usage in one such multiplayer video game, *The Witcher 3: Wild Hunt*, which was released in 2015 and sold nearly 10 million copies. It models a medieval fantasy world with castles, towns, and hamlets and many characters from all walks of life. Technologically, dialogues are realized by a very large set of predefined, pre-recorded utterances, which the player can choose from and receives as responses. Interestingly enough, the characters' performance in terms of their vowel and consonant usage and nonstandard grammar features is clearly and transparently sociolinguistically stratified, representing conventionalized stereotypical associations of specific regional and social dialect features with

speaker characteristics. For example, glottalization and *h*-dropping, features commonly associated with lower-class British dialects, are generally found in the speech of rural characters in the game.

### 3.2.4 *Discussion: Theoretical implications and applications*

When thinking about the theoretical implications these newly evolving forms of English bring with them, it is immediately clear that these pose a challenge for World Englishes theorizing and require what Seargeant and Tagg (2011) have labeled a “post-varieties approach”. As is obvious from the above examples, digital Englishes are very different from our traditional, nation-state-based varieties, in that they transcend nation state boundaries, both in their conditions of usage (i.e. involving speakers of different regional backgrounds) and their linguistic characteristics (involving different varieties and proficiency levels). Yet, there are also some similarities between these two types, both in social and in linguistic terms. Fanspeak, for example, deliberately establishes fannish in-group identity (cf. Herzog 2014), reminiscent of the role of identity conceptions in the nation-building processes found in postcolonial societies (cf. Schneider 2007). On the other hand, the use of fannish jargon is not homogeneous throughout the community. Even though fanspeak shares a common core of features, differences show between different fandoms and fanfiction communities (Herzog 2014: 151), similar to what we find in terms of variety-internal variation.

These similarities would probably call for an integrated model of World Englishes, which at the same time accounts for the differences between nation state-based and digital Englishes. While the internet is increasingly becoming multilingual,<sup>9</sup> the predominance of English, especially for transnational encounters and activities, illustrates its attraction and promotes its usage. This is where Schneider’s notion of Transnational Attraction shows in much the same way as in the emergence of non-postcolonial but still nation-state-based Englishes. In the electronic domain, the forces at work and the motivations for using English in, for example, multiplayer online games are also both extra-territorial (e.g. globalization in the form of long-distance personal communication, participation in an international player community, sociolinguistic stereotypes of who uses what type of language working on the creation of specific speech snippets) and intra-territorial (e.g. individual enjoyment and linguistic choices of the individual players), at least on an abstract level. However, this would require a reformulation into external and internal forces to get rid of the nation state-oriented perspective encoded in the term “territorial” (or a metaphorical understanding of the notion of “territory”). In

---

9. Multiplayer video games like *The Witcher* are often also released in dubbed versions at least for major languages.

fanfiction writing, we also find external and internal forces at work in the creation of a new variety (or at least jargon). These, for example, manifest in the wish to create a specific in-group identity and to set oneself apart from the original texts and traditional models of writing (i.e. identity construction and group pressure as internal factor), and again in globalization as an external factor. In both cases, the acceptance of globalization, a clearly internal force in the development of nation state-based varieties, here turns into an external force as individuals are dependent on decisions on internet accessibility made by the governments of each country the user lives in.

As the examples show, approaching digital Englishes along the lines of external and internal forces is generally conceivable, as is the idea that an electronic form of English might develop its own “nativized” features over time. It is far from impossible that digital Englishes can follow similar developmental routes like nation-based varieties, i.e. a development of special varieties under the influence of external and internal factors working on the users and their motivation and way of acting in the specific community of practice. However, it is not yet possible to state whether these Englishes could fully and unequivocally be captured by the EIF framework; such a claim requires more detailed analyses of these Englishes which also take into consideration their diachronic aspects, i.e. their development and the emergence of special linguistic characteristics over time. We hope that this section will strengthen research into these newly emerging forms of English, including both synchronic and diachronic aspects of their existence.

#### 4. Conclusion

In our attempt to cover and understand what has been happening to the English language in the recent past, essentially in the 21st century, we have covered quite some ground. This is a most vibrant ongoing process, most interesting to witness and highly complex in its various facets. Clearly, there is an intrinsic relationship between what is happening in the real world, i.e. the unsteered, dynamic evolutionary processes of English, on the one hand, and how linguists attempt to account for this, i.e. activities at model-making and theorizing, on the other.

We have seen that English has been moving into novel domains, far beyond the remains of its origins and colonial diffusion and also transcending a rather neutral status as learners’ foreign language. It has been argued (and to some extent rightly so) that national boundaries are no longer as determinative for linguistic developments as they once used to be and that it is time for a post-national perspective, yet many aspects of this expansion can still be tied to specific regions. We have looked into examples from regions as widely apart from each other as South-East Asia,

Namibia and Germany, Cyprus and Greece, and Bosnia and Herzegovina. In all these locations we have found English growing, expanding, and becoming more important (and partly more distinctive in its formal properties), albeit under varying circumstances and in different ways. The expansion has a social and sociolinguistic dimension: English is acquired in various ways also outside its traditional domain of formal education, and it diffuses across cultural contexts in all kinds of ways and forms (such as grassroots diffusion, local appropriations of nonstandard language in Hip Hop, or lingua franca usage with widely varying proficiency levels and irrespective of norm concerns). And it has a “medium” dimension, with cyberspace, social media and all forms of computer-mediated communication contributing substantially to English, which occurs in various forms and in innovative contexts.

We have considered and compared a range of recent theoretical frameworks and their power to explain these ongoing developments. Essentially, we have found that static, classificatory models, while useful and insightful in many respects, fail to account for the vibrancy of recent processes and need to be supplemented by frameworks which allow an understanding of processes and changes. The Dynamic Model is useful to account for the colonial diffusion of English, the growth of new postcolonial varieties, and some early postcolonial developments. It lacks the tools to explain these recent developments, especially in non-postcolonial contexts, and thus it needs to be supplemented by other lines of thinking. The notion of “Transnational Attraction” is appealing and powerful but rather generic, not suitable for explaining details and different facets. The EIF model, and in particular the more advanced version suggested above, which factors in internal heterogeneity of varieties, seems most suitable and promising to build an overarching framework. It will need to be developed further, with new forces to be identified and built in.

Let us end with a word of caveat. Nobody knows what the future will bring, of course. But some political developments of the very recent past (the British Brexit vote; the election and first moves of the new American President; and the growth of nationalist and authoritarian governments in a few countries in and near Europe) raise the question of whether the age of globalization, and with it possibly this strong transnational role of English, are gradually coming to a close. We do not believe so, since these are isolated, if strongly visible, developments, and in contrast to them in all countries today we find transnationally oriented, internationally experienced, educated and mostly young people whose world view is global and who will hopefully shape the future. But it remains to be seen what will happen in global politics, and to the English language, and, consequently, to linguists’ theorizing on developments of Englishes. The topic remains vibrant and exciting.



## References

- Ammon, U. 2006. The status and function of English in Germany. *Revista Canaria de Estudios Ingleses*. 53: 27–33.
- Arnaut-Karović, K. 2016. ‘English for Survival Purposes’ and the predicament of self-taught language teachers. In *The Status of English in Bosnia and Herzegovina*, L. Buckingham (ed.), 31–48. Bristol: Multilingual Matters.
- Bailey, R. W. & Görlach, M. (eds). 1982. *English as a World Language*. Ann Arbor MI: University of Michigan Press.
- Blommaert, J. 2010. *The Sociolinguistics of Globalization*. Cambridge: CUP.  
<https://doi.org/10.1017/CBO9780511845307>
- Bruthiaux, P. 2003. Squaring the circles: Issues in modeling English worldwide. *International Journal of Applied Linguistics* 13(2): 159–178. <https://doi.org/10.1111/1473-4192.00042>
- Buckingham, L. (ed.). 2016. *The Status of English in Bosnia and Herzegovina*. Bristol: Multilingual Matters. <https://doi.org/10.21832/9781783095971>
- Buschfeld, S. 2013. *English in Cyprus or Cyprus English? An Empirical Investigation of Variety Status*[Varieties of English Around the World G46]. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/veaw.g46>
- Buschfeld, S. 2014. English in Cyprus and Namibia: A critical approach to taxonomies and models of World Englishes and Second Language Acquisition research. In Buschfeld et al. (eds), 181–202.
- Buschfeld, S. Forthcoming. Towards an integrative approach to postcolonial and non-postcolonial Englishes: What Cyprus and Greece can tell us.
- Buschfeld, S., Hoffmann, T., Huber, M. & Kautzsch, A. 2014. The evolution of Englishes: The Dynamic Model and beyond. In Buschfeld et al. (eds), 1–17.
- Buschfeld, S., Hoffmann, T., Huber, M. & Kautzsch, A. (eds). 2014. *The Evolution of Englishes: The Dynamic Model and Beyond* [Varieties of English Around the World G49]. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g49>
- Buschfeld, S. & Kautzsch, A. 2014. English in Namibia: A first approach. *English World-Wide* 35(2): 121–160. <https://doi.org/10.1075/eww.35.2.01bus>
- Buschfeld, S. & Kautzsch, A. 2017. Towards an integrated approach to postcolonial and non-postcolonial Englishes. *World Englishes* 36(1): 104–126. <https://doi.org/10.1111/weng.12203>
- Buschfeld, S. & Kautzsch, A. (eds). In preparation. *Modelling Current Linguistic Realities of English World-Wide: The Extra- and Intra-territorial Forces Model Put to the Test*. Edinburgh: Edinburgh University Press.
- Buschfeld, S. & Schneider, E. W. 2018. World Englishes: Postcolonial Englishes and beyond. In *World Englishes: Re-Thinking Paradigms*, E. L. Low & A. Pakir (eds). London: Routledge.
- Canagarajah, A. S. 2013. *Translingual Practice: Global Englishes and Cosmopolitan Relations*. London: Routledge.
- CIA. 2017a. The World Factbook. Germany. <<https://www.cia.gov/library/publications/the-world-factbook/geos/gm.html>> (3 March 2017).
- CIA. 2017b. The World Factbook. Namibia. <<https://www.cia.gov/library/publications/the-world-factbook/geos/wa.html>> (3 March 2017).
- Crystal, D. 2003. *English as a Global Language*, 2nd edn. Cambridge: CUP.  
<https://doi.org/10.1017/CBO9780511486999>



- Crystal, D. 2006. *Language and the Internet*, 2nd edn. Cambridge: CUP.  
<https://doi.org/10.1017/CBO9780511487002>
- Crystal, D. 2008. *Txtng. The Gr8 Db8*. Oxford: OUP.
- Deumert, A. 2014. *Sociolinguistics and Mobile Communication*. Edinburgh: EUP.
- Edwards, A. 2016. *English in the Netherlands. Functions, Forms and Attitudes* [Varieties of English Around the World G56]. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g56>
- European Commission. 2012. Special Eurobarometer 386. Europeans and their languages. <[http://ec.europa.eu/public\\_opinion/archives/ebs/ebs\\_386\\_en.pdf](http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf)> (26 May 2016).
- Evans, S. 2009. The evolution of the English-language speech community in Hong Kong. *English World-Wide* 30(3): 278–301. <https://doi.org/10.1075/eww.30.3.03eva>
- Evans, S. 2014. The evolutionary dynamics of postcolonial Englishes: A Hong Kong case study. *Journal of Sociolinguistics* 18(5): 571–603. <https://doi.org/10.1111/josl.12104>
- Frankfurter Allgemeine Zeitung. 2010. Deutsche Gerichte verhandeln nun auch auf Englisch. <<http://www.faz.net/aktuell/wirtschaft/recht-steuern/gerichtssprache-deutsche-gerichte-verhandeln-nun-auch-auf-englisch-1907201.html>> (3 February 2016).
- Friedrich, P. & Diniz de Figueiredo, E. H. 2016. *The Sociolinguistics of Digital Englishes*. London: Routledge.
- Gilquin, G. & Granger, S. 2011. From EFL to ESL: Evidence from the International Corpus of Learner English. In *Exploring Second-Language Varieties of English and Learner Englishes. Bridging a Paradigm Gap* [Studies in Corpus Linguistics 44], J. Mukherjee & M. Hundt (eds), 55–78. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.44.04gra>
- Görlach, M. 1990[1988]. The development of Standard Englishes [1988]. In *Studies in the History of the English Language*, M. Görlach (ed.), 9–64. Heidelberg: Carl Winter.
- Gupta, A. F. 1997. Colonisation, migration, and functions of English. In *Englishes Around the World, Vol. I. General Studies, British Isles, North America. Studies in Honour of Manfred Görlach* [Varieties of English Around the World G18], E. W. Schneider (ed.), 47–58. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g18.o8gup>
- Hellekson, K. 2009. A fannish field of value: online fan gift culture. *Cinema Journal* 48(4): 113–118. <https://doi.org/10.1353/cj.o.0140>
- Herzog, A. 2014. ‘We, the fans’: Power in the democratic archive of fanfiction. Regensburg: University of Regensburg PhD dissertation.
- Hilgendorf, S. K. 2005. “Brain Gain statt [instead of] Brain Drain”: The role of English in German education. *World Englishes* 24: 53–67. <https://doi.org/10.1111/j.0883-2919.2005.00387.x>
- Hilgendorf, S. K. 2007. English in Germany: Contact, spread, and attitudes. *World Englishes* 26(2): 131–148. <https://doi.org/10.1111/j.1467-971X.2007.00498.x>
- Hilgendorf, S. K. & Erling, E. J. 2006. Language policies in the context of German higher education. *Language Policy* 5(3): 267–293. <https://doi.org/10.1007/s10993-006-9026-3>
- Huber, M. 2014. Stylistic and sociolinguistic variation in Schneider’s Nativization Phase: T-affrication and relativization in Ghanaian English. In Buschfeld et al. (eds), 86–106.
- Imamović A. & Delibegović Džanić, N. 2016. The status of English in Bosnia and Herzegovina: Past and present. In *The Status of English in Bosnia and Herzegovina*, L. Buckingham (ed.), 9–30. Bristol: Multilingual Matters.
- Kachru, B. B. (ed.). 1982[1992]. *The Other Tongue: English across Cultures*. Urbana IL: University of Illinois Press.
- Kachru, B. B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the World. Teaching and Learning the Language and Literatures*, R. Quirk & H. G. Widdowson (eds), 11–30. Cambridge: CUP.

- Kachru, B. B. 1992b. World Englishes: Approaches, issues and resources. *Language Teaching* 25(1): 1–14. <https://doi.org/10.1017/S0261444800006583>
- Kachru, B. B. 1992c. Teaching World Englishes. In Kachru (ed.), 355–365.
- Kautzsch, A. 2014. “English in Germany. Spreading bilingualism, retreating exonormative orientation and incipient nativization?” In *The Evolution of Englishes: The Dynamic Model and Beyond*, S. Buschfeld, T. Hoffmann, M. Huber, M. & A. Kautzsch (eds), 203–227. Amsterdam: John Benjamins.
- Kautzsch, A. & Buschfeld, S. Forthcoming. Theoretical models of English as a world language. In *Cambridge Handbook of World Englishes*, D. Schreier, M. Hundt & E. W. Schneider (eds). Cambridge: CUP.
- Kautzsch, A. & Schröder, A. 2016. English in multilingual and multiethnic Namibia: Some evidence on language attitudes and on the pronunciation of vowels. In *Anglistentag Paderborn 2015. Proceedings*, C. Ehland, I. Mindt & M. Tönnies (eds), 277–288. Trier: Wissenschaftlicher Verlag Trier.
- Kirkpatrick, A. 2010. *English as a Lingua Franca in ASEAN. A Multilingual Model*. Hong Kong: Hong Kong University Press. <https://doi.org/10.5790/hongkong/9789888028795.001.0001>
- Lee, J. S. 2011. Globalization of African American vernacular English in popular culture: Blinglish in Korean Hip Hop. *English World-Wide* 32(1): 1–23. <https://doi.org/10.1075/eww.32.1.01lee>
- Lee, J. S. & Moody, A. (eds). 2012. *English in Asian Popular Culture*. Hong Kong: Hong Kong University Press.
- Mair, C. 2013. The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34(3): 253–278. <https://doi.org/10.1075/eww.34.3.01mai>
- McArthur, T. 1987. The English languages? *English Today* 3(3): 9–11. <https://doi.org/10.1017/S0266078400013511>
- Meierkord, C. 2012. *Interactions across Englishes. Linguistic Choices in Local and International Contact Situations*. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139026703>
- Mesthrie, R. & Bhatt, R. M. 2008. *World Englishes: The Study of New Varieties*. Cambridge: CUP.
- Moag, R. F. 1982[1992]. The life cycle of non-native Englishes: A case study. In Kachru (ed.), 233–252.
- Mukherjee, J. 2007. Steady states in the evolution of New Englishes: Present-day Indian English as an equilibrium. *Journal of English Linguistics* 35(2): 157–187. <https://doi.org/10.1177/0075424207301888>
- Pennycook, A. 2007. *Global Englishes and Transcultural Flows*. London: Routledge.
- Percillier, M. 2016. *World Englishes and Second Language Acquisition. Insights from Southeast Asian Englishes [Varieties of English Around the World G58]*. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g58>
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1972. *A Grammar of Contemporary English*. London: Longman.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Samland, B. M. 2011. *Übersetzt du noch oder verstehst du schon?* Freiburg: Herder.
- Schneider, E. W. 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79(2): 233–281. <https://doi.org/10.1353/lan.2003.0136>
- Schneider, E. W. 2007. *Postcolonial English. Varieties around the world*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511618901>
- Schneider, E. W. 2011. *English around the World. An Introduction*. Cambridge: CUP.

- Schneider, E. W. 2013. Leisure-activity ESP as a special kind of ELF: The example of scuba diving English. *English Today* 29: 47–57. <https://doi.org/10.1017/S026607841300031X>
- Schneider, E. W. 2014. New reflections on the evolutionary dynamics of world Englishes. *World Englishes* 33(1): 9–32. <https://doi.org/10.1111/weng.12069>
- Schneider, E. W. 2016a. Hybrid Englishes: An exploratory survey. *World Englishes* 35: 339–354. <https://doi.org/10.1111/weng.12204>
- Schneider, E. W. 2016b. Grassroots Englishes in tourism interactions. *English Today* 32(3): 2–10. <https://doi.org/10.1017/S0266078416000183>
- Schneider, E. W. 2016c. World Englishes on YouTube: Treasure trove or nightmare? In *World Englishes: New Theoretical and Methodological Considerations* [Varieties of English Around the World G57], E. Seoane & C. Suárez-Gómez (eds), 253–281. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g57.11sch>
- Seargeant, P. & Tagg, C. 2011. English on the internet and a ‘post-varieties’ approach to language. *World Englishes* 30(4): 496–514. <https://doi.org/10.1111/j.1467-971X.2011.01730.x>
- Seidlhofer, B. 2011. *Understanding English as a Lingua Franca*. Oxford: OUP.
- Strang, B. M. H. 1970. *A History of English*. London: Methuen.
- Trudgill, P. 2004. *New-dialect Formation. The Inevitability of Colonial Englishes*. Edinburgh: EUP.
- United Nations Department of Economic and Social Affairs. 2015. *Demographic Yearbook 2015*. <<https://unstats.un.org/unsd/demographic/products/dyb/dysets/2015.pdf>> (3 March 2017).
- Weiß, M. 2016. Sociolinguistic Stratification in a Simulated Society: A Linguistic Investigation of the Video Game *The Witcher, 3: Wild Hunt*. BA thesis, University of Regensburg.
- Werner, V. 2016. Overlap and divergence: Aspects of the present perfect in World Englishes. In *World Englishes: New Theoretical and Methodological Considerations* [Varieties of English Around the World G57], E. Seoane & C. Suárez-Gómez (eds), 113–142. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g57.06wer>
- Weston, D. 2011. Gibraltar’s position in the Dynamic Model of Postcolonial English. *English World-Wide* 32(3): 338–367. <https://doi.org/10.1075/eww.32.3.04wes>
- W<sup>3</sup> Techs. Web Technology Surveys. 2017. Usage of content languages for websites. <[https://w3techs.com/technologies/overview/content\\_language/all](https://w3techs.com/technologies/overview/content_language/all)> (9 March 2017).
- 22blue. 2010. Dragonflies. <<https://www.fanfiction.net/s/6195171/1/Dragonflies>> (9 March 2017).

# Stabilising domains of English-language use in Germany

## Global English in a non-colonial languagescape

Christian Mair

University of Freiburg

The growing impact of English in Germany since World War II has largely been dealt with in terms of lexical borrowing. In contrast to this, the present contribution will focus on emerging domains of regular use of English, be it as a *lingua franca* or as part of multilingual repertoires. Two of these domains, English as Medium of Instruction (EMI) in higher education and English in business and industry, relate to the activities of social elites and have already attracted considerable scholarly attention. Data analysis will, therefore, focus on the diverse types of English-German language mixing which can be observed among two less prestigious and often marginalised groups, namely followers of urban youth cultural movements and the socially disparate group of recent arrivals in Germany from “Anglophone” West Africa. Taken together, these uses of English represent a challenge to the traditional monolingual sociolinguistic order of the nation state because they undermine it from without (increasingly globalised “markets” in business, education and the culture industry) and from within (increasing linguistic diversity of the country’s resident population). My analysis will draw on the “World System of Englishes” model (Mair 2013), enriched by the concept of transnational *languagescapes*, which I develop on the basis of work on cultural globalisation by Appadurai 1996; Loven 2009; and Pratt 2011.

**Keywords:** African Englishes, migration, multilingualism, diaspora, Germany

### 1. Introduction

Over the past three decades, the global spread of neoliberal capitalism, increased currents of voluntary and forced migration and – not least – the digital revolution have led to new uses of English and new transnational connections between existing varieties of English. These are difficult to conceptualise in widely accepted standard models of English as a World Language, such as Kachru’s (1985) “Three Circles of

English” or Schneider’s (2007) “Dynamic Model,” which were primarily concerned with documenting the sociolinguistic legacy of colonialism and early postcolonial nation-building. In Mair (2013) I proposed a systems-theoretical approach to World Englishes which was intended to fill this gap, the empirical test case being provided by the language practices found on web forums maintained by globally dispersed digital communities of practice of West African and Caribbean heritage. The present chapter extends the approach to cover a non-colonial environment, by analysing English as part of the linguistic ecology of present-day Germany.

The argument will be developed as follows. Section 2 will review the major changes in the status and function of English in Germany since the end of World War II. The general trend to be noted is that English has developed from a foreign language, which was learned for its cultural prestige and its usefulness in international communication, into a lingua franca which, in several domains of communication, has assumed essential functions not only on the international but on the national level, as well. Pace Kautzsch (2014), I do not regard these developments as harbingers of an emerging new variety of German English. Rather, I take them as proof that the English language has become a key component in the linguistic ecology of a country which has become increasingly diverse and multilingual internally and increasingly connected globally. Section 3 briefly revisits Mair (2013), suggesting some refinements of the model and pointing out relationships to recent related work by others. Sections 4 and 5 will provide empirical support and illustration for the central argument of the present chapter. The focus of Section 4 is on the ever deeper “Anglicisation” of German pop music and youth culture over the past six decades, while Section 5 exemplifies lingua franca use of English and various types of mixed English-German language practices among the country’s growing African immigrant and refugee communities.

## 2. Globalisation and the changing role of English in Germany

There is no doubt about the growing presence of English in Germany, and the growing impact of English on German, since the end of World War II. Most publications, however, have focussed on two relatively narrow and superficial aspects, namely on *Anglizismen*, borrowings from English into German, and on *Anglizismendebatte*, the language-ideological debates which have arisen in the wake of their striking recent increase in frequency (see Göttert 2013 and Spitzmüller 2005 for documentation). By comparison, “there has been little examination of how, why, and with whom Germans use English” (Hilgendorf 2007: 144).

The first among a small number of studies, usually produced by scholars based outside Germany, which have approached English in Germany from a more

international, “World Englishes” perspective was Berns (1988). Berns deplores the bias towards the study of lexical borrowings and argues that the role of English should be studied in a much wider cultural and social context. In hindsight, she can be credited with recording the first signs of developments which have intensified dramatically since the time of her writing, for well into the 1980s Germany remained a good fit for the Expanding Circle, as Kachru defined it in his tri-partite taxonomy of World Englishes (e.g. Kachru 1985).<sup>1</sup> English was taught and learned as a foreign language mainly in the school system. Teaching was expected to enable German speakers to communicate with native speakers of English and to gain access to their culture (with an original strong focus on Britain that gradually widened to include other parts of the English-speaking world). Lingua franca use, i.e. use of English as a link language among non-native speakers, was confined to core transnational domains such as diplomacy, international trade, air traffic, tourism, and publication in the natural sciences. It was generally not addressed in the teaching of English in school, but only in subsequent professional training, under headings such as “English for Specific Purposes” (ESP) or “Business English.” In Germany, English did not show any signs of the sociocultural nativisation typically associated with what Kachru referred to as the “Outer Circle” varieties.

By the time Hilgendorf (2007) published her analysis of the status and functions of English in Germany (2007), lingua franca uses of English, which were largely confined to communication between professional elites and international partners, had become routine for increasing numbers of people in their daily activities within the country itself. Hilgendorf emphasises both the extended *range* of communicative domains in which English is used and the increased *depth* of its impact on the local linguistic ecology, which is chiefly reflected in the growing percentage of the population using English on a regular basis:

At the beginning of the 21st century, English has now spread to numerous domains in the German context. These include politics, law, business, advertising, science and research, the media (popular music/radio, television, film, the Internet), and education [...] Whether at work dealing with clients, at school taking classes, or at home relaxing and being entertained, for a growing number of Germans contact with the English language is a frequent, if not daily, occurrence.

(Hilgendorf 2007: 135)

---

1. The other two Circles are the “Inner Circle” of natively spoken varieties (e.g. Britain, US, Australia) and the “Outer Circle” of second- or official-language varieties, i.e. emerging postcolonial standards, which – though usually not spoken natively but alongside other languages in multilingual repertoires – nevertheless show signs of sociocultural nativisation and institutionalisation in the relevant communities (e.g. India, Nigeria, Ghana).

Germany's higher-education system is a case in point. By the 1980s English had become the default language of scientific publication in the natural sciences, and was in the process of establishing a similar position across most other academic disciplines (Tsunoda 1993; Ammon 2006). Its role in teaching was still marginal, however. This has changed considerably. English is now the language of instruction for increasing numbers of students, both German and international, who are enrolled in English-taught programmes (Gürtler and Kronewald 2015, Gundermann 2014: 2–9, Hilgendorf & Erling 2006). As a result, even university administrations, traditionally solid bastions of German, are having to cope with English.

Striking changes can also be noted in the media. Before the digital revolution of the 1990s films and television were produced, distributed and consumed in institutional frameworks which were tightly regulated at the national level. For the vast majority of German viewers, Hollywood movies were accessible in dubbed German versions only.<sup>2</sup> American and British politicians' and celebrities' voices were rarely heard in the original but only in brief snatches through the German voice-over.<sup>3</sup> Original English-language media content from outside Britain and the US was even more of a rarity. This is clearly different from the transnational and multilingual *mediascapes* (Appadurai 1990, 1996) created by CNN, cable and satellite television, Internet streaming services and Youtube. National borders and language boundaries have become much more permeable, which has not only vastly increased exposure to English among large sections of the population, but – to the extent that the new media encourage audience participation – led to more active use of English, usually in mixed or truncated form.

In view of the fact that the use of English as a lingua franca in elite social domains such as academia, higher education and international business has already received a fair amount of scholarly attention (Ammon 2006; Hilgendorf 2007, 2010; Hilgendorf & Erling 2006), the focus of the present chapter will be on the use of English, often as part of multilingual repertoires, in non-elite domains.

One striking illustration of the changing status of English in one such domain, youth culture, is provided by the almost instant borrowing and adaptation of “global English slang” (Coleman 2014). Piggybacking on the spread of Standard English as a lingua franca, American slang terms and expressions have become

---

2. Cf., in this connection, Hilgendorf's (2013) study of the changing conventions in the translation of the titles of internationally distributed Hollywood movies.

3. One notable exception is the broadcasting activity of AFN, the United States' Armed Forces Network, in Germany from 1945 to the early 1990s, whose role is also mentioned by Berns (1988: 39). AFN broadcasts such as DJ Casey Kasem's “American Top Forty” had a considerable following among young German-speaking listeners.



available as a linguistic resource in German youth culture in ways that French or Spanish slang have not.

The inevitable presence of English in contemporary Germany can also be detected in the language habits of some of the country's new immigrant communities. While "Deutsch als Fremdsprache" (DaF, German as a Foreign Language) is learned without a detour through English in the established Turkish community, temporary residents and recent immigrants from "Anglophone" Africa have learner varieties which show transfer and interference from English (as will be exemplified in Section 5 below). More importantly, mixed use of English and German remains a linguistic survival strategy for this group for many years in communication with German institutions and the resident German-speaking population. This poses considerable (if different) challenges for both parties concerned.

For many individual learners/users, both in Germany and in the Expanding Circle at large, English continues to be a foreign language which is learned with varying degrees of motivation and effort and mastered at widely varying levels of individual competence. At the societal level, however, the world's lingua franca has long ceased to be a foreign language like any other. *How ya gonna keep 'em down on the farm (after they've seen Paree)* was the satirical question which a popular 1919 song asked about rural American youth who had been exposed to the experience of World War I in the US army.<sup>4</sup> *Why you ain't gonna get it back into the Expanding Circle after it's gone global* is the question which underpins the present study of the new roles of English in a traditionally foreign-language environment.

### 3. Remapping the "Expanding Circle": New types of contact among varieties of English, new multilingual practices

Over the past half-century English has rapidly developed from being one of several competing world languages into the undisputed global lingua franca. The historical foundations of this process were laid by British colonialism, but colonialism and its aftermath have had little to do with its final stages, which played out in world regions such as East Asia and the successor states of the Soviet Union, which were affected by British colonialism only indirectly or hardly at all (see Northrup 2013: 137–160). As the world's lingua franca, English has gained a historically unique position and has stopped being a foreign language like any other, in Germany as well as in most other nations placed in the "Expanding Circle" in Kachru's model.

---

4. Sam L. Lewis and Joe Young (lyrics), Walter Donaldson (music).



English is the only language in the world which is a potential contact language for all others. By implication, English is therefore also the only language which – at least in principle – is universally available as a component in multilingual repertoires. This degree of linguistic omnipresence and integration needs to be taken into account in any viable sociolinguistic model of World Englishes. A systems-theoretical approach to global multilingualism which provides a good starting point is the “World Language System” proposed by de Swaan (2001, 2010). In this model, the world’s ca. 6,000 languages are integrated into a single system which is both comprehensive and strictly hierarchical. The world’s linguistic diversity is stratified into four layers:

1. *one single hyper-central language, functioning as the “hub” of the World Language System: English*
2. *a very small number of super-central languages: Arabic, Chinese, English, French, German, Hindi, Japanese, Malay, Portuguese, Russian, Spanish and Swahili<sup>5</sup>*
3. *a somewhat larger but still relatively small number of central languages: e.g. Dutch, Finnish, Korean, ...*
4. *peripheral languages, i.e. the vast majority of the 6,000+ languages of the world.*

For de Swaan, the World Language System works as “a surprisingly efficient, strongly ordered, hierarchical network, which ties together – directly or indirectly – the 6.5 billion inhabitants of the earth at the global level” (2010: 56). Alongside the political, economic, ecological and cultural dimensions, it represents the linguistic dimension of globalisation, drawing attention to the impact of globalisation on language use and at the same time recognising the partial autonomy of the linguistic dynamics of globalisation from the other dimensions.

What makes the “World Language System” truly global is that it has a single hub today, namely English. One tier below English we find *super-central languages*, i.e. major standard languages with transnational reach (many of which have been or still are being referred to as “world languages”). On the third tier, there are the world’s *central languages*, usually standardised and institutionally recognised at the national level. The fourth tier comprises the *peripheral languages*, which tend to be spoken by small communities and often lack a written standard, media presence and other institutional support. It is obvious that, as we move down the hierarchy, the number of languages gets larger: one at the top, a dozen or so on the second tier, 150 at most on the third, and the vast majority at the bottom.

As has been pointed out, the World Language System is highly efficient. As with other aspects of globalisation, there is no guarantee that efficient systems will automatically work fairly for those participating in them. So clearly there is scope for improving the system through language policy and planning. However, for

---

5. In the updated 2010 version de Swaan adds Turkish to this list (2010: 57).

language policy and planning measures to be successful, the hierarchical stratification of the current global linguistic ecology has to be recognised as the inevitable point of departure.

Inspired by de Swaan's World Language System, Mair (2013) proposed a similar systems-theoretical analysis of the global "English Language Complex" (McArthur 2003: 56; Mesthrie & Bhatt 2008: 1–3, 12–17): the "World System of Standard and Non-Standard Englishes." On the one hand, this World System of English is simpler than the World Language System, because it accounts for one language only. On the other hand, it is more complex, because it distinguishes between standard, nonstandard and lingua franca uses, thus recognising the fact that it is not only Standard English which is globally dispersed, but also nonstandard uses:

### World System of Standard and Non-Standard Englishes

- *hyper-central variety/hub* of the World System of Englishes: Standard American English
- *super-central varieties*:
  1. *standard*: British English, Australian English, Indian English, South African English, Nigerian English
  2. *non-standard*: African American Vernacular English, popular London English, Jamaican Creole, and a very small number of others
  3. *regionally unrestricted, but domain-specific ELF uses*: e.g. science, business, international law
- *central varieties*:
  1. *standard*: Irish English, Scottish (Standard) English, Canadian English, Jamaican English, Ghanaian English, Kenyan English, Sri Lankan English, Pakistani English, New Zealand English, and a small number of others
  2. *non-standard*: Northern British English urban koinés, US Southern English, Nigerian Pidgin, and a small number of others
  3. *regionally restricted and domain-specific ELF uses*: e.g. "Euro-English" (i.e. stabilised lingua franca use of English in the European institutions)
- *peripheral varieties*:
  1. *standard*: Maltese English, St Christopher and Nevis English, Papua New Guinea English, and others
  2. *non-standard*: all traditional rurally based non-standard dialects, plus a large number of colonial varieties including pidgins and creoles

**Figure 1.** The World System of Standard and Non-Standard Englishes (adapted and expanded from Mair 2013: 264)

As Figure 1 shows, Standard American English is the hub which integrates the system at the global level. This can be demonstrated easily. Just as in the World Language System other languages are more likely to borrow words from English than English is to borrow from them, so in the World System of English other varieties are more likely to follow American usage than American English is to follow

developments in other varieties. This asymmetry has also been characteristic of the relationship between American English and British English for some time now, which is why Standard British English, in spite of its continuing role as one of two globally relevant standards in EFL teaching, has been listed as the most powerful of the super-central standard varieties here.<sup>6</sup>

Similarly asymmetric relations hold between the other standard varieties listed. Australian English has been classified as super-central and New Zealand English as central because the latter is more likely to borrow from the former than the other way round; also, Australian English has a wider sphere of influence, for example as a foreign-language teaching model, than New Zealand English in the Asian-Pacific region. Note that the model also provides for internationally largely irrelevant peripheral standard varieties of English (no value judgment implied, of course). These are associated with communities which lack demographic weight, are geographically isolated, economically weak or fragmented in other ways. Their ability to develop and maintain endonormativity is precarious. As an example, consider the Federation of St. Christopher and Nevis, one of the small Caribbean nation-states, with a population of ca. 52,000 people. Formal and educated use of English in this polity is influenced by Jamaican and general Caribbean norms, which themselves are developing in a force field defined by the competing pressures of British English, the old colonial norm, American English, the currently dominant standard in the region, and emerging local norms, which are shaped by contact with the local English-lexifier creoles. In addition, a considerable part of the population of St. Christopher and Nevis tends to reside abroad at any given time. Transnational family networks and complex patterns of circular migration are factors of linguistic influence to reckon with.

In the absence of empirical research, it will always be difficult to decide whether a variety of formal English is a peripheral standard in this sense, and I am therefore happy to add that the classification of the three examples listed in Figure 1 is provisional. However, they stand as a reminder of the importance of a question which has not been asked in World Englishes linguistics yet: What are the minimal demographic, economic, and institutional prerequisites for the endonormative stabilisation (Schneider 2007) and standardisation of New Englishes?

---

6. There are other indicators of the present-day power-differential between the British and American standards. American spellings (*center, favorite*, etc.) are preferred when English words are borrowed into other languages. Some British academic publishers demand American orthography from their authors, while the reverse is never done. British fiction (famously, the Harry Potter series in recent times) is linguistically edited for the American market, while British audiences are expected to cope with American fiction and films in the original versions.

As Figure 1 shows, there is no non-standard global hub, i.e. a globally relevant non-standard variety of English exerting potential influence on *all* others. However, several non-standard varieties can lay claim to super-central status because they have added transnational functions to their purely vernacular local uses. African American Vernacular English (AAVE) is the prime case in point. Within the US, it is important as a contact vernacular for Hispanics and recent immigrants from Africa and the Caribbean (including territories which are not historically Anglophone). Even more widespread are mediated and commodified uses of AAVE, for example in the global hip-hop scene. In the guise of Hip Hop Nation Language (HHNL, Alim 2015), AAVE elements have been taken up by youth of all ethnic backgrounds in the US and the rest of the world, both English-speaking and non-English-speaking (see Alim, Ibrahim & Pennycook (eds) 2009). Similar points can be made about Jamaican Creole (where the cultural and media vectors in the spread have also been popular music subcultures) and Nigerian Pidgin (where one of the vectors has been Nigeria's popular "Nollywood" movie industry).

Note that there are cases such as Jamaica in which the local relations of prestige between the standard and the vernacular have been turned around in the global linguistic ecology. For example, Jamaican Standard English dominates Jamaican Creole in the sociolinguistic order of the post-colonial nation-state. In Figure 1, however, Jamaican Standard English appears on the third tier, as a typical example of a central standard, its role uncontested within the boundaries of the postcolonial nation-state, but exercising little influence beyond that. Jamaican Creole, on the other hand, is classified as a super-central non-standard variety, due to a strong transnational presence in the Caribbean diaspora, in several domains of youth culture, and the associated media and entertainment industries. For proof, it is not necessary to go further than the OED Online. A search for all entries with a Jamaican etymology first attested between 1950 and 2016 yields 37 examples, of which 33 originate from Jamaican Creole (including its subcultural and slang offshoots such as Rastafarian dread talk) and colloquial/mesolectal Jamaican English, to wit: *bashment*, *bloodclaat*, *bluebeat*, *criss*, *cruft*, *def*, *dub*, *dubplate*, *irie*, *junglist*, *livity*, *locksman*, *natty*, *nice* (v), *one-drop*, *pan*, *peeny-wally*, *picky*, *punani*, *pussyclaat*, *ragga*, *Ras*, *rass*, *rassclaat*, *reggae*, *rootsy*, *rudie*, *sing-jay*, *ska*, *stylee*, *sufferation*, *toaster* (n3), *yardie*.<sup>7</sup>

Paraphrasing de Swaan (2010: 56, and quoted above), we can say that the World System of Englishes constitutes a surprisingly efficient, strongly ordered, hierarchical network, which ties together – directly or indirectly – the two billion regular users of the language at the global level. The fact that some privileged varieties of

7. Conditions set in the Advanced Search mode: *Jamaican* in "Etymology" OR *Jamaican* in "Full Text" AND "1950–2016." This resulted in 44 hits, of which seven were removed as irrelevant.

English – whether standard or nonstandard – are now indeed everywhere is highlighted in a poignant scene from Zadie Smith’s novel *Swing Time*. The first-person narrator, a British-Jamaican woman, encounters a local in The Gambia:

‘You sound American,’ I said, but that was only one thread of the rich tapestry of his voice. Many different movies and adverts were in there, and a lot of hip-hop, *Esmeralda* and *As the World Turns*, the BBC news, CNN, Al Jazeera and something of the reggae that you heard all over the city, from every taxi, market stall, hair-dresser. An old Yellowman tune was playing right now, from the tinny speakers above our heads. (Smith 2016: 373)

The irony, as it turns out, is that the local man, whose English reflects the whole world, has never been abroad. However, he may well be a literary representation of the Gambian *bumster*, an informal mediator between sun-seeking European and American winter tourists and the local population, and hence very much a part of the globalised economy (see Lawson & Jaworski 2007).

Like the World Language System, the World System of Englishes may be very flexible and efficient, but efficiency does not necessarily imply fairness:

Accepting a naïve conceptualization of English as a monolithic entity putting everyone on the same footing is risky because we cannot disregard the practical relevance of the sociolinguistic concept of linguistic variety. For example, being a rich native British speaker abroad, say, in Spain, cannot be viewed as equivalent to being a poor African-born speaker of English who migrated to that country. Speaking English as such therefore does not mean all that much if we do not clarify *who* speaks this language with *which* accent and *where*.

(Gazzola & Wickström 2016: 13–14)

The above discussion should have made clear where the advantages of the “World System of Englishes” model are for the study of the status and functions of English(es) in contemporary Germany. Rather than merely deal with English *tout court*, we can refine the question and ask who speaks (or draws from) which variety to whom, and on which occasion. Also, we can more easily handle the complication that in an Expanding-Circle locale, we will constantly be witnessing encounters with, appropriations of and responses to Inner- and Outer-Circle Englishes, both standard and nonstandard, in addition to the foreign-language and lingua-franca uses expected.

Since the publication of Mair (2013), a number of studies have appeared which address similar issues and usefully complement the approach advocated here. Schneider (2014) has also noted a recent and very dynamic restructuring of the Expanding Circle, characterised by “the widespread emergence of hybrid mixes between local languages and English and phenomena of ‘poststructuralist diffusion,’ English being adopted by whatever means, in fragments and unconstrained of norm concerns, driven by strongly utilitarian considerations” (2014: 9). He investigates

whether these developments can be fully accounted for in the “Dynamic Model” (Schneider 2007), but finds that:

despite some similarities [the Dynamic Model] is not well suited to grasp the vibrant developments of the Expanding Circle. Instead, the notion of ‘transnational attraction’ is defined and proposed as an appropriate conceptual framework.”  
(2014: 9)

*Transnational attraction* is clearly a factor which is at play in the German case studies to be discussed in Sections 4 and 5 below. Also starting out from the Dynamic Model, and in response to the same perceived blurring of the boundary between postcolonial and non-postcolonial Englishes, Buschfeld and Kautzsch (2017) propose an “EIF Model” accounting for *Extra-territorial* and *Intra-territorial Forces* governing the spread of English in the era of globalisation; their case study is Namibia, where – given the complex colonial history of the territory – the extremely rapid and uncontroversial move to English as de facto official language and interethnic lingua franca may have come as a surprise.

There is also a small but growing body of research which involves the notions of *linguascapes* or *languagescapes* – both concepts which are highly compatible with Mair’s and Schneider’s approaches to the new developments in the Expanding Circle. Mary Louise Pratt, who has used the term *languagescape* informally, points out that:

In the global languagescape, new forms of linguistic distribution are in play. In electronic form, any language can travel anywhere any time. With access to tools, anyone can appropriate, broadcast, download, study any language they want, for any purpose they want, without asking permission. (Pratt 2011: 279)

A more technical use of the term had been proposed even earlier, by Loven (2009) in an ethnographic study of television viewing habits in Indonesia. She introduced *languagescapes* with explicit reference to the five *-scape* neologisms coined by Appadurai (1990, 1996) to account for the transnational flows of cultural globalisation. The global *technoscape* and the global *finanscape* refer to the hard-wiring of globalisation. The world’s changing *ethnoscap*es, *ideoscap*es and *mediascap*es, on the other hand, are shaped by the intensifying transnational flows of people, their ideas and cultural artefacts and cover the “soft” social, cultural and psychological dimensions of globalisation. Languages and language variation play obvious and direct roles in all three of them, which opens up very promising perspectives for interdisciplinary cooperation between linguistics, history, the social and political sciences, and cultural studies. It is with this aim in mind that I have advocated the systematic use of the term *languagescape* in research on World Englishes and globalisation (Mair 2017), and it is in this sense that it is used in the title of the present

chapter.<sup>8</sup> Grounded in Appadurai's comprehensive theory of cultural globalisation, the concept is helpful because it redresses four potential biases in existing research on World Englishes:

1. *Territorial bias*: Most of the prevailing classifications of World Englishes – from Kachru's (1992) "Three Circles" to Schneider's (2007) "Dynamic Model" – are organised around national varieties and geographical boundaries. Languagescapes acknowledge the territorial factor, but accept that boundaries can be fuzzy and shaped by human experience.
2. *Vernacular bias*: For the pioneering "First Wave" (Eckert 2012) sociolinguists, the most authentic data was to be obtained from spontaneous face-to-face interaction in closely knit vernacular communities. Spontaneous speech of this kind provides the baseline of any languagescape, as well. But languagescapes are also shaped by stylised, performed and mediated language.
3. *Colonial bias*: The study of "varieties of English around the world" took shape under the long shadow of colonialism. Neither colonialism nor decolonisation, however, account for the growing presence of English in Germany, Russia or Japan. Languagescapes are agnostic with regard to whether a particular use of English has a colonial history or not.
4. *Monolingual bias*: The study of World Englishes has always recognised the sedimented traces of language contact – in the shape of lexical borrowings and substrate influence in pronunciation and grammar. There has been relatively little interest, by comparison, in the multilingual settings in which varieties of English are used – and even less in the multilingual practices of their speakers. This bias is redressed in the languagescapes, which are multilingual by default and monolingual only in the marked case.

This is the perspective through which I will explore the changing roles of English in Germany's increasingly multilingual languagescape. Section 4 below will discuss samples of popular music lyrics from the last 50 years to show the transition from lexical borrowing to linguistic hybridisation. Section 5 will analyse data from language-biographical interviews with Africans residents in the Freiburg area to define research priorities in this domain.

---

8. Pennycook (2003: 523) suggested *linguascape* as an appropriate term to add an explicitly linguistic dimension to Appadurai's five other *-scapes*. To the best of my knowledge this suggestion was not followed up in any detail by either himself or others until Dovchin's recent research on mixed language practices involving English in Mongolian hip hop (e.g. Dovchin 2017).



#### 4. From borrowing to linguistic hybridisation: 50 years of English in German popular music

Defeat in World War II and the subsequent period of American and British occupation put paid to the ideology of German supremacy and triggered an unprecedented wave of lexical borrowing from English. Vocal resentment by cultural conservatives notwithstanding, many of these new Anglicisms were soon firmly entrenched in popular usage, as “symbols of American attitudes, values and modernity” (Berns 1988: 39). One of the domains in which Anglicisms were conspicuous from the start was popular music and youth culture. The following is a 1958 German adaptation of US singer Connie Francis’ (= Concetta Rosa Maria Franconero, b.\* 1938) popular hit “Stupid Cupid.”<sup>9</sup> Direct lexical borrowings from English are printed in bold; forms displaying more indirect or potential English influence are italicised:

*Sexy Hexy* sagt der *Jo* zu mir,  
*Sexy Hexy* sagt auch *Tom* zu mir,  
 doch ich find es einfach lächerlich,  
 dieser Name passt doch nicht für mich.  
*Hey Hey*, rufen sie,  
*Sexy Hexy*, doch ich höre nie.

Auf der netten **Party** gestern Nacht  
 haben alle **Boys** mich angelacht,  
 denn mein neuer *Pulli* steht mir gut,  
 und die **Boys** bekamen plötzlich Mut.  
*Hey Hey*, rufen sie,  
*Sexy Hexy*, doch ich höre nie.

Ob ich im *Sportdress* mal vor unserm *Clubhaus* steh'  
 [...]

[translation: *Jo says Sexy Little Witch to me / Tom also says Sexy Little Witch to me / but I simply find it ridiculous / this name doesn't fit me at all / Hey hey, they shout / Sexy Little Witch, but I never listen // At the nice party last night / all the boys were smiling at me / cause my new jumper suits me fine / and the boys started taking courage / Hey hey, they shout / Sexy Little Witch, but I never listen // If I stand in front of our club house wearing sports gear ...]*

*Boy*, *sexy* and *Party* are obvious lifestyle Anglicisms, in frequent use today. *Pulli* is a clipped form of *pullover* and, like *Sportdress* (“sportswear”) and *Clubhaus*, also fully established usage today. *Hexy* (*Hexe* “witch” + diminutive suffix) is complex,

9. Da[n]ny Mann, “Sexy Hexy” [originally “Sexie Hexy”], 1958. An mp3 recording is available on Youtube: <https://www.youtube.com/watch?v=pC85M4gpKTK>.



as *i*-diminutives have a native history in German, but are likely to have increased in frequency under contact influence from English. None of the Anglicisms used in the song are unusual in themselves, though their concentration in such a short span of text may be. Like the performer herself (*Dany Mann*), the proper names *Jo* and *Tom* are anglicised or given a form which is phonologically compatible with either German or English. As in the song, most post-World War II Anglicisms have been nouns and adjectives. All this is undramatic, and any shock value such English-influenced language use may have had in the 1950s is difficult to understand from the vantage point of the present.

As is the case in many other languages, lifestyle Anglicisms have not fully replaced their native synonyms in German, but co-exist with them in a relationship of functional diglossia. Where the German member of the pair signals practicality and utilitarian value, the English one stands for conspicuous consumption and hedonism (cf. *einkaufen* – *shoppen*). Where the German word signals tradition and sincere emotional attachment, the English one reduces commitment (cf. *am Wochenende kommt meine Familie auf Besuch* – *am Wochenende kommt die Family auf Besuch*, “on the weekend my family will be visiting”). Where the German word signals elite culture, the English hints at commercialised entertainment (cf. *Opernergebnis* – *Opernevent*, “great night at the opera”).

The approach to English is rather different in the following example of (Austrian-)German hip hop, “Kein Limit,”<sup>10</sup> produced more than five decades later (in 2015, to be precise). The same notational conventions apply as in the preceding case:

[...]

[Part 2: Money Boy]

Kein **Limit**, **no limit**, **no**, wie bei **Master P**

**Yeah**, ah, **pass** mir mal den **Hennessy**

Wir **flippen** die **Packs**, **wippen** das **Crack** und falls diese **Feds** **watchen**

**Ferragamo-Belt**, **Louis-Sneakers**, wir sind **fresh as fuck**

**Undercover-Cops** **liken** die **Pictures** jetzt auf **Instagram**

Wie wir grade **posen** mit den **Choppers**, wir sind **Kings**, **verdamm**t

Mach das alles für die **Hater** da, es **ging** an mit 'nem **Major-Plan**

Und heute gehen wir auf **Tour** und **spiel'n** vor 16.000 Mann

Füllen Hallen auf wie **Luftballons**, **die Venues** sind jetzt **packed**

Aber trotzdem **slang** ich weiter auf der **Avenue** das **Crack**

Und **gette Money** wie ein **Motherfucker**, **hard work** **pays off**

**24/7** und ich **take** keine **Days off**

10. Money Boy (feat. Spinning 9 & Hustensaft Jüngling), 2015; source: <http://genius.com/Money-boy-kein-limit-lyrics>, with adaptations.

[No limit, no limit, no, like with Master P / Yeah, ah, pass me the Hennessy / we flip the packs, whip the crack and if these feds are watching / Ferragamo belt, Louis sneakers, we're fresh as fuck / undercover cops are liking the pictures now on Instagram / how we're posing with the Choppers, we are kings, damn / Do all this for those haters, it started with a major plan / and today we go on tour and play for 16,000 people / fill the concert halls like balloons, the venues are packed / but on the avenue I still go on slanging the crack / And get money like a motherfucker, hard work pays off / 24/7 and I don't take days off]

There are obvious differences between the way the English elements are embedded in the German text in the two songs. The foremost is contextual: The reference variety here is not Standard or even colloquial English, but African American Vernacular English (AAVE) in the specific form of Hip Hop Nation Language (Alim 2015), which – as will be remembered – has been classified as the internationally most prominent of the *super-central non-standard* Englishes above. This nonstandard orientation is reflected in the relatively unrestrained use of taboo vocabulary (*fresh as fuck*, *motherfucker*), street talk and drug users' slang. In addition, there are specific intertextual allusions to classic works of the hip hop canon, which will be recognised and appreciated by insiders. *Falls diese Feds watchen*, for example, takes up a motif from 2Pac Shakur's "All eyez on me:" "The feds is watchin', niggas plottin' to get me/Will I survive, will I die, come on let's picture the possibility."<sup>11</sup>

*Feds*, short for agents of the Federal Bureau of Investigation (FBI), refers to federal as opposed to state or local law enforcement in the US. Cut off from this specific institutional context, the term has become general slang, potentially referring to any kind of law enforcement agency anywhere. Its use in this sense has been common for some time in urban British youth language, and this is the sense in which it is used in the German text analysed here. This particular instance is just one example of a whole set of highly mobile language practices which hip hop communities of practice have used to create a languagescape which transcends national borders and language boundaries.

In terms of language structure, the major difference between the 1958 and 2015 texts is that borrowing is no longer restricted to nouns and adjectives, but includes verbs (*likened*, *getten*, *taken*, ...) and idiomatic chunks (*hard work pays off*). Among the verbs, only *likened* conforms to the traditional profile of English lexical borrowings into German. The verb *like* is not used in its general sense, but in its specific social-media use, which originated in English and was then taken over into colloquial German. *Getten* and *taken*, by contrast, retain their wide English meanings. Here, borrowing cannot be traced back to a specific material cause, the

11. See <http://genius.com/2pac-all-eyez-on-me-lyrics> for a transcription of the complete song.

motivation is purely rhetorical, and, unlike *liken*, these two verbs are not in general use in German.

There are hardly any antecedents for this type of heavily mixed language practice in the national German tradition of dealing with English (though similar degrees of integration are attested historically for German and Latin).<sup>12</sup> On the other hand, there are clear structural parallels with the following sample of Camfranglais, an urban contact vernacular from Cameroon, which shows colloquial and non-standard French heavily laced with English and Pidgin English elements:

Quand je *waka* [= walk] dans la rue ils me *look* comme un potentiel Ben Laden, les yeux te disent preske “*Back* chez toi”!... Mais il faut aussi *understand* ke comme les Camers se cherchent ici, ils “*goent* au front”, c’est comme ça ke les *whites* se chechent [sic!] au Camer...<sup>13</sup>

Several linguistic features of this short text index an international orientation toward global modernity. There are, for example, the nonstandard <k> spellings of words such as *que* and *presque*, which are typical of French digital orthography. Equally transnational, but sociologically more specific, is the use of English for the terminology of race. *Les whites* is used instead of the default French *les blancs* to refer to white people. *Les blacks* could have been used in exactly the same way. This usage is widespread throughout the postcolonial *francophonie* and also found in France itself among diasporic communities and their youth-cultural and sub-cultural affiliates. The use of English is best understood as recognition of the pioneering role of African American and Caribbean artists, activists and intellectuals in the quest for the emancipation of Africa and its diaspora. Note that while most English terms are given in their standard citation forms, one (*goent*) is inflectionally integrated into French (although this particular inflection only exists in spelling and would not be sounded). One verb appears in its Cameroonian Pidgin form (*waka*). Taken together, this mix of linguistic features is not random or mere individual playfulness. Rather, it hints at the existence of a set of transnational and multilingual languagescapes associated with the global African diaspora.

The rich cultural and linguistic substratum of a postcolonial diaspora is not something which an Austrian hip hopper can draw on, but there are clear parallels between the two texts with regard to the formal strategies of language mixing, and also with regard to some of the indexical social values associated with the English element in the mix.

---

12. See the extracts from Martin Luther’s “Tischreden” analysed in Auer 1999.

13. This example is from a corpus of posts from a Cameroonian diasporic web-forum (see Mair 2013: 278).

What remains difficult to assess is the degree of realism of the representation when we take spontaneous language use as the benchmark for comparison. On that score, the contrast between the use of English by a contributor to a Cameroonian web forum and an Austrian rapper is considerable. The Cameroonian's posting style is authenticated by widely documented similar practices in face-to-face interaction in certain urban milieux (Schröder & Rudd 2017). How much of Money Boy's German-English language mixing survives off-stage or offline is an open question. It is easy to find evidence of the performer retaining his style as a trademark in rehearsed interviews: "wir hamm a nice Video dazu geschootet [...] das Movement is am Moven/wir getten das Money/wir sind booked out/wir spielen Shows/wir sind booked out bis Ende des Jahres" (<<http://genius.com/Money-boy-kein-limit-lyrics>>, embedded video comment).<sup>14</sup> However, even rehearsed occasions occasionally require spontaneity, and it is on such occasions that the performer tends to revert to largely monolingual (Austrian) German.

At the time of writing, there are signs that Money Boy's trademark style of English-German language-mixing has entered the active use not only of his own immediate followers, but other young speakers, as well. Most widespread is the symbolic use, mostly as interjections, of hip hop slang items such as *swag*, *fly* and *burr*, which he has propagated actively (e.g. in his "Dreh den Swag auf" [= "turn on the swag"]). Such single-word or single-phrase insertions into German discourse are not without antecedent. For example, *no way*, from Standard English, has served as an emphatic variant of German *nein* for a number of years among wide segments of the (younger) population. The interjection *Joke!* is widely current among young people in the sense of "Don't worry, I'm just kidding." More elaborate mixing of the type *jemand muss das beer getten* ("somebody has got to get the beer") or *wie kann man nur so ein lames video shooten!* ("how can anyone shoot such a lame video!") can be heard, too.

Systematic corpus-based research on these practices is difficult, however, as they are largely restricted to the spoken language. If one consults the Web for the trigram *wir getten das* (i.e. a sentence frame illustrating the use of *get* as an inflected transitive German verb), one will find around 200 results, most of which repetitively refer to hip hop song lyrics by Money Boy or other performers. In spite of their enormous size, the standard reference corpora of Modern German are not helpful, because they comprise large amounts of written language and do not cover the immediate present except in the newspaper genre. This is the sole example of *getten* from the 29-billion-word DeReKo (Deutsches Referenzkorpus), made available by the Institut für Deutsche Sprache (IDS):

---

14. Translation: *We have shot a nice video about it. [...] The movement is moving ahead. We are getting the money. We are playing shows. We are booked out through the end of the year.*

Und man sollte bereit sein, all seine Coolness zusammenzuraffen, wenn die mit Anglizismen gemästete Sprache ihren Enddarm nach außen stülpt: „You can get it if you really want. Ich wante vermutlich nicht really genug. Auf der anderen Seite wante ich zumindest genug, um ordentlich unzufrieden zu sein, es nicht zu getten.“

(U09/APR.01360, original source: *Süddeutsche Zeitung*, 9 April 2009, p. 14; “Gegen Plappern helfen keine Pillen” [Review of Sarah Kuttner, *Mängelexemplar*] [*Readers should be ready to pluck up all their coolness to confront a language stuffed with Anglicisms moving its bowels. “You can get it if you really want. Probably I don’t really want enough. On the other hand I at least want enough to be thoroughly unhappy about not getting it.*]

Sarah Kuttner (b. 1979), considered by some to be an authentic voice of hip and trendy multicultural urban German youth culture, is taken to task by the reviewer for her excessive use of Anglicisms. The passage quoted in evidence, however, fails to serve the intended purpose. Formally, *you can get it if you really want* is a sentence-length code-switch into English. Functionally, however, it is not part of the spontaneous formulation, but rather an allusion to the title of a globally successful early 1970s reggae hit, which even to those who do not know about the source has become available as a cliché.<sup>15</sup> The verbs *wanten* and *getten* are created *ad hoc* to refer back to this fixed expression.

Parallels between Sarah Kuttner’s and Money Boy’s strategies are therefore not very close. The hip hop lyrics and the novel merely show that knowledge of English is widespread enough today among German youth to make the language available for creative linguistic play. Obviously, young people who do not buy into international pop-cultural trends to the same extent as the two performers discussed here would be unlikely to avail themselves of such strategies.

## 5. English and German in language-biographical interviews with African immigrants

The present section discusses extracts from language-biographical interviews with members of Freiburg’s African community which have been collected by the author and his team since January 2017. Interviewees are recruited from among African doctoral students and members of the city’s refugee population, who are approached in German-language classes (which are often taught by student volunteers) or through other local networks such as church groups.

---

15. Written by Jimmy Cliff, performed by Cliff himself, Desmond Dekker, and others.

Interviewees are free to choose the language of the interview.<sup>16</sup> Interviews begin with structured questions about essential biographical facts, including migration history (where relevant), language competence, arrival and early stay in Germany and thence lead on to topics such as language-related misunderstandings experienced in Germany, interviewees' language learning and coping strategies, communicative networks, and so on. Interviewees are invited to bring up further topics, but are not pressed to provide information on sensitive issues such as experiences of racial discrimination and harassment during migration. The plan is to analyse the interview data both for language form and for content. English-language interviews are expected to reveal how speakers of African varieties of English accommodate to the foreign-language/lingua franca environment in Germany. German-language interviews will be analysed for German-English code-switching and direct and indirect English influence on the interviewees' learner German. In terms of content, both English and German interviews provide ethnographic information on Freiburg as a multilingual immigrant languagescape, on informants' language attitudes and ideologies, and on their local and transnational communicative networks.

To illustrate the potential of the data, I will quote from one English and one German interview. The English sample interview, 02MGAM, was conducted by Dr. Axel Bohmann (English Department, University of Freiburg) in May 2017, with a young man from The Gambia in his early twenties, who arrived as an asylum seeker in Germany in early 2015. The interview is particularly rich in biographical, ethnographic and ethnolinguistic information, as it lasted more than two hours (rather than the average 45 to 60 minutes). The interviewee is a fluent speaker of Mandinka and Wolof and reports having learnt his English in school, which he left after grade 9. According to his own words, he left The Gambia to join the migrant coach trek through the Sahara in early 2014, was left stranded for four months in Libya in late 2014, survived a traumatic ship passage to Sicily, where he stayed for six months, before he arrived in Germany by train in the spring of 2015. During his stay in Libya he picked up some Arabic, which he reports speaking in Germany with security personnel in his hostel. Some input from Italian is evident in place names mentioned in the narrative (on which see below).

With regard to language form, the speaker's English is fairly fluent, though – given his educational background – somewhat short of the acrolectal level. This is a typical passage:

---

16. In practice German elements show up in the English interviews, and the German interviews feature direct or indirect influence from English. For interviews outside the academic community, where limited competence in both English and German may be a problem, a third option is provided, namely interviews with the help of interpreters.

02MGAM: In the Sahara side also it's like that – you, sometimes you have accident. But you will be in Libya for, like, some some are there like three years, two years, four years. They are bad luck, they will not even reach in the capital Tripoli. Like when you finish [= /finis/] the desert, you have like places called /ba'he: ga'dron/, I forget the die anderes [German: “the other,” lack of agreement between plural determiner and neuter singular pronoun] other names. You reach there from there to Sabha, there's a place called Sabha, and from Sabha also to Libya. And all that is smuggling you know, you don't go there directly, you go take the highway small. The rest all is desert, so not controlled. In that desert also you have some controls there where you pay more – more money. If you don't have money, they beat you very seriously. Upon issue for dayses [hypercorrect double plural marking for *days*?] sometimes your car leave [= /lrv/] you there. You start again fresh [= /fres/], yeah. Because when you are in Africa, you heard Libya money is good, viel Arbeit [German, “a lot of work”] is there. So, here nothing – you wake up, so seeing your mates, so that is something inspired you also. You say: let me go. Then conditions are not the same. Some are living, some other some countries are living good. With some also they're not living good. So that's why many people take this hard risk journey, you know. But you, even if you heard about it, you think, even for me I think it was very easy, like I will be inside a bus for Libya. But it's not like that. You reach through Mali, you start paying, like every two hundred metres, checkpoint, you pay money too much money. [...]

Or if your car get accident you have to make call. And that call, also that's – those places it also depends where your car get breakdown. /ift/ you are in other parts, you don't get network, you must walk like six seven [hundred?] metres to have network and call, because before you leave [= /lrv/], you buy many credit units. So if you walk after that you call, so they have to prepare other things, like when they have tyre problem, they break anything, before they also reach. Some days, you don't have food in those dayses. Some died, some get so tired, ill, you know. And like for some weeks, or two weeks, or one week they don't even take shower [= /sauʔ/]. You don't brush [= /brɔs/] your mouth. Even sometimes you have a little bit water, you have your toothpaste, you cannot brush [= /brɔs/] your mouth.

(02MGAM, time in the recording [mins:secs] 44:38–47:20)

With regard to accent, which is not represented in the orthographic transcription, the extract is within the expected range. West African accents are best characterised as showing “typical features associated with the whole region, but also some sub-regional features and, predictably, some quasi-exclusive features” (Peter, Wolf & Simo Bobda 2003: 48). The interviewee's language has the expected general West African features, such as almost categorical *th*-stopping in words such as *mouth* and *that*, lots of consonant-cluster simplification, and also some neutralisation of the contrast between the vowels in *leave-live*<sup>17</sup> and *walk-work*. The most distinctive pronunciation

17. Peter, Wolf and Simo Bobda (2003: 49) do not consider this a common feature in Gambian English, as the major local contact languages, Mandinka and Wolof, have a quantitative vowel distinction here (2003: 49).



characteristic of Gambian English is the substitution of /s/ for /ʃ/, which is said to have “acquired the status of a national norm” (Peter, Wolf & Simo Bobda 2003: 50). This is due to the absence of such a distinction in Wolof and Mandinka, the two major local contact languages, and – unsurprisingly – richly attested in the interviewee’s English. The words *finish*, *fresh*, *brush* and *shower* are pronounced as /fɪnɪs/, /fres/, /brʊs/ and /sauə/. Similar treatment is given to the voiced fricative, as is evidenced by the pronunciation of *Niger* (name of the country, /niʒe/) as /nize/ (not part of the transcribed passage).

The same substitutions can also be noted in the pronunciation of a number of German words which are used in the interview. For example /ʃraŋk/ (*Schrank*, “cupboard”) is pronounced as /sraŋk/, and /ʃmerts/ (*Schmerz*, “pain”) as /smas/, with reduction of the consonant cluster. This feature is presumably unproblematical in the use of English inside The Gambia. However, it is bound to cause comprehension difficulties once Gambian English is on the move to Europe. Lacking the appropriate frame of reference, Germans are unlikely to decode expressions such as *take sour* (“take a shower”) easily in lingua franca interaction. Given that /s/ is a common substitution for the voiceless dental fricative /θ/ in German learner English, some confusion may also arise in the reverse direction, particularly if minimal pairs such as *thank-sank* or *thick-sick* are concerned:

In terms of linguistic ethnography, one of the many things which deserve comment is how the interviewee deploys his multilingual resources to map a difficult and confusing journey from The Gambia to Germany to himself, and to the listener. Evidently, he did not start out with a fixed English-language toponymy which would have prepared him for a journey through *Sicily* in *Italy* to *Munich* in Germany. Sicily and Italy are usually referred in Italian, as *Sicilia* (with the internal affricate /tʃ/ being rendered as /s/). For Munich, the German *München* is used, in a learner pronunciation (/muniçen/) which is rather different from the target /myŋçn/. Mobile speakers with partly truncated linguistic repertoires need to draw on resources from several languages to turn space into place.

Legal status, life experience in Germany and language skills are rather different in the case of the second interviewee, a female doctoral student from Nigeria, with Igbo as her native language, who chose to be interviewed in German. The following extract shows the speaker operating in three communicative domains with drastically different language requirements: (1) English-dominant academia, (2) the largely English-free environment of an old people’s residential home, and (3) a multilingual restaurant kitchen. For better orientation, the relevant portion of the interview, which was conducted in German, is given in a summary English translation before I move on to a detailed analysis of the original text. The speaker is identified by the code 02FNG assigned to the interview:



[Context: The interviewee discusses the usefulness of English in the university and in two places in which she does temporary work to supplement her income, a residential home for the elderly and a restaurant.]

**02FNG:** Initially I always used German and English together. When I wanted to ask something and didn't know the word in German, I said it in English first, and the other person perhaps understood a little English as well and said: "Oh, do you mean that?" And I said: "Yes, that's what I mean!"

*I: Do you feel that Germans speak English well enough to understand what you mean, when you mix English and German?*

**02FNG:** It depends on the people I'm talking to. I think everybody has learned a bit of English in school or in college, but when I'm in the old people's home, many of them don't understand English. We have no contact in English.

*I: And in the kitchen of [NAME OF RESTAURANT]?*

**02FNG:** In the kitchen it's nice, because many of them are students. There are so many students there, and sometimes we speak English.

*I: And how do you speak to your Nigerian colleague, the cook?*

**02FNG:** In Igbo. He's Igbo. We sometimes speak Igbo.

*I: Do the two of you also speak to each other in Pidgin?*

**02FNG:** Yes.

*I: And in English?*

**02FNG:** Yes we do. Once we spoke in Pidgin and a colleague who was there said: "What language is that?" And we said: "English." And he said: "No, that isn't English. I want to understand what you're saying."

*I: [...] So you speak English, German, Igbo and Pidgin. Can you say a few sentences, to show me what it sounds like when you and the cook talk. [...]:*

**02FNG:** Maybe an order has just come in and I say: "Chef, dem want chop-o [they want to eat]." And he says: "Wetin dem want chop [what do they want to eat]?" [And I say] "Chef, na you side-o, no be salad, na you side [it's your department, it isn't salad, it's your department]." Then he will take a look at the order and prepare the food.

Interviews conducted in German can be used to study the influence of ESL English on the acquisition of L3 German.<sup>18</sup> Below is the German original of the passage summarised above in orthographic transcription, with a short passage added at the beginning. Typical features of colloquial German, such as loss of the inflectional ending *-e* in verbs in the first person singular of the present tense (*ich sage* → *ich sag*), reduced pronunciations of grammatical words (*nicht* → *nich*), and regional variants (*das* → *des*) are italicised. If the pronunciation of German words shows influences from formally similar English equivalents, this is indicated in square brackets. For example, *Studenten* indicates a pronunciation of the initial cluster as [ʃt], whereas the anglicised pronunciation is rendered as [st]udenten.

18. Parallel research on Igbo speakers' Italian has been carried out by Goglia (2009).

02FNG: ja

am Anfang als ich angefangen bin

ich *hab* immer <hesitation> Deutsch und Englisch zusammen so  
wenn ich will was fragen und ich kenne diese Wort *nich* auf Deutsch  
dann *sag* ich erst auf Englisch

und vielleicht diese Person auch verstehen ein bisschen [1]nglisch  
und sagen

oh meinst du *des* da

Ja das mein ich ja

*I: Hast Du das Gefühl, dass die Deutschen gut genug Englisch können, um Dich zu verstehen – wenn Du Englisch und Deutsch mischt?*

02FNG: <hesitation> manchmal

kommt darauf

die Leute das ich rede mit

wenn sie sind wenn sie sind auf <hesitation> in der Uni es ist OK

weil sie verstehen

ich glaube alle hat ein bisschen [1]nglisches gelernt auf Hochschule oder Grundschule  
aber die alte Leute wenn ich bin Altenheim und viele versteh kein [1]nglisches ich  
*denk* wir haben kein Kontakt

*I: Und in der Küche in der Küche bei [NAME OF RESTAURANT]?*

02FNG: <laughter> in der Küche es ist schön

weil viele viele sind [st]udenten

in diese dieser Betrieb wir haben soviel [st]udenten

und manchmal reden wir auf [1]nglisch

*I: Und wie redest Du mit Deinem nigerianischen Kollegen, dem Koch in der Küche?*

02FNG: der Koch

auf Igbo

er ist auf Igbo <hesitation> von Igbo

wir reden manchmal auf Igbo

*I: Sprecht ihr Nigerian Pidgin, Du und er?*

02FNG: ja ja

*I: und Englisch?*

02FNG: und Englisch.

einmal haben wir auf Pidgin geredet und eine Kollege auch war da

und sagt Äh? was für Sprache ist das?

wir haben gesagt wir reden auf [1]nglisches

er sagt nein das ist kein englisches

ich will verstehen

*I: Da muss es ja bei Euch in der Küche lustig zugehen. Ihr redet Englisch, Deutsch, Igbo und Pidgin. Kannst Du mal ein paar Sätze sagen, wie das klingt, wenn Du mit dem Koch redest. I want to get out your acting talent now [...]:*

02FNG: er sagt

vielleicht eine Bon ist gekommen und dann ich *sag*

chef dem want chop-o

<I & 02FNG: laughter>

und er sage  
 wetin dem want chop  
 chef na you side-o  
 no be salad na you side  
 dann kommt er schaut Bon an und macht er das Essen

(time in the recording [mins:secs] 11:29 – 15:07)

In young African academics' lives, official languages such as English and French have a particularly strong presence vis à vis the mother tongues, which represents an interesting, yet under-researched constellation for research on L3 or Ln acquisition.<sup>19</sup>

English interference is obvious in the frequent vacillation between English and German initial vowels in *Englisch*, a word which is mostly used in the fossilized form of *Englisches* by this speaker (“Pidgin ist kein Englisches ...”, “auf Englisches ...”). Similarly, the initial cluster of *Student* is often articulated as [st]. This speaker's mother tongue is Igbo, which – like English – has SVO basic word order. Thus, both languages can be held responsible for the common use of SVO order in the place of German main-clause verb-second order.

Verb-second is indeed a major challenge for foreign learners of German of most L1 backgrounds. The finite verb occurs in second position in the main clause, regardless of what is the first constituent. *Ich las das Buch gestern* (= “I read the book yesterday”) is fine, and so are *das Buch las ich gestern* and *gestern las ich das Buch*. What is not possible is \**gestern ich las das Buch* (i.e. the structural parallel to English “yesterday I read the book”) or \**das Buch ich gestern las*. As can be seen from the transcript, this learner applies the verb-second order only very unsystematically, usually in highly frequent frames such as *dann sag ich ...* (“then say I ...”). Mostly, however, XSV and XSVO are not transformed to correspond to the German standard norm:

- (1) vielleicht diese Person auch verstehen ein bisschen [ɪ]nglisch  
 perhaps this person also knows a little English  
 [correct version:<sup>20</sup> vielleicht versteht diese Person auch ein bisschen Englisch]
- (2) in diese dieser Betrieb wir haben soviel [st]udenten  
 in this workplace we have so many students  
 [correct version: in diesem Betrieb haben wir so viele Studenten]

19. If Pidgin is subsumed under the “English Language Complex,” L3 is the appropriate category for the present interviewee. As many students are multilingual in African indigenous languages, Ln is more appropriate in general terms.

20. Many of the inflectional endings in German nouns, verbs and adjectives are wrong in the original. This is not relevant in the present context, and such errors are tacitly corrected.

As both Igbo and English have SVO syntax and are strongly analytical, it is often impossible to distinguish transfer from one or the other language. In other words, this type of error would be expected in monolingual Igbo speakers' German as well as in monolingual English speakers'. Errors in specific constructions, however, do point to the specific role of English, for example the following clause, which is modelled closely on English preposition stranding constructions:

- (3) die Leute das ich rede mit  
 the people that I speak with  
 [correct version: die Leute, mit denen ich rede, i.e. "the people with whom I speak"]

Another clear case of interference from English is the following German calque on the English present perfect (19:40 in the same interview):

- (4) für ungefähr zwei Jahre ich habe keine englische Fernsehen gesehen  
 for about two years I have no English television watched  
 [correct version: seit ungefähr zwei Jahren schaue ich kein englisches Fernsehen]

Formally, this sentence echoes English "For about two years I have not watched any English television" – an idea which is rendered in the present tense in German and many other European languages, as "seit zwei Jahren schaue ich kein englisches Fernsehen mehr." Note also that if a time interval is expressed which began in the past and extends to the moment of speaking, the preposition of choice in German is *seit*, regardless of whether the following noun refers to a point in time ("seit 2015" = "since 2015") or a period of time ("seit zwei Jahren" = "for two years").<sup>21</sup> Like many other main clauses in the text, this one has not got verb-second order, either (*\*für ungefähr zwei Jahre ich habe ... → für ungefähr zwei Jahre habe ich ...*). On the other hand, exbraciation, the discontinuous German "Satzklammer" between the auxiliary (*habe*) and the lexical verb (*gesehen*), has been mastered.

Complex though it may be, the investigation of transfer, interference and other learner features is merely a first step in the analysis, and the data should not be reduced to this dimension. Equally necessary is a sociolinguistic analysis which focusses on the way speakers use multilingual and often "truncated" repertoires (Blommaert 2010: 103–106) creatively to cope with the challenges of the migrant situation. Obvious cases in point are *ad-hoc* insertions of English lexical items such as "weil das ist unsere *official* Sprache [because that is our official language]" (2:30)

21. *Für ungefähr zwei Jahre* exists, but has a different meaning, namely "for any period of about two years," and not "two years in relation to the moment of speaking" or "two years in relation to a contextually given other point of time." For an example, consider *sie sperrten ihn für ungefähr zwei Jahre ein* (= "they locked him away for about two years").

or “*full mit Wasser*” [full of water]” (35:00). More intriguing are cases such as “Die Google Translate manchmal gibt eine nicht richtige Translation [translatʃo:n]” (17:50), where, probably helped by English *translation*, the rare and formal German word is articulated correctly, but of course completely out of place in the informal spoken register, in which *Übersetzung* would be expected. This creativity routinely produces uses of language in which the question of right or wrong becomes irrelevant. In the extract reproduced above, for example, the word *Chef*, used to refer to the Nigerian cook, evokes several meanings simultaneously, and all of them fit the situation. The recorded pronunciation ([ʃef]) is compatible both with German *Chef* (= “Boss”) and French/German/English *chef* (= “master cook”). Ultimately, what such examples show is that it is at least as instructive to look at the German data not as mere learner language, but as socially embedded successful communication by means of truncated repertoires. On the individual level, observations on points of linguistic detail add up to documenting processes of language acquisition, in sometimes difficult and disadvantageous circumstances. On the sociolinguistic level, they show how the world’s *lingua franca* leaves its indelible imprint in the transidiomatic practices which emerge in diasporic communication, among migrants themselves and among migrants and the German resident population:

Transidiomatic practices are often the products of linguistic innovations grafted onto an English structure, but any number of other languages could be involved in these recombinations. The social world is increasingly composed of settings where speakers use a mixture of languages in interacting face-to-face with known and unknown people; these settings become “transidiomatic” when the participants *habitually* read English and/or other global languages on their computer screens, watch local, regional, or global broadcasts, listen to pop music in various languages, and interact via cellular phones with nonpresent contacts. In these environments, speakers use mobile, real-time communication devices (from laptops to cell phones to wi-fi enabled tablets) to enhance everyday social interactions, producing a massively fluid, layered communicative style that relies on access to multiple communicative channels to achieve its shape. (Jacquemet 2015: 342)

This is the truly globalised, rather than the merely postcolonial dimension of the use of English as a world language which, as I have argued above, will require more attention in future research on World Englishes.

## 6. Conclusion

The four domains of regular English-language use in Germany distinguished above (business and industry, academia, pop culture, new migrants) differ significantly with regard to the extent to which they subscribe to ideologies of monolingualism. In the two “elite” professional domains, speakers tend to keep English and German separate. Mixed language use is not generally encouraged. In the two “non-elite domains,” by contrast, multilingualism and language mixing are tolerated or even celebrated. Cutting across the “elite”–“non-elite” distinction is the classification based on communicative usefulness. Use of English is largely driven by communicative needs in business, academia and among migrant communities. In youth culture, on the other hand, use of English often takes the shape of staged or otherwise mediated performance in communities of practice whose members are German-dominant offline and offstage.

With the expansions and modifications suggested above, the “World System of Englishes” model has proved fully adequate to the complex and openly or covertly multilingual data to be accounted for in the present study. In its first programmatic presentation (Mair 2013), the model postulated the role of a “central” nonstandard variety for Nigerian Pidgin. The continuing usefulness of Pidgin in the migration diaspora was empirically confirmed from the interview with the doctoral student from Nigeria. Gambian English, which in terms of the model is best classified as a peripheral standard, lacks supra-national prestige in the West African context and is a totally unknown quantity in the German ELF environment. Any prestige it may have had in the national Gambian context is bound to dissipate quickly in migration, and the interview with the young Gambian refugee displayed specific phonetic features which were likely to even become dysfunctional by giving rise to misunderstandings in the new environment. On a more general level, the model posits “regionally restricted and domain-specific ELF uses.” The example given is “Euro-English,” i.e. the type of *élite* lingua franca associated with politicians and professionals working in and for the European institutions.<sup>22</sup> In the EU institutional environment, a high degree of competence in English can be assumed for

---

22. An idea of “Euro-English” can be gleaned from Gardner (2016), a 59-page-long brochure listing a large number of English words and expressions with limited currency outside the EU context. Many of these usages are likely to be due to contact influence from French, another major EU working language. Others are well-known lingua franca features, such as the pluralisation of nouns which are “non-count” in the major L1 varieties. As the brochure is written in a moderately prescriptive spirit, the institutional legitimacy of Euro-English as a lingua franca is, of course, not a concern.

the higher echelons in the bureaucracy. When it comes to communication between African refugees and German administrators and social workers at the local level, this degree of competence cannot be taken for granted in all cases; the emerging lingua franca will reflect these limitations and deserves description in its own right. A central claim of the model concerns the major role of nonstandard varieties, often through media dissemination, in the global spread of English, the example being African American Vernacular English, in the shape of its mediated derivative Hip-hop Nation Language. While this does not play a major role in the two interviews selected for discussion here, the data as a whole support this claim, too.

To conclude with the obvious, the entire constellation investigated here is situated in the communicative space “beyond and between the Three Circles” (Mair 2016) of Kachru’s model. It takes up the story of English after the first two crucial chapters, namely (i) colonial spread and imposition and (ii) postcolonial appropriation and adaptation in the context of identity formation and nation-building. In this third chapter, the story of English moves from the long shadow of colonialism to contemporary dynamics of globalisation. In Germany as well as many other countries without a colonial past, English is challenging the traditional monolingual sociolinguistic order of the nation state. It undermines it from without, in that the increasing use of English is an inevitable response to increasingly globalised “markets” in business, higher education and pop culture and the entertainment industry; and it undermines it from within, because the increasing linguistic diversity of the resident population challenges the near monopoly German used to have in public communication and in the media. In the past few decades, English has been firmly integrated into the increasingly multilingual languagescape of contemporary Germany. In its turn, this has integrated Germany more closely into several fast-evolving transnational languagescapes shaping our globalising world.

## References

- Alim, H. S. 2015. Hip hop nation language: Localization and globalization. In *The Oxford Handbook of African American Language*, J. Bloomquist, L. J. Green, & S. L. Lanehart (eds), 850–863. Oxford: OUP. <https://doi.org/10.1093/oxfordhb/9780199795390.013.49>
- Alim, H. S., Ibrahim, A. & Pennycook, A. (eds). 2009. *Global Linguistic Flows: Hip Hop Cultures, Youth Identities, and the Politics of Language*. New York NY: Routledge.
- Ammon, U. 2006. The status and function of English in Germany. *Revista Canaria de Estudios Ingleses* 53: 27–33.
- Appadurai, A. 1990. Disjuncture and difference in the global cultural economy. In *Global Culture*, M. Featherstone (ed.), 295–310. London: Sage.
- Appadurai, A. 1996. *Modernity at Large: Cultural Dimensions of Globalization*. Minneapolis MN: University of Minnesota Press.



- Auer, P. 1999. From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism* 3: 309–332. <https://doi.org/10.1177/13670069990030040101>
- Berns, M. 1988. The cultural and linguistic context of English in West Germany. *World Englishes* 7: 37–49. <https://doi.org/10.1111/j.1467-971X.1988.tb00213.x>
- Blommaert, J. 2010. *The Sociolinguistics of Globalization*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511845307>
- Buschfeld, S. & Kautzsch, A. 2017. Toward an integrated approach to postcolonial and non-postcolonial Englishes. *World Englishes* 36: 104–126. <https://doi.org/10.1111/weng.12203>
- Coleman, J. 2014. Understanding slang in a global context. In *Global English Slang: Methodologies and Perspectives*, J. Coleman (ed.), 1–14. London: Routledge.
- de Swaan, A. 2001. *Words of the World: The Global Language System*. Cambridge: Polity.
- de Swaan, A. 2010. Language systems. In *The Handbook of Language and Globalization*, N. Coupland (ed.), 56–76. Malden MA: Blackwell. <https://doi.org/10.1002/9781444324068.ch2>
- Dovchin, S. 2017. Translocal English in the linguascape of Mongolian popular music. *World Englishes* 36: 2–19. <https://doi.org/10.1111/weng.12189>
- Eckert, P. 2012. Three waves of variation study: The emergence of meaning in the study of variation. *Annual Review of Anthropology* 41: 87–100. <https://doi.org/10.1146/annurev-anthro-092611-145828>
- Gardner, J. (ed.). 2016. *Misused English Words and Expressions in EU Publications*. Brussels: European Court of Auditors. <[http://www.eca.europa.eu/Other%20publications/EN\\_TERMINOLOGY\\_PUBLICATION/EN\\_TERMINOLOGY\\_PUBLICATION.pdf](http://www.eca.europa.eu/Other%20publications/EN_TERMINOLOGY_PUBLICATION/EN_TERMINOLOGY_PUBLICATION.pdf)>
- Gazzola, M. & Wickström, B-A. 2016. Introduction. In *The Economics of Language Policy*, M. Gazzola & B-A. Wickström (eds), 1–18. Cambridge MA: The MIT Press.
- Gazzola, M. & Wickström, B-A. (eds). 2016. *The Economics of Language Policy*. Cambridge MA: The MIT Press.
- Götter, K-H. 2013. *Abschied von Mutter Sprache: Deutsch in Zeiten der Globalisierung*. Frankfurt: Fischer.
- Goglia, F. 2009. Communicative strategies in the Italian of Igbo-Nigerian immigrants in Italy: A contact-linguistic approach. *Sprachtypologie und Universalienforschung* 62: 224–240.
- Gürtler, K. & Kronewald, E. 2015. Internationalization and English-medium instruction in German higher education. In *English-Medium Instruction in European Higher Education*, S. Dimova, A. K. Hultgren, & C. Jensen (eds), 89–114. Berlin: De Gruyter Mouton.
- Gundermann, S. 2014. *English-medium instruction: Modelling the role of the native speaker in a lingua franca context*. Freiburg: Universitätsbibliothek. <<https://freidok.uni-freiburg.de/data/9795>>
- Hilgendorf, S. K. 2007. English in Germany: Contact, spread, and attitudes. *World Englishes* 26: 131–148. <https://doi.org/10.1111/j.1467-971X.2007.00498.x>
- Hilgendorf, S. K. 2010. English and the global market: The language's impact in the German business domain. In *Language and the Market*, H. Kelly-Holmes & G. Mautner (eds), 68–80. Houndmills: Palgrave-Macmillan. [https://doi.org/10.1007/978-0-230-29692-3\\_7](https://doi.org/10.1007/978-0-230-29692-3_7)
- Hilgendorf, S. K. 2013. Transnational media and the use of English: The case of cinema and motion picture titling practices in Germany. In *Neue Medienpraktiken: Die sprachliche Dimension / New Media Practices: The Language Dimension / Nouvelles pratiques de média: La Dimension linguistique* [Sociolinguistica 27], U. Ammon, J. Darquennes, & S. Wright (eds), 167–186. Berlin: De Gruyter.



- Hilgendorf, S. K. & Erling, E. J. 2006. Language policies in the context of German higher education. *Language Policy* 5(3): 267–293. <https://doi.org/10.1007/s10993-006-9026-3>
- Jacquemet, M. 2015. Language in the age of globalization. In *Routledge handbook of linguistic anthropology*, N. Bonvillain (ed), 327–347. London/New York, Routledge.
- Kachru, B. 1985. Standards, codification, and sociolinguistic realism: The English language in the Outer Circle. In *English in the World: Teaching and Learning the Language and Literatures*, R. Quirk & H. G. Widdowson (eds), 11–30. Cambridge: Cambridge University Press.
- Kachru, B. 1992. World Englishes: approaches, issues and resources. *Language Teaching* 25: 1–14.
- Kautzsch, A. 2014. English in Germany. Spreading bilingualism, retreating exonormative orientation and incipient nativization? In *The Evolution of Englishes: The Dynamic Model and Beyond* [Varieties of English Around the World G49], S. Buschfeld, T. Hoffmann, M. Huber & A. Kautzsch (eds), 203–227. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g49.12kau>
- Lawson, S. & Jaworski, A. 2007. Shopping and chatting: Reports of tourist-host interaction in The Gambia. *Multilingua: Journal of Cross-Cultural and Interlanguage Communication* 26: 67–93. <https://doi.org/10.1515/MULTI.2007.003>
- Loven, K. 2009. *Watching Si Doel: Television, Language and Cultural Identity in Contemporary Indonesia*. Leiden: Brill.
- Mair, C. 2013. The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34: 253–278. <https://doi.org/10.1075/eww.34.3.01mai>
- Mair, C. 2016. Beyond and between the Three Circles: World Englishes research in the age of globalisation. In *World Englishes: New Theoretical and Methodological Considerations* [Varieties of English Around the World G57], E. Seoane & C. Suárez-Gómez (eds), 17–36. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g57.02mai>
- Mair, C. 2017. World Englishes in cyberspace. In *Cambridge Handbook of World Englishes*, D. Schreier, M. Hundt, & E. W. Schneider (eds). Cambridge: CUP.
- McArthur, T. 2003. World English, Euro-English, Nordic English. *English Today* 19(1): 54–58. <https://doi.org/10.1017/S0266078403004097>
- Mesthrie, R. & Bhatt, R. M. 2008. *World Englishes: The Study of New Varieties*. Cambridge: CUP.
- Northrup, D. 2013. *How English Became the Global Language*. Houndmills: Palgrave Macmillan. <https://doi.org/10.1057/9781137303073>
- Pennycook, A. 2003. Global Englishes, Rip Slyme, and performativity. *Journal of Sociolinguistics* 7: 513–533. <https://doi.org/10.1111/j.1467-9841.2003.00240.x>
- Peter, L., Wolf, H-G. & Simo Bobda, A. 2003. An account of distinctive phonetic and lexical features of Gambian English. *English World-Wide* 24: 43–61. <https://doi.org/10.1075/eww.24.1.04pet>
- Pratt, M. L. 2011. Comparative literature and the global languagescape. In *A Companion to Comparative Literature*, A. Behdad & D. Thomas (eds), 273–295. Malden MA: Blackwell. <https://doi.org/10.1002/9781444342789.ch18>
- Schneider, E. W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511618901>
- Schneider, E. W. 2014. New reflections on the evolutionary dynamics of World Englishes. *World Englishes* 33: 9–32. <https://doi.org/10.1111/weng.12069>
- Schröder, A. & Rudd, P. W. 2017. Language mixing and ecology in Africa: Focus on Camfranglais and Sheng. In *Linguistic Ecology and Language Contact*, R. Ludwig, S. Pagel & P. Mühlhäusler (eds). Cambridge: CUP.
- Smith, Z. 2016. *Swing time*. Harmondsworth: Penguin Books.

- Spitzmüller, J. 2005. *Metasprachdiskurse. Einstellungen zu Anglizismen und ihre wissenschaftliche Rezeption*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110201987>
- Tsunoda, M. 1993. Les langues de publications scientifiques au 20e siècle. In *Actes du XV<sup>e</sup> Congrès de Linguistes* [Québec 1992], Vol IV, A. Crochetière, J-C. Boulanger, & C. Ouellon (eds), 43–46. Quebec: Presses Universitaires Laval Sainte-Foy.



# Hybridity, globalisation and models of Englishes

## English in South African multilingual digital repertoires

Bertus van Rooy<sup>1</sup> and Haidee Kruger<sup>1,2</sup>

<sup>1</sup>North-West University / <sup>2</sup>Macquarie University

Current models of Englishes face empirical challenges, such as multilingualism, hybrid varieties, complex identities, online communication and other consequences of globalisation, alongside a number of conceptual and methodological challenges. In this chapter, we explicate these challenges, and offer a corpus-linguistic analysis of interactive, online data from South Africa, in an attempt to expand current models. The data reveal that a shared pool of English resources form the core of the online interactions, but this is supplemented by resources from a global online repertoire, global and local non-standard English forms, and forms from other South African languages. Users make selections from these resources, to communicate and align with other users, and in the process, express their hybrid, complex identities by combining these resources. An adequate model has to provide for a diverse resource pool, selection processes and the eventual diversity of text types emerging from communicative interactions.

**Keywords:** computer-mediated communication, South Africa, globalisation, multilingual repertoires

### 1. Introduction

A core theme in world Englishes scholarship is that the spread of English has led to a diversification of forms and functions in the language, which various models aim to describe and explain. Colonial expansion is usually identified as the first major force that led to the transplantation of English beyond its original territory, followed by the continued use of English in former British colonies after independence in the second half of the twentieth century (Schneider 2007: 1). The influential models of

Kachru (1985) and Schneider (2003, 2007) address this process, but with slightly different goals.

Kachru's work, especially in the foundational stage in the 1970s (collected in the 1986 volume *The Alchemy of English*) is to a significant extent directed at pedagogical problems, in particular the question of what constitutes valid norms for language education in newly independent postcolonial states. Kachru (1985) develops the model of three concentric circles, where he advocates the distinction between an Outer Circle that is norm-developing and on its way to establishing endonormativity, and an Expanding Circle, which is norm-dependent and remains exonormative in its reliance on the norms of the Inner Circle (Kachru 1985: 160). While he recognises variability in English and its embeddedness in local multilingual ecologies, he places strong emphasis on an educated variety of English as the norm for the local system, as opposed to the usage of less proficient users.

Schneider's (2003, 2007) model covers the varieties that the Inner and Outer Circle cover in Kachru's (1985) model. Schneider (2007: 5) argues that there is a uniformity to be observed in the processes and the sequence of developments that he uncovers across different postcolonial settings, which he captures with the Dynamic Model of postcolonial English. The model seeks to generalise about and explain linguistic outcomes in a more comprehensive way than the largely pedagogical and normative focus of Kachru's Concentric Circle Model allows.

There are three kinds of challenges to existing models of Englishes. In the first instance, as Schneider (2014, 2016) and Mair (2013a) note, there are **empirical challenges**: As the world changes, and the use of English in the world changes, models that were adequate at a certain point in the past are no longer adequate. For instance, certain similarities between Outer and Expanding Circle Englishes have been identified (e.g. Buschfeld & Kautzsch 2017; Edwards & Laporte 2015) that are not directly accounted for in the Dynamic Model. The second challenge is **conceptual**: Seargeant and Tagg (2011: 496) question whether the variety-centric approach is in principle adequate to provide an account of the use and spread of English in multilingual settings, based on Blommaert (2010: 4–5), who rejects what he terms the artefactual view of language as static and reified, and unable to account for the language repertoires and language use of multilingual people in a mobile world. The third challenge is **methodological**: On the one hand, quantitative data from corpora or surveys typically isolate English from the multilingual settings in which it naturally occurs; on the other, attempts to capture the "wider semiotic repertoire" (Seargeant & Tagg 2011) of multilingual users more accurately have mainly proceeded qualitatively. This leaves unanswered questions about the representativeness of the conclusions, which are a precondition for a model that makes any claims to wider generality, but conversely, it is not necessarily given that a quantitative approach would be methodologically compatible with the complexity and nuance of the data.

These challenges have been partially addressed by recent proposals for revised models. Mair's (2013a) concept of a world system of Englishes is one attempt to expand the scope of models of Englishes. This model gives recognition to the power inequalities and status differences between standard varieties in different countries, and also incorporates non-standard varieties. The direction of influence among different varieties, and the hybrid communication strategies in globalised communication in online forums also receive attention. "Transnational attraction", proposed by Schneider (2014: 28), is construed as an attractor within a dynamic field, understood metaphorically in terms of chaos theory, but this is not yet embedded in a larger model. However, these proposals are still premised on the concept of "varieties" as unit of analysis, which is what Seargeant and Tagg (2011) call into question.

The present contribution aims to add to recent reflections on ways of modelling the global diffusion of English, and to offer a concrete proposal for the understanding of the interaction of linguistic resources in multilingual communicative settings. The empirical focus of the contribution is South Africa, a multilingual country that continues to pose a challenge to existing models of Englishes due to the demographic complexity of multilingualism alongside a sizeable native-speaker population, and its ambivalent embeddedness in the larger African and global environment.

The next section analyses the three main categories of challenges to current models of Englishes outlined above, in order to articulate the challenges that a new proposal for models of Englishes needs to meet. Thereafter, the research method is presented, followed by a discussion of the findings from the empirical study, which leads to an evaluation of the findings as a means to offer a solution (if necessarily partial).

## 2. Recent challenges to modelling Englishes

### 2.1 Empirical challenges

While the thrust of the world Englishes enterprise has been to expand the notion of what counts as "Englishes", there has been a strong emphasis on the educated local varieties, without affording non-standard varieties or informal contexts of communication a place in the models (Mair 2013a: 254–255). Mair (2013a: 257–259) proposes to include non-standard varieties, and at the same time, to consider the different degrees of centrality of standard and non-standard varieties alike in the model of a world system of Englishes, which is necessary since the non-standard features also participate in global diffusion.

Taking a further step, Mair (2016, forthcoming) points out that a national variety is no longer the only possible outcome of linguistic decolonisation, particularly for states where a kind of national identification is not well established, or where

the provision of education is inadequate to give sufficient access to the local standard, and the development trajectory of a local standard might be overtaken by the forces of globalisation. In diasporic communities, furthermore, Mair (2013b, 2016) illustrates that Nigerian Pidgin and Jamaican Creole are more important identity markers than the standard(ising) varieties of the countries of origin.

The importance of non-standard varieties as identity markers raise important questions about identity. A national, collective identity is central to both Kachru and Schneider's models of Englishes, and a nativised, endonormative English is seen as an expression of such an identity. Such an identity is construed in the first place as a statement against the colonial masters, an act of national (and in its wake linguistic) self-determination. Such stable, national identities are exactly what globalisation calls into question. Behind such a national identity construct and its association with an educated, indigenised form of English lies not only the exclusion of users who do not have access to the resources of the prestige variety of the language and who develop their own linguistic practices, drawing on the resources at their disposal (Blommaert 2010: 80–101), but also the non-recognition of the diverse linguistic practices across a range of contexts of all multilingual English speakers, including those who have access to the resources of a standard variety. Moreover, Coetzee-Van Rooy (2014b) and Makalela (2014) point out that the multiplex identity construction of multilingual citizens does not hinge on English, and that other languages play a more central role than English does for the majority of the participants in their studies.

The most dramatic empirical challenge to models of Englishes, however, is accounting for the consequences of globalisation. Two very pertinent challenges to current models of Englishes merit attention:

- the severing of a stable link between language and community, in favour of a more fluid and mobile view of linguistic resources; and
- new forms of widespread communicative practice enabled by interactive online technologies.

Both of these aspects escape analytical frameworks that attempt to define the linguistic varieties of identified speech communities in terms of communicative norms or conventions that stabilise over time.

The “national focus” of conventional work on models of Englishes, according to Mair (2013a: 255–257), has to yield to post-national approaches, where individuals draw on a wide range of resources, including global and local Englishes and other languages and styles, to project their identities, which are much less fixed and stable than traditionally conceived of by models of Englishes, as outlined above.

Computer-mediated communication (CMC) is a very important aspect of globalisation, which enables extensive interaction between individuals who are not in

face-to-face contact. Not only may online communities that develop some shared practices emerge, but novel linguistic forms may also develop and subsequently spread to other registers. Seargeant and Tagg (2011: 501–503) note that the earliest research on CMC adopted a varieties approach, in which electronic communication was approached as another variety of English with its own emergent linguistic features. However, they argue that the use of CMC leads to much more diverse and complex patterns, because various resources are drawn into actual communicative practice, of which the semiotic possibilities offered by CMC is but one. Mair (2013a) draws attention to the way online communication also acts as agent of diffusion, alongside other global media, of linguistic resources beyond the communities/ places of origin, becoming “a second site of super-diversity” (Mair 2013a: 257). However, Biber and Egbert (2016: 96) point out that much of the descriptive linguistic research on CMC has focussed on the affordances and concomitant linguistic innovations of the new medium, whereas a general quantitative picture of its properties is still lacking.

In summary, the empirical challenge to models of Englishes is the extent to which other languages, non-standard Englishes, CMC, hybridity, multiplex identities, and trans-nationalism can be incorporated. The sources of linguistic innovation, but also the processes by which linguistic features disseminate are both affected by the consequences of globalisation and technology. If it is not possible to extend current models to cover the wider linguistic ecology of which the standard(ised) local Englishes form part, then the search is on for a new model if one wants to increase the adequacy of the explanation for ongoing changes in the form and use of English.

## 2.2 Conceptual challenges

Models of Englishes are based on explicit or implicit assumptions about the phenomenon of language and the differentiation among languages. Seargeant and Tagg (2011) present a challenge to conceptualisations that underpin current models, drawing on Blommaert (2010) and Seargeant (2010). At the heart of this challenge is the claim that existing scholarship on world Englishes assumes the existence and development of reasonably stable varieties that are located in reasonably stable communities. Seargeant and Tagg (2011: 498) acknowledge the contribution of world Englishes scholarship to challenge the myth of a single, unified English language, and to legitimise its various local forms. However, they argue that a fixed code view underlies the linguistic analysis in the varieties-based approach, where researchers attempt to discover “a recognisable system of linguistic features which can be associated with a community of speakers” (Seargeant & Tagg 2011: 499). As a consequence of the more fluid, dynamic practices and trans-national diffusion of linguistic



resources, a model is required that does not operate on the assumption of a stable link between linguistic practices (“language”) and a delineated speech community.

Seargent (2010) argues that stable linguistic conventions are imposed by linguistic analysis, based on ontological assumptions that derive from particular historico-cultural perspectives, such as the romantic nationalism of Herder. Otsuji and Pennycook (2010), arguing from a similar angle, note that fluidity and fixity are in constant tension in actual communicative interactions. They propose the concept of metrolingualism, whose “focus is not on language systems but on languages as emergent from contexts of interaction” (Otsuji & Pennycook 2010: 246). However, in the actual analysis of data, both Otsuji and Pennycook (2010) and Seargent and Tagg (2011) continue to denote particular linguistic features in terms of the named languages they are commonly associated with, as Blommaert (2010) largely does as well.

Benor (2010) examines a related notion, the ethnolect, and points out that speakers who are associated with “ethnic dialects” differ from one another in the degree to which their speech display the distinctive ethnic markers, and also display intra-speaker variation. She proposes as solution that the reified ethnolectal concept should be replaced by the notion of an ethnolinguistic repertoire, from which speakers can draw to varying degrees. However, at the same time, Benor (2010: 170–173) identifies an irresolvable analytic challenge, which is the need to assume some abstract or concrete norm in terms of which the use of distinctive features making up the ethnolinguistic repertoire can be identified, and the density of their selection be quantified. This is so, because she believes that “it is impossible to discuss ethnic language variation without incorporating the notion of markedness” (Benor 2010: 172).

Meierkord (2012: 195–199) agrees that the consequence of globalisation and migration is that the stable social structures in which linguistic conventions traditionally developed have been dissolved. However, she cautions that outcomes of globalisation are not uniform, and that this dissolution of structures and its effect on languages is a matter of degree. While intra-national interactions across Englishes, such as the ones she investigates in South Africa (Meierkord 2012: 95–131), do yield stable and converging linguistic outcomes, due to the relatively higher stability of the patterns of social interaction, international interactions (Meierkord 2012: 158–194) are of a more fleeting nature, often once-off, and do not lead to similar stable linguistic patterns.

Thus, while we agree in principle with the emergentist view espoused by Blommaert (2010), Otsuji and Pennycook (2010), and Seargent (2010), we do not assume that this is incompatible with the possibility implied by Meierkord’s (2012) work that the degree of stability in linguistic conventions is correlated with the degree of continuity in the composition of the interaction participants. Likewise, following Benor (2010), we allow for the possibility that users behave in terms of a

relatively neutral shared code against which particular choices from other parts of their repertoire can be interpreted as more salient identity markers. We leave this question open to be decided by the patterns that emerge from the data, rather than to assume a particular answer in advance.

### 2.3 Methodological challenges

Previous research aimed at meeting some or most of the challenges posed by a changing, globalising world and reified concepts of language tends to focus on exemplars of usage, aimed at identifying the range of resources that have been used. The study of hybrid languages in face-to-face usage is characterised by a strong qualitative approach, as is evident in Mesthrie and Hurst (2013) and Makalela (2014) on Tsotsitaal and Kasi-taal, for instance. The same is true for most research on mobile communication, which tends to rely on qualitative analyses supplemented by limited quantitative analyses (as in Crystal 2011; Deumert 2014; Mair 2013b; but see Meierkord 2012 and Mair 2016 for exceptions).

The problem is that to the extent that the most illustrative and exemplary data are selected, such research draws on the methodological strategies typical of generativism: to describe the grammar of possibility – what people can do with language, what the most complex and previously unsuspected ways of combining resources are that can be discovered or imagined, in order to determine the limits of possible combinations. Thus, Møller (2008: 218) asks “What if they [participants] instead orient towards a linguistic norm where all available resources *can be used* to reach the goals of the speaker?” (emphasis added), which is quoted approvingly by Otsuji and Pennycook (2010: 247) as part of their critique against the counting of “languages” in a repertoire as if they are discrete enumerable objects.

The qualitative approach adopted by scholars aligned with the traditions of the sociolinguistics of globalisation (including metrolingualism, polylingualism, translinguaging and the post-varieties approach) succeeds in expanding on the range of possibilities and illuminating some of the identity positions emerging from the choice of particular features. However, if we follow through on the emergentist position, then we must acknowledge that the most frequently encountered instances of language use form the basis for the inductive grammatical representations that speakers develop, which in turn sanction their use of language (Bybee 2006; Croft 2000; Langacker 1987: 99–146). Thus, in order to come to a complete understanding of the “linguaging practices” of people, it is necessary but not sufficient to only identify, by way of example, what the resources in the repertoire are, and what the range of possible choices from these resources are. For the development of a model that can account for the functioning of Englishes within multilingual communication settings, it is important to identify the distribution and interaction of

the resources quantitatively, to get a sense of how people *typically* leverage their resources in communicating in such complex new contexts, rather than exploring how people could *potentially* use these resources. Furthermore, adopting a quantitative approach allows one to approach data in a more probabilistic fashion, to distinguish between more and less frequent strategies, and allows for a better account of the variability in linguistic choices and social parameters. Writing from the perspective of corpus linguistics more generally, Leech (2015: 146–147, 159) notes that corpus linguistics, through its rigorous quantification of large samples of text, deals very explicitly with variation in language, but also uncovers the systematicity of variation in terms of both linguistic and extra-linguistic factors.

This raises questions, in the first instance, about data selection. Selecting data from the “extremes” of communicative hybridity without at the same time keeping an eye on data with less extensive hybridity runs the risk of overgeneralising the degree of instability of norms. This would amount to elevating an empirical difference along a continuum to a categorical or even a conceptual difference, as pointed out earlier in connection with concepts of “language”. By contrast, selecting English-only data, abstracted away from the authentic, multilingual communicative ecologies of its multilingual users, risks a caricature in the opposite direction. Such a strategy falls into the trap that a model of Englishes does not take sufficient account of the interaction between English and other languages at the level of actual language use, and continues to address this problem only at the broad functional level of choices between different languages in different domains.

### 3. Methodology

#### 3.1 Data

To meet the methodological challenges outlined earlier in this chapter, a new dataset had to be collected. The data had to meet a number of criteria: it had to draw on a range of Englishes, while at the same time offering room for the use of other languages/resources. Contexts characteristic of the recent wave of globalisation, and opportunities and incentives to develop local identities and norms, needed to form an integral part of the communicative setting for data collection. Interactive online data were required, rather than published print data, or static, possibly edited web content. In view of these considerations, we opted for the interactive user comments that accompany the “soapie teasers” (a daily summary of the content of the most popular South African television soap operas) on a popular website, TVSA (<https://www.tvsa.co.za/blogshome.aspx?tag=Soapie%20Teasers>). The popularity of the soapies is evident from the viewership numbers in Table 1. Only registered

users can contribute blogs and comments, although the content is fully accessible without any restrictions. At the time of writing (17 March 2017), almost 120,000 registered users participate on this website.

TVSA has an internal ranking system for blogs devoted to individual soapies and related blogs. The rankings are based on recent activity and not over the entire lifetime of the blog – some of which date back to 2006. Among the top ten blogs (out of a total of 4,211 blogs), eight are devoted to specific individual soapies, and their information (collected on 17 March 2017) is presented in Table 1. Viewership numbers are based on the audited viewership for television viewers older than 15 years from June 2016 (<http://www.brcsa.org.za/june-2016-top-20-tv-programs/>, date of access 16 March 2017), the last month of data collection.

**Table 1.** Popularity of and blog traffic for soapies among the top ten blogs on TVSA

Soapie	Peak viewership (in millions), in June 2016	Number of post views on blog (in millions) on 17 March 2017	Number of replies on blog on 17 March 2017
1. Muvhango	5.8	10.2	11,776
2. Generations	8.0	10.3	61,461
3. Skeem Saam	7.5	3.8	713
4. Uzalo	8.3	3.0	611
5. Sewende Laan	2.1	4.7	6,602
6. Scandal	4.9	4.1	4,710
7. Isibaya	1.0	3.8	9,354
8. Rhythm City	3.6	4.0	10,530

The soapies are all, to various degrees, multilingual. On the most un-English side of the spectrum are *Uzalo* and *Isibaya* in isiZulu, and *Sewende Laan* in Afrikaans, although they all include English dialogue, and often venture into other languages with smaller portions of the dialogue. *Muvhango* was in Tshivenda when it was first launched in 1997, but has increasingly migrated to a multilingual format. The other soapies are much more mixed, with larger parts English, but other South African languages contribute some of the dialogue in each one of them. All these soapies make use of subtitles into English, however, to provide wider access to content. The soapie teasers on the TVSA website are written in English, irrespective of the languages used in each one of them.

The data cover the period 2006, when the website was launched, to June 2016, to contain a snapshot of the “linguaging practices” of television watchers engaging in online conversation for exactly one decade. The total size of the corpus, in orthographic word tokens, excluding all metadata, is 5.33 million words. While demographic data about the participants are not directly accessible from the metadata,

an inspection of the actual textual content of the website and the profiles created by bloggers who contribute frequently show that the majority of participants who choose to share (demographic) information about themselves are black South Africans, and mostly female.

### 3.2 Analysis

To address the challenges we identified above, we adopted a quantitative corpus approach, but very specifically a corpus-driven or inductive approach. We do not make a priori assumptions about the resources in the corpus, but rather use statistical techniques that allow the identification of the most frequent orthographic tokens, which are then classified inductively into major resource types, and sub-classified in terms of classes that appear to cover the observations in a motivated fashion.

A wordlist for the entire corpus was compiled using WordSmith Tools version 7 (Scott 2016). Upon inspection of the wordlist, it became clear that the vast majority of the words were words in regular English orthography and thus it was established that resources conventionally associated with the named language “English” are extremely dominant in our data. However, non-standard English orthographic forms, particularly CMC-related forms, and orthographic tokens from other South African languages were also present. These are again labelled in terms of the conventional classes that emerge from the data. This practice is not alien to the participants in the TVSA discussion, as shown by Example (1).

- (1) Commenter 1: Hey Nonny lwill take this to your guestbook, geez l am scared of Gen. fans they will attack us Sizo le thola noma leka sifihla(my xhosa or zulu sucks)

Commenter 2: Sizo le thola noma leka sifihla(my xhosa or zulu sucks) LMAO at least wa-traya utswana lena ke-tryer sotho/tswana... panel beating ...haibo Shuga babe

Translation (using italics for translated text):

Commenter 1: Hey Nonny lwill take this to your guestbook, geez l am scared of Gen. fans they will attack us *We will find you even if you hide* (my xhosa or zulu sucks)

Commenter 2: *We will find you even if you hide* (my xhosa or zulu sucks) LMAO at least *you are trying you are just like me I am trying sotho/tswana...* panel beating ...*wow* Shuga (personal name) babe

In Example (1), the first participant, who is presumably from a Sotho background, tries to code-switch to an intermediate Nguni variety, but instead of *sizone*, writes *sizo le*, using the Sotho suffix for *will*; also *leka* is Sotho for *can*, the Nguni would

be *nika*. The second participant, presumably from a Nguni language background, repeats the non-standard Nguni expression, and then continues in a hybrid form with the English verb *try* embedded in Nguni and Sotho morphology, and various spelling errors from the perspective of the standard varieties. This extensive mixing is accompanied by a clear awareness of the named languages (or language families) from which the resources are drawn.

Two complementary analyses of the wordlist were undertaken. In the first instance, using the KeyWords function, a comparison was undertaken between the wordlist of the TVSA corpus and the British National Corpus (BNC, see Leech, Rayson & Wilson 2014). The BNC, as the modern variant of the historical input variety to South African English, represents a fairly conservative basis of comparison, which is useful for the purposes of this chapter, as it highlights differences between the resources typical in a representative corpus of standard English with the data in our corpus. The statistical technique itself is blind to the provenance of resources, it simply uses a calculation to find orthographic forms that are significantly more frequent in the TVSA corpus than in the control corpus. The 200 words that displayed the highest keyness, as measured by a log-likelihood score, were submitted to a detailed classification (where  $\lambda > 4529$ , and all orthographic types included in the list were attested at least 763 times in the TVSA corpus). We specifically utilised the KeyWords comparison to identify words from a global shared resource of Englishes that function in statistically exceptional ways, as a means of exploring the dynamic use of the resources from the standard language in the data, beyond simply noting the static presence of such resources. While many non-standard English and non-English orthographic tokens were also identified by the same statistical technique, we specifically focus on forms that are conventionalised forms in standard English in our exposition of this part of the analysis, while the non-English forms and non-standard English forms were both primarily identified by the second analysis.

The second analysis was a manual classification of the most frequent orthographic forms into their presumed linguistic origins: the conventional linguistic system with which they are typically associated. This does not imply that we assume speakers intended to use these named languages or deliberately mixed them. Instead, we use this method as a descriptive tool to organise the data, prior to interpreting their use in functional terms. In other words, our main interest is in the dynamic of use, but informed by a quantitative overall perspective, rather than proceeding from the basis of selected extracts in a qualitative analysis. We decided to include in the analysis all types with a normalised frequency of 100 or more instances per million words, which yielded a cut-off point of the first 1,041 items of the frequency-ordered wordlist. Of these, a total of 231 are not conventionally associated with standard English. A number of these were discarded because they

were personal names in the indigenous languages, typically names of characters in the television programmes. Further items discarded from the analysis include single letters with various uses; initials, abbreviations and emoticons; and single letters following apostrophes. Eventually, 152 forms were submitted to analysis, of which 99 fell in the broad category of non-standard English forms (including CMC), and 53 were from other South African languages.

Categories of analysis were developed inductively. Apart from drawing on our own knowledge of the various non-standard English and indigenous language forms, and dictionaries and grammars of the languages, we also worked with three language consultants, Lebohang Mathibela, Keabetswe Mothlodi and Johannes Mahlasela, whose contribution we acknowledge with gratitude. Once a provisional classification of subtypes in the data was made, we consulted previous research on code-switching, hybrid languages, CMC and multidimensional studies of register variation to refine the classification and embed it in relevant previous scholarship. A number of items were ambiguous, and in these cases, a brief inspection of the concordance lines was done to determine its majority function, and classification was then simplified to the most frequent function for every orthographic type. This sacrifices some precision, but should not change the overall quantitative or qualitative picture.

Where appropriate, we draw on the Global Web-based English corpus (GloWbe; Davies 2015), to determine if a particular noteworthy usage is specific to South Africa, or is also attested in other countries where English is widely used. GloWbe is a 1.9 billion word corpus of internet data from 2012–2013 that represents English-language websites from 20 countries, identified via web-domains. For any given search result, it can display results broken down per country, and indicate whether a form is relatively more frequent than the average in each country by means of colour-coding. Given the nature of the data we use, GloWbe is a very appropriate basis for comparison, since it contains data very similar to the main blog pages whose comments we analyse as corpus in this chapter (see Mair 2016 for a similar use of GloWbe). All frequencies except in Table 1 are normalised frequencies per million words (henceforth pmw).

In reporting examples from the corpus, forms under discussion are underlined. Material from languages other than English is printed in italics, followed immediately in square brackets by a translation and an indication of the source language. Often, with the Nguni and Sotho families, the same form can be used by multiple members of the family, in which case the individual language, e.g. isiZulu or Setswana, is not indicated, but just the family itself. Complete morphological analysis and glossing of examples are limited to those cases where the detail of the grammatical structure of the text is germane to the discussion.



## 4. Results

Three different sets of results are discussed in this section. We first analyse resources from standard English, based on the keywords analysis. Subsequently we focus on non-standard English forms and forms from indigenous languages, based on the classification of the 1,041 most frequent orthographic types from the wordlist. The total number of tokens represented by the selection, as well as their origins, is represented in Table 2, to give an estimation of the overall diversity of resources in the entire corpus. The most frequent forms together account for 4.18 million (78%) of all orthographic tokens in the corpus.

**Table 2.** Overall frequencies of resource types among the 1,041 most frequent tokens

Resource type	Tokens	Percentage
Standard English	3,727,378	89.1
Words containing numbers	59,082	1.4
Other non-standard English forms	228,603	5.5
Indigenous languages	76,781	1.8
Proper noun (indigenous content)	89,753	2.1
<b>Total</b>	<b>4,181,597</b>	

The estimate of the indigenous language resources might be too low, given that the agglutinative morphological structure of these languages mitigates against high type frequencies. Thus, another estimate was derived by comparing forms that index the first person singular subjects: the English pronoun *I* (28,285 per million words) to the corresponding forms in the major indigenous languages used in the corpus: the Afrikaans pronoun *ek* (42 per million words), and the subject concord forms *ke* (868 per million words) for the Sotho languages and *ngi-* (373 per million words) for the Nguni languages. This yields an estimate of 4% of indigenous language content. It is therefore reasonable to assume that the proportion of indigenous language use ranges between 2–4%. Example (1) illustrates an exchange with relatively high mixing, across a range of languages.

### 4.1 Keywords

The analysis of keywords yielded 54 forms that are conventionally associated with standard English in the 200 words with the highest keyness, alongside 59 non-standard English forms, 28 forms from indigenous South African languages, and 59 proper nouns that refer to characters, soapies or bloggers in the data. These 54 forms are sub-classified into five subsets, which give a sense of the functions



they perform in communicative interaction: personal interaction, mental verbs, evaluation, hedges and generalised language, and content words (the latter are not analysed in this chapter). Table 3 presents the classification of these forms.

**Table 3.** Standard English forms with significantly higher frequency in TVSA corpus than in BNC (in order of descending  $\lambda$ -value per subcategory)

Category	Orthographic tokens
Interaction	<b>First person:</b> <i>I, am, my, me</i> <b>Other:</b> <i>guys, please, morning, why, welcome, thanks</i>
Mental verbs	<i>love, think, like, hope, wonder, hate, guess, wish, missed, agree, tell, feel, miss, wants</i>
Evaluative lexis	<i>shame, boring, sorry, happy, sexy, poor, cute</i>
Hedges and generalised language	<b>Hedges:</b> <i>just, maybe, really</i> <b>Generalised nouns:</b> <i>guy, girl, thing, someone</i>
Content words	<b>Nouns:</b> <i>blog, bloggers, soapie, baby, episode, teasers, show, mom, TV, drama, storyline</i> <b>Verbs:</b> <i>watch, wait, going, watching, let</i>

The orthographic tokens that are identified as particularly frequent compared to the reference corpus are mainly features of interaction that are typically associated with oral communication, and correspond to feature classes on the positive side of the first three dimensions of Biber and Egbert's (2016) analysis of web registers. The features are specifically associated with a web register that they label interactive discussion. The first person subject pronoun *I* is the second most frequent word in the entire corpus after the definite article *the*, identical to the results for the wordlist of the spoken part of the BNC, but very different from the overall corpus results that are dominated by written language. This also clearly relates to the general purpose of the comments, to convey stance, hence the high frequency of first person pronouns and the copular verb *am*, alongside mental verbs,<sup>1</sup> as illustrated by Example (2):

- (2) Such an amazing show. It's a pity that it plays at the same time as an established show such as Scandal. I hope another season is in the pipeline.

Comments serve to evaluate various aspects of the soapies, either negatively with *boring*, *sorry*<sup>2</sup> and *poor*, as illustrated by Example (3), or positively with *happy*, *sexy* and *cute*.

1. The verb *tell* is a communication verb, rather than a mental verb, but ties in with the purpose of framing the expression of stance in the data, and is therefore grouped with the mental verbs.
2. *Sorry* is often used as discourse marker in South African English to signal general hearer misfortune, rather than an apology by the speaker that he/she is somehow responsible for the misfortune. This usage is also very widely attested in the data.

- (3) Dumi shame, i feel sorry for u brother, *phela* ['after all', isiZulu] Khethiwe wasn't begging in a sense, she was telling him that she will be staying with him.

The use of hedges like *just*, *maybe* and *really*, illustrated by (4), and very general nouns or shell nouns, like *guy*, *girl* and *thing*, illustrated by (5), is similar to interactive face-to-face speech (Biber 1988), but has not been identified as typical of web registers by Biber and Egbert (2016). The use of these resources betrays the planning pressure under which spoken language is typically performed, but also shows that users treat the digital medium with the same degree of immediacy as spoken language.

- (4) um glad um rytin' ur name down coz um not sure if maybe able to pronounce it properly
- (5) as 4 bernad it serves him right wateva bad thing that's happening to him... tshidi *yo* [exclamation, general South African] she wants to be a high school drop-out *nogal* ['surprisingly', Afrikaans]

Another feature of spoken interaction, which serves useful functions in the comments register where participants are engaged in dialogue, is the use of greetings (*morning*, *welcome*), terms of address (*guys*) and politeness markers (*thanks*, *please*). These elements are identified as extremely frequent by the keywords analysis, and are illustrated in Example (6).

- (6) Morning guys, i miss all the talk shows, please update me what is actually happening in generation, i do not know when was the last time ive watched the soapie.

The comment data rely to a very substantial degree on resources from the standard English repertoire (as shown in Table 2 – although admittedly based an approximation based on the 78% of the data covered by the most frequent orthographic tokens, a picture that is subject to slight revision if the entire corpus were to be classified). While there are clear correspondences to the representative, broad international sample of English web registers analysed by Biber and Egbert (2016), features of spoken interaction are even more prominent in the TVSA corpus than in their analysis. The data suggest that the resources of a shared English core form a baseline for the communication in the corpus, against which other resources from the repertoires of the speakers are added, consistent with the view advocated by Benor (2010). The participants do not consistently opt for the radical translanguaging that Makalela (2014) identifies in his South African conversational data, or the picture that emerges from the online data analysed by Seargeant and Tagg (2011), but remain closer to the kind of selections that Meierkord (2012) attributes to interactions across Englishes. This choice can be understood in similar ways to Meierkord's account: writers engage with many other writers, who are mostly not

known to them personally, and therefore they opt for resources that they anticipate will be widely understood by the conversation partners. At the same time, very few complete utterances rely only on the resources of standard English. The next two subsections of results present a closer look at the additional resources used. Their contribution to the coherent purpose of the conversational setting of the TVSA comments is shown, alongside additional layers of complexity where these usages reveal how multiple global and local linguistic resources interweave in the construction of complex, hybrid identities.

## 4.2 Non-standard English forms

The 99 non-standard English forms identified among the 1,041 most frequent orthographic tokens were classified inductively into two groups: forms linked to CMC specifically (drawing primarily on classifications from Deumert 2014), and other non-standard forms, not exclusively associated with CMC. The two categories with their subcategories are represented in Table 4.

**Table 4.** Classification of non-standard English forms

CMC-forms	Number of types
Phonetic spellings ( <i>gonna, hav, dat, shem</i> )	27
Consonant writing ( <i>thnk, bt, hw, cnt</i> )	18
Rebus writing ( <i>U, M, l, C, UR, b4</i> )	15
Apostrophe omission ( <i>Im, whats, don</i> )	8
Acronyms ( <i>LOL, LMAO, OMG</i> )	3
Contracted forms ( <i>abt, pple</i> )	2
General reduction ( <i>realy</i> )	1
Emphatic-aesthetic orthography ( <i>soo</i> )	1
Onomatopoeia ( <i>hahahaha</i> )	1
<b>Subtotal</b>	<b>76</b>
Other non-standard forms	
Abbreviated forms not exclusive to digital writing ( <i>peeps, ex, pic, congrats, etc, aka</i> )	10
Colloquial lexis ( <i>shame, wow, dude, chick</i> )	8
Lexis: forms of address ( <i>ms, ma, maan, bra</i> )	4
Spelling: one word ( <i>atleast</i> )	1
<b>Subtotal</b>	<b>23</b>
<b>Total</b>	<b>99</b>

As is evident from Table 4, CMC forms account for the majority of non-standard English types, while non-standard English spelling and lexis are relatively less important as a resource. The selection of non-standard English forms reflects the underlying dynamics of online communication as a hybrid register that straddles spoken and written registers: speakers draw on a fuller range of resources from their repertoires than just those ones associated with conventional monologic writing. There is an interplay between features imitating speech (as in phonetic spellings), and features optimising the economy offered by the written medium (as in consonant and rebus writing). This interplay is clearly evident in the extract in (7), where, apart from examples of rebus writing (*b4*, *u*), consonant writing (*dd*, *tnx*), and phonetic spellings (*wat*, *wen*), there is also onomatopoeia in the representation of laughter (*hahahahahahaha*) and exclamations (*uuuuuuu*).

- (7) Commenter 1: Hey Des tnx lala travel by train wat were u using b4? FanieN nothing interesting on Gen lately so ya no  
 Commenter 2: @ Sexy wat happened wu dd Ronnie tell? i was lolling wen Naomi was throwing herself herself @ Hermmie baby hahahahahahaha @ Inge Maldi's always HOT hey uuuuuuu

The selection of CMC features draws in diverse and often unpredictable ways on a global feature pool or repertoire for online communication. In doing so, users align themselves with the community of netizens, citizens of the net, but also incorporate local innovations reflecting complex, hybrid processes of identity formation. These CMC features combine in equally complex ways with English, Afrikaans and African-language utterances, drawing on resources that go beyond the confines of even typical English CMC, to convey local content in a glocal rhetoric, as in Example (8).

- (8) LMAO, I know Cngle, and mind u lomuntu ushaye [‘this person is displaying a’, Nguni] i-colgate [noun class prefix, Nguni] smile, knowing very well unez-ingovolo zika nogwaja [s/he has two big teeth of a rabbit]

Due to limitations of space, it is not possible to discuss all the features in Table 4 in detail. We therefore focus on two categories, phonetic spellings and lexis (including colloquial lexis and forms of address), which highlight the complex local and global embeddedness of the participants’ repertoires, as well as the strong interactive nature of the register, tying in with the pattern already established in the analysis of the use of standard English resources.

### *Phonetic spellings*

The most frequent phonetic spellings identified in the corpus, and the normalised token frequency of these spellings are shown in Table 5.

Table 5. Normalised frequency of phonetic spellings

Form	Standard equivalent	Token frequency (normalised pmw)
coz	<i>(be)cause</i>	879
gonna	<i>going to</i>	710
da	<i>the</i>	666
dat	<i>that</i>	650
wat	<i>what</i>	559
gal	<i>girl</i>	431
de	<i>the</i>	382
guyz	<i>guys</i>	359
cos	<i>cause</i>	356
wen	<i>when</i>	282
gud	<i>good</i>	282
wanna	<i>want to</i>	257
wil	<i>will</i>	256
lyk	<i>like</i>	245
shem	<i>shame</i>	235
dis	<i>this</i>	226
luv	<i>love</i>	207
tym	<i>time</i>	187
jus	<i>just</i>	163
wit	<i>with</i>	160
shud	<i>should</i>	141
neva	<i>never</i>	130
kinda	<i>kind of</i>	125
gona	<i>going to</i>	120
hav	<i>have</i>	120
wud	<i>would</i>	119
wats	<i>what's</i>	106

In analysing these phonetic spellings with a view to understanding their provenance, and how they might potentially reflect identities as constructed in the repertoires of these users, we carried out two analyses. We first used a phonetic analysis to determine whether these spellings may plausibly reflect the typical pronunciation features of Black South African English (BSAfE), as described by Van Rooy (2004). Subsequent to this, we turned to GloWbe to identify the countries where the forms are most widely used, thus attempting to trace the global paths of influence of these forms.

The phonetic analysis yielded only limited support for an interpretation that the phonetic spellings reflect the influence of BSAfE. For example, *dat* ('that') might reflect typical BSAfE features in the shift from /ð/ → [d], and /æ/ → [a],

although [ɛ] would be more frequent. Likewise, the substitution /ɛɪ/ → [ɛ] reflected in *shem* ('shame') is in line with BSafE pronunciation. However, in most other cases phonetic spellings are unlikely to reflect BSafE pronunciations. For example, *luv* ('love') does not reflect the likely pronunciation [laf], and *gal* ('girl') similarly does not reflect the typical pronunciation [gɛl]. The spelling *gel* for 'girl' is attested, but with a frequency of only 10 pmw, compared to the 431 pmw of *gal* and 698 pmw of *girl* itself. Thus, while imitations of the pronunciation of BSafE are present in the repertoire, speakers generally select these variants less often than they choose non-standard spellings that reflect other sources.

We next turned to GloWbe to trace the sources of the phonetic spellings in the repertoire of the TVSA-users. For the analysis, the frequency of each phonetic spelling across different countries was checked in GloWbe to determine where these forms are proportionally more frequent. Based on this, the phonetic spellings were classified into five main patterns of influence. The first category includes phonetic spellings that are **globally widespread**, so that the South African usage falls in line with international usage. These include *wanna*, *gonna*, *kinda* and *wit* – although it should be noted that the latter is noticeably more frequent in Nigeria. The second category is composed of phonetic spellings that are **more prominent in Outer Circle usage** (Africa and Asia) than in Inner Circle Usage. In this group, there are two sub-categories: usages that are widespread across the Outer Circle generally (*coz*, *wat*), and usages that are widespread across the Outer Circle but much more frequent in Nigeria. The latter category includes a substantial number of phonetic spellings: *cos*, *wen*, *gud*, *wil*, *shud*, *hav* and *wud*. The third category includes forms where **African usage is more prominent** than elsewhere. Again, there are two sub-categories: usages which are widespread across Africa (*lyk*, *tym*), and ones where Nigeria leads the generally higher frequency across Africa (*wats*, *guyz*). The fourth category is composed of usages that are distinctly more prominent in **Nigeria and Jamaica**. *dat* is notably more frequent in both varieties, whereas *dis*, *luv* and *neva* are most frequent in Nigeria, then in Jamaica, and then followed by countries elsewhere in Africa and Asia. *gal* and *jus* are most frequent in Jamaican usage, followed by African usage more generally. The fifth group of usages are those that appear to be **distinctly associated with South Africa**. *da* and *gona* are particularly prevalent in South Africa, whereas *shem* is most frequent in South Africa, but also occurs elsewhere in Africa.

The high frequencies of these forms in the TVSA corpus data suggest that these South African users draw in eclectic ways on resources available to web users across the world. The selection of phonetic spellings does not reflect, in any singular or consistent way, a straightforward anchoring of language to a consistent "local" identity. Rather, phonetic spellings in the TVSA corpus reflect a kind of bricolage of linguistic resources functioning in overt or covert ways as a hybrid constellation of identity markers. Orientations towards general globally shared usages are certainly

evident, but users draw particularly often on resources that seem to be very strongly represented in Outer Circle, and specifically African contexts, with a particular influence emanating from Nigeria, but also Jamaica. It should also be kept in mind that these phonetic spellings occur alongside other resources, including the use of African languages as in Example (9) with *coz*.

- (9) ok guys, im working as a receptionist @ a security company, so one didn't get paid so he called the officce & i answered the fone and he said all the rude things 2 me calling me a bitch and all, i lost my temper *ngamthuka nami* ['I also insulted him', Nguni], he then said *ngilale emsebenzini ngingaphumi* ['I should sleep at work and never get out', isiZulu] *coz* he will wait 4 me outside, now im afraid of going

### *Non-standard lexis: Colloquialisms and forms of address*

Non-standard English forms not exclusively associated with the digital context are a much less pervasive feature of the TVSA corpus than features associated specifically with CMC. In this section we discuss two groups of non-standard lexis not specific to CMC which occur frequently in the TVSA corpus: colloquialisms and forms of address. The token frequencies (pmw) of the items categorised in these groups are shown in Table 6.

**Table 6.** Normalised frequency of non-standard English lexis

Form	Token frequency (normalised pmw)
<i>Colloquialisms</i>	
hey	1,189
shame	881
oh	776
wow	529
yeah	207
okay	156
<i>Forms of address</i>	
ma	343
ms	303
maan	184
bra	146
dude	129
chick	106

These resources clearly function to manage interpersonal interaction, in similar ways as the standard English forms that were identified in the keyness analysis. As is the case for the phonetic renderings, the selection of these items indexes multiple

identity orientations. In the case of these lexical items, we find more typically local South African usages (*shame, ma, maan, bra*), alongside usages that are (based on an analysis of GloWbe) associated particularly with American English but disseminated more widely across varieties, often with particularly high frequencies in Inner Circle varieties (*dude, chick, yeah, okay, hey, oh, wow*). However, the local forms, particularly for the forms of address, are considerably more frequent.

We briefly consider in more detail two locally grounded forms: *shame* and *maan*. *shame* is a pragmatic marker with peculiar South African usage, and occurs in the standard spelling about four times more often than in the non-standard spelling *shem*. Its use cuts across South African varieties of English, and it is also often used in informal conversation and online communication by Afrikaans speakers. For example, it occurs more than 50 times pmw in Afrikaans online comments, represented in the Maroela comments corpus (Maroela Media 2015), where it is rendered as *shame* or in the Afrikaans phonetic spelling *sjym*. It expresses various sentiments, broadly related to sympathy, pity, or commiseration, as evident in Examples (10) and (11), where it is used clause-finally and clause-initially, respectively. Its high frequency in the corpus (second only to *hey* among the colloquial forms, at 881 pmw) reflects not only the importance of the interpersonal function in the data, but also suggests a particularly local pragmatic grounding.

- (10) Meant i did'nt *hayi* ['no!', Nguni] *bathong!*['Oh my word!', Sotho] he is tryin his best shame.
- (11) Shame man, how do u tell ur heart not to love someone? Eish *mina* ['T, Nguni] I wish all this was over now. Why do u think Pasons cheated?

*maan* is likely a phonetic spelling of *man* pronounced [ma:n], used as a vocative. Conversation often uses vocatives for managing interaction, and they also express an attitudinal function (Biber et al. 1999: 1108–1111). This function of *maan* is very clear in the TVSA corpus, as evident in Example (12).

- (12) i wudn't blame her its better than *ukujola no* ['dating with', Nguni] Khaphela and *amadevu akhe* ['his beard', Nguni] sies ['gross', Afrikaans] maan, cant the man shave already!!!

The collocations of *maan* are particularly revealing in this respect. Its most frequent L1 collocates (in raw frequency) are:

- related variants *hayi* (67), *hay* (27), and *haai* (24): a discourse marker in both Afrikaans (spelled *haai*) and the Bantu languages (where the original meaning is 'no'), roughly equivalent to English *hey*
- *soka* (39) and *suka* (13): from isiZulu, meaning 'go away/depart', often in the expression *suka wena* ('go away, you')



- *sies* (27), *sis* (15), variants of Afrikaans *sies*, meaning literally ‘gross’ or ‘disgusting’, but used as an interjection equivalent to *for shame* or *fie* in English (see Example 12)
- English *no* (17), together with Afrikaans counterpart *nee* (11)

The most frequent R1 collocates of *maan* are three pronouns (*I* (49), *she* (16) and *U* (16)), as well as the acronym *LOL* (11). These collocations clearly suggest the interpersonal function of *maan*, but also demonstrate how users draw on various resources within the broader local constellation of languages particularly for pragmatic purposes, even including languages (like Afrikaans) that would index a cultural identity that is unlikely to connect with the users’ cultural backgrounds. Users draw on the hybridity also within the local environment, so that the hybrid nature of the discourse is not just a simple opposition between global and local: the local is in itself differentiated, and users select resources in ways that are not self-evidently related to their cultural background.

The analysis of the non-standard English part of the repertoire of TVSA bloggers reveals two different sources: international resources, mainly CMC and non-standard usages associated with African, Outer Circle and to a lesser degree the African diaspora in the global North, and local innovations, including but not limited to phonetic spellings. Apart from signalling their identity as global netizens, some local English innovations, of the non-standard type that are unlikely to be candidates for conventionalisation within a local standard, are selected by bloggers to align themselves with other users. This latter function is also performed by forms from other languages, to which we turn next.

### 4.3 Forms from other languages

The resources from other languages can be classified linguistically on a continuum from those that are grammatically more integrated to those that are both morphologically independent and syntactically flexible. On the most grammatical side, affixes (mainly verbal prefixes in the Sotho languages) and demonstratives are closely linked to their syntactic class and usually combine with material from the Bantu language that the element is drawn from. They signal instances where participants choose to express themselves in relatively complete utterances in different Bantu languages. Conjunctions and complementisers are less integrated, as juncture phenomena, and allow for combination with English and indigenous language material. The same applies, somewhat unexpectedly, to emphatic pronouns, which also combine with English text. Terms of address and discourse markers are the most flexible categories, and can be used without any apparent grammatical constraint on the

combination possibilities, mainly to achieve interaction and alignment with other users. An overview of resources from non-English origins is presented in Table 7.

**Table 7.** Classification of material from languages other than English

Class	Examples	Type frequency	Token frequency (normalised pmw)
Affixes and demonstratives	<i>ke, re, wa, di, sa, ga, ba, mabloggers, ko</i>	13	4,047
Prepositions	<i>kwa, ka</i>	2	974
Conjunctions and complementisers	<i>en, le, gore, ukuthi, mara, kodwa</i>	6	1,668
Terms of address	<i>sana, mama</i>	2	449
Emphatic pronouns	<i>nna, mina, mna, yena, wena</i>	5	1,124
Discourse markers	<i>eish, ha, hayi, tjo, yo, ne, neh, nje, bathong, phela, kanti</i>	25	6,143
<b>TOTAL</b>		<b>53</b>	<b>14,405</b>

The most grammatical of Bantu language material in the classification are the **affixes** of the Sotho verbs, which are written as separate orthographical words, as shown by Example (13). There are corresponding affixes in the Nguni languages, where they are written as part of the same orthographic unit with the verb stem.

- (13) *ke a bona gore o* confused LaPan  
 SC-tense-see COMP SC-confused LaPan [gloss, Setswana]  
 I see that you are confused LaPan (LaPan = commenter's name)

Beyond verbal affixes, the hybrid noun *mabloggers*, which consists of the plural (class 2) noun prefix *ma-* and the English pluralised noun *bloggers* was the other form with a grammatical affix in the high-frequency list. This hybrid pluralisation strategy is attested in lower frequencies in the corpus, with this noun the only one making it to the top of the frequency list.

In general, the use of verb prefixes goes hand in hand with longer parts of text in the Sotho languages, as Example (13) illustrates, and there are also corresponding longer parts of text in the Nguni languages. Users draw on the resources of languages other than English to share their comments with other users, in line with the findings by Coetzee-Van Rooy (2014a, 2014b) and Makalela (2014) that English is not sufficient to fulfil functions of identity expression and alignment for users. The selection of these resources is also premised on the assumption that they are widely available to other users too and would not function in an exclusionary manner, which makes sense against the background of widespread multilingualism in South Africa.

A number of **emphatic pronouns** that are grammatically independent words are used in the data. These pronouns are not grammatically compulsory, since the subject concord indicates person and number distinctions. However, speakers have the choice to make the subject emphatically visible, with forms like *mma* or *mina/mina* for the first person in Sotho and Nguni respectively, or *wena* for the second person singular and *yena* for the third person singular in both families.

The emphatic use of these pronouns displays a varied range of combinations. They combine frequently with English text, as in Example (14), where *mma* simply functions as only pronoun, and Example (15), where it serves a genuine emphatic function.

- (14) hi guys am here tjoo talking of fasting *mma* ['T, Sotho] am giving it a rest lmao am joking i was in a camp from 2nd january
- (15) my boyfriend 's family invited me to join them for family picknik , but *mma* ['T, Sotho] i don't know *gore ke reke* ['whether I should buy', Sotho] food from the shops , or bake

In Example (15), the emphatic pronoun *mma* combines with the English pronoun *I*. This is not an isolated instance – of the 875 instances (raw frequency count) in the corpus, 38% have *I* immediately following *mma*, more frequent than the expected *mma + ke* in the Sotho languages, which is the second most frequent collocate, accounting for 12% of all cases. The option of using an emphatic pronoun, which is a grammatical construction in the Bantu languages, is transferred to the hybrid communicative context of the comments. As grammatical construction, it becomes schematic enough to allow semantically compatible material from any of the languages used in actual texts. The selection of a hybrid construction for referring to the self or the addressee in an indigenous language and English simultaneously can be interpreted as a mechanism to convey the hybrid sense of identity very explicitly.

**Conjunctions** (especially coordinating conjunctions) are combined freely with Bantu or English text, including *kodwa* (Nguni – ‘but’), *mara* (from Afrikaans, with vowel epenthesis, also ‘but’), as well as *le* (Sotho – ‘and’) and *en* (Afrikaans – ‘and’) (see Example (16) and (17)). It is interesting that Afrikaans conjunctions are more prominent than resources from Afrikaans generally, and these are often used to conjoin two different English conjuncts, at phrasal or clausal levels, as illustrated in (17).

- (16) they ended up in his flat *en* ['and', Afrikaans] you know biology
- (17) He would have told me *.Kodwa* ['but', Nguni] i will ask him *ntambamba* ['evening', Nguni] at home

The category of **discourse markers** is the area where the corpus makes the most frequent use of material from other South African languages. The Nguni form *yazi*,

equivalent to ‘you know’, functions as a comment clause very similar to English *I think* or *you know*, as in Example (18), although it obviously combines extensively with text in the Nguni languages too:

- (18) love is a b!tch *yazi* [‘you know’, Nguni]!! i wonder why we want it if it’s so cruel!!!!

What makes the choice of this discourse marker very interesting is that fact that the keyword analysis points to the extremely high frequency of a number of English mental verbs (*think, guess, hope, feel*) that are often used as comment clauses (see discussion above). Thus, the function of interpersonal alignment is performed by resources from both the English and indigenous parts of the repertoire.

Interjections are widely used in spoken discourse in the Bantu languages. These markers of orality are then transferred to the digital world of TVSA, and are supplemented with a number of expressions that have their origin in Afrikaans, but have become naturalised in BSAfE, as well as elements from Tsotsitaal, a hybrid subcultural urban language (Mesthrie & Hurst 2013). Similar to what has been shown for *maan* earlier, creativity, but also incipient constructionalisation, can be observed for some of these forms, for example in combinations with *hayi* and *bo*. Both are in origin from the Nguni languages, where *hayi* means ‘no’, but *bo* has no denotative meaning: it functions as supplement to exclamations and can be used on its own as exclamation. In the corpus, *hayi* has all kinds of spelling variants, including *hai* and *haai*. The latter resembles the Afrikaans discourse marker and greeting equivalent to ‘hi’, but in general the usage is in line with the isiXhosa origins (see also the discussion of these spelling variants in relation to their collocation with *maan*). Altogether, these variants make up 800 tokens per million words. An interesting combination attested a number of times in the corpus is the one in Example (19), where the Nguni *hayi* combines with a Sotho exclamation *bathong*, which originally means ‘people’ with a locative suffix – ‘to the people’, but as exclamation, it expresses shock or surprise, an emergent hybrid construction:

- (19) Generations!! Generations!! 1.Shazz and Sam u kids ,, just kissing each and evey episode, *hayi bathong* [‘no people’, Nguni and Sotho]!!!!

The single most frequent form among the interjections, however, is a relatively recent South Africa innovation, *ish*, which has diffused widely among different groups of speakers in South Africa. The origins of the interjection are unclear (though often attributed to Tsotsitaal). An example from the corpus is presented in Example (20), where the pronoun *I* follows the interjection, as it does in about a third of all examples in the corpus.

- (20) U must be right u know...*ish* i dnt know...

The keywords analysis points to some English interjections that are also significantly more frequent than in the BNC, especially *shame* and *please*, both discussed earlier. The general intent to manage the conversation at an interpersonal level, drawing on the rich resources of oral culture, cuts across all languages that participants have access to. However, forms like *hayi* and *eish* are actually more frequent than the most frequent English-language equivalents, which clearly shows that the users in some cases find resources from other languages more effective in fulfilling interpersonal functions.

Resources from the repertoire of Bantu languages are widely selected, while the examples from Afrikaans are generally less frequent. The vowel affixes are typically associated with longer utterances in the Bantu languages, but other elements are more independent and allow more flexible combinations of resources from different linguistic origins. Complex expressions of identity and interpersonal alignment are particularly prominent in the creative use of emphatic pronouns and discourse markers, where expressions of self and ways of managing interaction are conveyed by an interplay of resources from the repertoire. To the extent that the combination of English and Bantu language resources to fulfil these functions already constitute relatively high-frequency forms, some conventionalisation of these strategies in usage is evident. In respect of how these forms relate to notions of identity, it is crucial to note that these usages do not represent a counter-cultural statement against some dominant other, but alignment with other users drawing on the full range of available resources.

## 5. Conclusions: Summary and implications for modelling Englishes

The bloggers in the data choose resources from English most frequently. They do not exclusively choose resources from standard international or a standard local English, but rather a shared linguistic “core” that is presumed to be easily accessible to other users. The texts that result from their choices resemble interactive online communication (Biber & Egbert 2016), and even to a degree, face-to-face interaction (Biber 1988) in terms of their linguistic features. These selections also correspond to Meierkord’s (2012) finding that in interactions across Englishes, users select resources in terms of their assumptions about what their interlocutors will understand. The selection from such a general English core can be interpreted in a similar way as Benor’s (2010) view of the “standard English” that is assumed in the background of research that attempts to interpret the indexical use of resources from a repertoire to go beyond the identity associated with “standard English”. However, the strong register adjustments that our analysis uncovers indicate that this shared English core does not correspond so much to the elevated notion of

a “standard language”, and is rather to be conceived of as a selection of features appropriate to informal interaction.

In these users’ online communicative practices, they also draw on the resources offered by digital communication, and specifically the international English conventions of digital communication. Features of digital communication do not appear to significantly disperse to the Bantu languages, however, and remain strongly anchored in English, as also found by Deumert and Masinyana (2008). The way in which the users draw on the interface between the shared English core and the features offered by digital communication reflects a globalised, urbanised and technologised identity emergent in the interaction between local users.

Local and international non-standard English forms further contribute to the repertoire of resources. The selections in the data show evidence of alignment with a global African or African diaspora identity, where forms strongly present in Jamaican and Nigerian English are adopted by the TVSA commenters. The data do not allow for an analysis of how this dispersion may take place; however, while the possibility of face-to-face contact exists, the nature of these forms (mostly phonetic spellings) suggests a dispersion mechanism through written usage, possibly reflecting support for Mair’s (2013a) argument that online communication also acts as agent of diffusion of linguistic resources far beyond communities of origin. At the same time, decidedly local South African English innovations, not so much in the form of phonetic spellings or imitations of speech, but in resources of interpersonal communication, like forms of address and interjections, are frequently selected as well.

Local (English) innovations are complemented by the indigenous languages in the repertoire of resources, especially as far as terms of address, including emphatic pronouns, and discourse markers are concerned. They reflect interpersonal coordination, a kind of identity alignment *with* others, rather than in opposition to some mainstream identity, unlike the more typical finding from work on migrant communities (e.g. Blommaert 2010; Mair 2013a, 2013b). Intersubjective alignment therefore appears to be an important functional driver for frequency of use, contributing to the propagation of forms from the Bantu languages alongside forms from the English core, CMC, and non-standard English forms. It is important to note that, in some cases, there is evidence of forms from the Bantu languages interacting morphologically and syntactically in particular with forms from the English core, suggesting not so much mere co-existence of constructions from the different resources, but potential or incipient constructionalisation processes involving a type of linguistic gene flow that leads to hybrid micro-constructions.

In drawing these findings together with a view to their implications for a model of Englishes, a few key points emerge. The overarching model we propose has three aspects to it: a repertoire of **resources**, **selection** processes and the eventual diversity

of texts emerging from communicative interactions. The resources that the users have access to form a linguistic feature pool in the broad sense of evolutionary approaches to language change (Croft 2000; Mufwene 2008), but also consistent with the more dynamic concept of resources in the post-varieties approach advocated by Seargeant and Tagg (2011). This pool is not populated by objectified variety-objects as much as by users' knowledge of constructions, expressions and words from languages, in line with emergentist, constructional approaches to language (including more or less complete bits of language in the sense of Blommaert 2010). Users' knowledge of these resources includes knowledge of the sociolinguistic value of the elements and a sense of possible combinations with other elements. From here they select elements which they assume to be understood by the other users that they communicate with, and which fulfil a desired function within the communicative context of an online forum. The population of users is also open, equally in the sense that new participants may join, bringing new features into the pool, and in the sense that each TVSA user is embedded in other social networks, offering opportunities for the dissemination of features outside the forum by allowing for these features to enter an even larger feature pool.

The most extensive resource in the repertoire that TVSA users select from is the resources that form a subset of 'standard English' – that subset that users assume to be readily available, in line with Meierkord's (2012) findings. This subset is likely to be reasonably neutral in terms of identity construction, although it might project some of the positive attributes of English for black South Africans (Coetzee-Van Rooy 2014b). Combined with this, users draw on other resources from their repertoire, which index in complex, interwoven ways their embeddedness in an online culture, their expression of multiple subjectivities, and their alignment with fellow South African users. CMC language, a "bit" in the repertoire that only makes sense when it is integrated in a general English core, is used as a second important resource, exploiting the affordances of the medium while expressing a globalised and digitalised dimension of the users' identity. The grammatically more flexible resources from the indigenous languages, and those that are particularly directed at managing interaction, such as pronouns, terms of address and discourse markers, are among the most salient ways in which the participants' express their alignment with a local identity in online interaction. The selection of these resources emerges from the interplay between participants' exploitation of the conventions of the register of interactive online communication, and the ways in which they express complex local and global identity alignments. Local identity emerges from inter-subjective alignment processes with the immediate digital conversation partners, but it is at the same time interlaced with the global – but especially a very broad international African global.



When we consider the texts that emerge as products from the selections operating on the repertoire of resources, we find evidence of an English core, which users in their act of communicating with other users supplement with combinations of other elements, drawing on their own linguistic experience. They do so within the framework of intersubjective coordination within a digital community – where despite the absence of face-to-face contact, there is nevertheless both a degree of continuity, and social dynamics of intersubjective alignment.

The innovative patterns that arise in the pool may, for various reasons, conventionalise into new micro-constructions, as has been shown for *hayi* and *maan*. The pool is therefore not untouched and independent, and as combinations for multiple reasons yield certain “winning” patterns, they can become absorbed in one or more of the resources of the pool. At the same time, the caveat is that there is relative stability in this pool, with a common, if generous, core of English features that are useful for all varieties and contexts used, in terms of which marked choices become meaningful tools for identity expression and intersubjective alignment in the act of communication.

Most existing models, such as Kachru (1985) and Schneider (2007), focus on varieties as **objects**; even Mair (2013) operates with varieties, but affords more centrality to non-standard forms of English. In our model, fluid clusters of resources are embedded in a larger model where the resources form the first of three layers, alongside the next layers of a dynamic selection process operating on the resources, and textual products. We propose to focus on multilingual users and how their repertoires contribute to the development of feature pools and the potential conventionalisation of innovative features in dynamic **processes** of online interaction (or at least the records of these processes) – the feature pools are therefore dynamically influenced by usage. The kind of data investigated in this chapter, and the model proposed to account for it, make visible the interactive, multilingual communicative dynamics that may also be at the heart of the evolutionary processes of Englishes more generally. Drawing the view down to the usage level also highlights the instability, variability, and unpredictability of these processes. While these users are undoubtedly using English, that there is some evidence of hybrid practices that may be conventional(ising) in this English is indisputable.

At least two further sets of analyses are required in order to further reflect on the challenges to existing models of Englishes. To validate the approach outlined here, the introduction and diffusion of innovations over time in the corpus need to be modelled, taking into account the roles of particular agents (users). Secondly, to validate the more general proposal that the variety-based view of Englishes masks the widespread interaction between English and other languages, it is necessary to determine to what degree this hybridity extends outside the digital environment, and in particular is typical in authentic spoken interaction.



## Acknowledgements

We acknowledge the assistance of our consultants, Lebohang Mathibela, Keabetswe Mothlodi and Johannes Mahlasela with the interpretation of the data, and Martin Puttkammer for assistance with data collection. We also acknowledge the feedback from two anonymous reviewers, the editor and the audiences at the 2017 meeting of the Varieties of English in the Indo-Pacific – English in Contact (VEIP-EIC) research group in Sydney, Australia, and the 2017 UNESCO/CIPSH World Humanities Conference in Liège, Belgium. This work is based on research supported in part by the National Research Foundation of South Africa (Grant Number 109369). We are responsible for all remaining shortcomings, and take responsibility for all views expressed in this chapter.

## References

- Benor, S. B. 2010. Ethnolinguistic repertoire: Shifting the analytic focus in language and ethnicity. *Journal of Sociolinguistics* 14(2): 159–183. <https://doi.org/10.1111/j.1467-9841.2010.00440.x>
- Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: CUP.  
<https://doi.org/10.1017/CBO9780511621024>
- Biber, D. & Egbert, J. 2016. Register variation on the searchable Web: A multi-dimensional analysis. *Journal of English Linguistics* 44(2): 95–137. <https://doi.org/10.1177/0075424216628955>
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson.
- Blommaert, J. 2010. *The Sociolinguistics of Globalisation*. Cambridge: CUP.  
<https://doi.org/10.1017/CBO9780511845307>
- Bybee, J. 2006. From usage to grammar: The mind's response to repetition. *Language* 82: 711–733.  
<https://doi.org/10.1353/lan.2006.0186>
- Buschfeld, S. & Kautzsch, A. 2017. Towards an integrated approach to postcolonial and non-postcolonial Englishes. *World Englishes* 36(1): 104–126. <https://doi.org/10.1111/weng.12203>
- Coetzee-Van Rooy, S. 2014a. Explaining the ordinary magic of stable African multilingualism in the Vaal Triangle region in South Africa. *Journal of Multilingual and Multicultural Development* 35(2): 121–138. <https://doi.org/10.1080/01434632.2013.818678>
- Coetzee-Van Rooy, S. 2014b. The identity issue in bi- and multilingual repertoires in South Africa: Implications for Schneider's Dynamic Model. In *The Evolution of Englishes: The Dynamic Model and Beyond* [Varieties of English Around the World G49], S. Buschfeld, T. Hoffmann, M. Huber & A. Kautzsch (eds), 39–57. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/veaw.g49.03coe>
- Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach*. London: Longman.
- Crystal, D. 2011. *Internet Linguistics: A Student Guide*. London: Routledge.
- Davies, M. 2015. Introducing the 1.9 Billion Word Global Web-Based English Corpus (GloWbE). <<https://21centurytext.wordpress.com/introducing-the-1-9-billion-word-global-web-based-english-corpus-glowbe/>> (18 March 2017).
- Deumert, A. 2014. *Sociolinguistics and Mobile Communication*. Edinburgh: EUP.
- Deumert, A. & Masinyana, S. 2008. Mobile language choices – The use of English and isiXhosa in text messages (SMS): Evidence from a bilingual South African sample. *English World-Wide* 29(2): 117–147. <https://doi.org/10.1075/eww.29.2.02deu>

- Edwards, A. & Laporte, S. 2015. Outer and Expanding Circle Englishes: The competing roles of norm orientation and proficiency levels. *English World-Wide* 36(2): 135–169. <https://doi.org/10.1075/eww.36.2.01edw>
- Kachru, B. 1985. Standards, codification, and sociolinguistic realism: The English language in the Outer Circle. In *Collected Works of Braj B. Kachru*, Vol. 3, J. Webster, (ed.), 153–176. London: Bloomsbury.
- Kachru, B. 1986. *The Alchemy of English: The Spread, Functions, and Models of Non-native Englishes*. Urbana IL: University of Illinois Press.
- Langacker, R. 1987. *Foundations of Cognitive Grammar, Vol. I: Theoretical Prerequisites*. Stanford CA: Stanford University Press.
- Leech, G. 2015. Descriptive grammar. In *The Cambridge Handbook of English Corpus Linguistics*, D. Biber & R. Reppen (eds), 146–160. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139764377.009>
- Leech, G., Rayson, P. & Wilson, A. 2014. *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Routledge.
- Mair, C. 2013a. The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34(3): 253–278. <https://doi.org/10.1075/eww.34.3.01mai>
- Mair, C. 2013b. Corpus-approaches to the New English Web: Post-colonial diasporic forums in West Africa and the Caribbean. *Covenant Journal of Language Studies* 1(1): 17–31.
- Mair, C. 2016. Beyond and between the “Three Circles”: World Englishes research in the age of globalisation. In *World Englishes: New Theoretical and Methodological Considerations [Varieties of English Around the World G57]*, E. Seoane & C. Suárez-Gómez (eds), 17–36. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g57.02mai>
- Mair, C. Forthcoming. Crisis of the “Outer Circle”? – Globalisation, the weak nation state, and the need for new taxonomies in World Englishes research. In *Changing English: Global and Local Perspectives*, M. Filppula, A. Mauranen, J. Klemola, & S. Vetchinnikova (eds). Berlin: Mouton de Gruyter.
- Makalela, L. 2014. Fluid identity construction in language contact zones: Metacognitive reflections on *Kasi-taallanguage* practices. *International Journal of Bilingual Education and Bilingualism* 17(6): 668–682. <https://doi.org/10.1080/13670050.2014.953774>
- Maroela Media 2015. *Maroela Media-korpus 1.0*. Potchefstroom: CTextT, North-West University. <[viva-afrikaans.org](http://viva-afrikaans.org)>
- Meierkord, C. 2012. *Interactions across Englishes: Linguistic Choices in Local and International Contact Situations*. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139026703>
- Mesthrie, R. & Hurst, E. 2013. Slang registers, code-switching and restructured urban varieties in South Africa: An analytic overview of tsotsitaals with special reference to the Cape Town variety. *Journal of Pidgin and Creole Languages* 28(1): 103–130. <https://doi.org/10.1075/jpcl.28.1.04mes>
- Møller, J. S. 2008. Polylingual performance among Turkish-Danes in late-modern Copenhagen. *International Journal of Multilingualism* 5(3): 217–236. <https://doi.org/10.1080/14790710802390178>
- Mufwene, S. 2008. *Language Evolution: Contact, Competition and Change*. London: Continuum.
- Otsuji, E. & Pennycook, A. 2010. Metrolingualism: Fixity, fluidity and language in flux. *International Journal of Multilingualism* 7(3): 240–254. <https://doi.org/10.1080/14790710903414331>
- Scott, M. 2016. *WordSmith Tools*, Version 7. Liverpool: Lexical Analysis Software.
- Schneider, E. 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79: 233–281. <https://doi.org/10.1353/lan.2003.0136>

- Schneider, E. 2007. *Postcolonial English: Varieties around the World*. Cambridge: CUP.  
<https://doi.org/10.1017/CBO9780511618901>
- Schneider, E. 2014. New reflections on the evolutionary dynamics of world Englishes. *World Englishes* 33(1): 9–32. <https://doi.org/10.1111/weng.12069>
- Schneider, E. 2016. Hybrid Englishes: An exploratory survey. *World Englishes* 35(3): 339–354.  
<https://doi.org/10.1111.weng.12204/>
- Seargeant, P. 2010. The historical ontology of language. *Language Sciences* 32(1): 1–13.  
<https://doi.org/10.1016/j.langsci.2008.11.001>
- Seargeant, P. & Tagg, C. 2011. English on the internet and a ‘post-varieties’ approach to language. *World Englishes* 30(4): 496–514. <https://doi.org/10.1111/j.1467-971X.2011.01730.x>
- Van Rooy, B. 2004. Black South African English: Phonology. In *Handbook of Varieties of English, Vol 1: Phonology*, E. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (eds), 943–952. Berlin: Mouton.

# Placing ELF among the varieties of English

## Observations from typological profiling

Mikko Laitinen

University of Eastern Finland

This study investigates how (dis)similar ELF is structurally from the core native varieties of English, indigenized L2 varieties, and learner English. ELF is understood as second language use of English in settings where the interactants do not necessarily share a first language. The empirical part makes use of the method of typological profiling based on aggregate structural features. This method measures three indices (i.e. grammaticity, analyticity, and syntheticity), and it has been used previously to analyze a range of variety types but has not been applied to the assessment of ELF. The results provide quantitative evidence that places ELF on the map and shows that, on purely structural grounds, ELF is a distinct variety type among English varieties. Moreover, the observations show that ELF is structurally different from second language acquisition, and there is a quantitative basis for drawing a distinction between ELF and traditional learner data.

**Keywords:** English as a lingua franca, second language use, typological profiling, genres in ELF

### 1. Introduction

This article tackles the question of how different English as a lingua franca (ELF) is structurally from other varieties of English. It makes use of the method of typological profiling based on aggregate structural features (Szmrecsanyi 2009). This method measures three indices and has previously been used to analyze various native Englishes, indigenized L2 varieties and learner English. In a seminal study, Szmrecsanyi and Kortmann (2011: 182) use the method to draw a distinction between English as a foreign language (EFL) and English as a second language (ESL) on structural grounds. Up to today, the method has not been used to assess ELF, and my results provide quantitative evidence that places ELF on the map. ELF is understood here as second language use of English in settings where the interactants do not share a first language. As in Mauranen et al. (2015), my definition allows native

speakers to be involved. Throughout the article, I will use the term ‘ELF speaker’ in a generic sense to include both speakers and writers.

ELF is a comparatively new object of study, and it is not included in the current models in the study of World English, which are based on history and geo-political background (cf. Kachru 1985; Schneider 2007). Nevertheless, the sociocultural and sociolinguistic importance of ELF among the present-day English varieties is substantial. In his article on the World System of Englishes, Mair (2013) classifies ELF as a super-central variety, i.e. a variety that is transnationally relevant, carrying demographic weight and sociocultural importance. Mair (2013) argues that, unlike some other super-central varieties, ELF is restricted to domain-specific uses such as academia and international business/law, but as will be shown in the material used here, this is a far too restricted view (cf. also Pietikäinen (2017) on the uses of ELF in the family setting). Sociolinguistically, ELF is not a focused variety, in the sense of Milroy (1987: 182–183). It has no native speakers nor do speakers share a widely-held and recognizable set of norms at all levels of the language. Yet recent empirical evidence suggests that spoken ELF is gradually emerging as norm-developing, rather than being simply norm-receiving (Low 2016).

This line of quantitative typological research is novel in ELF, where the research has predominantly been qualitative. Mauranen et al. (2015) point out that grammatical variability has been the least researched area, and ELF scholars have primarily only investigated individual grammatical features without relating them to in-depth quantitative information and systematic comparisons with other varieties of English. These include studies of the reorganization in the relative pronouns *who* and *which* (Cogo & Dewey 2012), and the regularization of the third person *-s* (Breiteneder 2009). Large-scale corpus studies are fewer. Ranta (2013) has focused on shared non-standard features in grammar, *viz.* vernacular universals, such as the inverted word order in indirect questions, the extended use of the progressive, the use of *would* in hypothetical *if*-clauses, and the use of singular agreement in existential *there* structures (Ranta 2013). In addition, Laitinen and Levin (2016) and Laitinen (2016, 2017) have looked into how ongoing grammatical change is adopted in ELF. These studies consist of investigations of the changes in non-aspectual uses of the progressive and investigations of a broader set of features, such as core and emergent modal auxiliaries (*can*, *should*, *have*, *need to*, and *be going to*, etc.) in ELF. Nevertheless, it remains fair to say that little is known about the typological status of ELF.

Drawing her evidence from the *English as a Lingua Franca in Academic Settings* corpus (ELFA), Mauranen (2012: 247) points out that spoken academic ELF is in many ways similar to native speech and “the overwhelming majority of lexis, phraseology, and structures are indistinguishable from those found in a comparable corpus of educated ENL [English as a Native Language], including their frequency

distributions”. Her observations are based on n-grams and word lists. With regard to other non-native varieties, she argues that ELF is essentially dissimilar from learner English.

The research question is concerned with how different ELF is structurally from the core native varieties, indigenized L2 varieties and learner English. Since all ELF use has involved language learning at one point, my null hypothesis is that ELF is similar to learner data. The results respond to two questions: First, where on the unidimensional grammaticity index does *lingua franca* evidence fall; and second, where on the two-dimensional analyticity–syntheticity plane can I place ELF? Both of these questions will be explained below.

Section 2 details both the theoretical and the methodological basis of the profiling method adopted here. Section 3 discusses the ELF material analyzed, and the results are presented in Section 4. Lastly, Section 5 discusses the implication of the results for model building in World Englishes.

## 2. The method of typological profiling and ELF

The study employs the method presented in Szmrecsanyi (2009), who proposes that the typological notions of analyticity and syntheticity by Greenberg (1960) could be amended by large-scale quantitative corpus data. This method makes use of three indices. Firstly, the synthetic index is based on the frequency of select bound grammatical markers, and the numeric value in this index is the number of lexical items that carry at least one bound marker. A prototypical case would be verbal third person *-s*, which marks two meanings, *viz.* nonpast and third-person singular. Secondly, the analytic index is calculated on the basis of a range of function words, which are “defined as being members of closed word classes” (2009: 320). Thirdly, the grammaticity index is the sum of synthetic and analytic markers per sample.

The indices are presented in detail in Szmrecsanyi (2009: 326–327). The analytic markers consist of the following:

1. conjunctions, subjunctions, and prepositions
2. determiners, articles and *wh*-words
3. existential *there*
4. pronouns
5. analytic comparative and superlative markers
6. *to*-infinitive marker
7. modal auxiliaries
8. negator *not*, or *n't*
9. auxiliary *be*

10. auxiliary *do*
11. auxiliary *have*

The synthetic markers are:

12. *s*-genitive
13. synthetic comparative and superlative adjectives
14. plural nouns
15. plural reflexive pronouns
16. inflected verbs

Some of the categories, *viz.* the analytic categories 9–11 also load the synthetic side. Similarly, the synthetic items 15–16 load the analytic indices. This method results in frequencies that measure variability, but it is not variationist as such, in which case all of the markers would be used to express one meaning with two forms. Szmrecsanyi (2009) points out that for some analytic markers there is a clear synthetic alternative (i.e. the forms of adjectival comparison, or the analytic and synthetic genitives). For some, such as the negator *not*, or the plural noun marking, this is not the case.

It should be noted that the method should also be viewed critically, especially since the grammatical components are not weighed relative to their basic frequencies. The component categories vary in token frequency, and the relative weight of the high-frequency elements is substantial. Therefore, tiny alternations in prepositions, determiners and pronouns (all on the analytic side) or plural nouns (synthetic) will lead to considerable alternations in the normalized frequencies. One way to improve the precision of the method could be to assign a relative weight to all the components, but since the present aim is to compare ELF with previous results, this is left for future studies.

Since the method integrates various structural features, it offers a useful way of quantifying a variety that does not prioritize a single grammatical structure, thus limiting the bias inherent in single-feature studies. For instance, it has offered a more fine-grained picture of the interplay between synthetic and analytic tendencies in the history of English. As Szmrecsanyi (2009) points out, Standard English is often seen to be an analytic language *par excellence*, but the quantitative results obtained through the method have contested this monolithic myth, and his results show that both major varieties, *viz.* American English (AmE) and British English (BrE), became more synthetic and less analytic during the second half of the 20th century. The same holds for earlier periods; Szmrecsanyi (2012) applies the frequency-based indices for post-Old English corpus-data. His results show that analyticity has been on the decline since the Early Modern English period, and syntheticity has increased.

The method has also offered a way of quantifying structural differences between varieties. Firstly, when it comes to grammaticity, varieties differ substantially with regard to how overtly redundant they are. Traditional L1 vernaculars (low-contact varieties) exhibit more grammatical marking than high-contact L1 vernaculars (e.g., AmE, New Zealand English, etc.), which in turn exhibit more grammaticity than indigenized L2 varieties (Singapore English, etc.). Secondly, in terms of syntheticity, low-contact varieties show higher frequencies of syntheticity than high-contact varieties, and L1 varieties in general display more syntheticity than L2 varieties. Thirdly, among the L2 varieties, Southeast Asian Englishes (e.g. Singapore, the Philippines, and Hong Kong) are less analytic and less synthetic than those outside (e.g. IndianE, JamaicanE, East AfricanE). Most importantly for my purposes, learner English data exhibit “less syntheticity and more analyticity than Standard British English” (Szmrecsanyi & Kortmann 2011: 182).

Many of the features covered in the indices are directly relevant to aspects of ELF and could therefore offer empirical insights into how similar or different ELF is when compared to the other English varieties. A case in point is one of the hallmark characteristics of ELF, i.e. negotiating meanings through online processing. According to Mauranen (2012: 244), one characteristic of spoken ELF is enhanced transparency through structural simplification. Since the typological indices are closely connected with language complexity, the method offers a way of quantifying such transparency in ELF. On the one hand, the more analytic a language is, the more it tends to contribute to transparency and explicit nature of communication. On the other hand, increasing syntheticity tends to create a more economical output, while grammaticity contributes to explicit redundancy, meaning that the more grammatical markers there are, the less needs to be inferred from the contextual cues (Szmrecsanyi 2009).

While the present article applies the method that is readily available, it still involves a considerable theoretical component. It deals with ensuring empirical validity and enlarging the scope of ELF corpora, as the existing datasets cover only a small set of genres. We need new corpora that offer a multi-genre view to ELF. These new corpora should ideally be such that they enable comparisons with other (native and non-native) corpora.

### 3. First- and second generation ELF corpora as material

My material comprises two sets of corpora, which for the first time make it possible to access a broad range of ELF genres. The term genre is understood as a concept that points to the functions of communication, i.e. situation, audience, and the purpose (see Biber & Conrad 2009). The first set of data consists of the well-known



first-generation ELF corpora, *viz.* the spoken *Vienna-Oxford International Corpus of English*, VOICE (Seildhofer 2011), and the newly-released *Written English as a Lingua Franca*, WrELFA, corpus (Mauranen et al. 2015).

VOICE is a one-million-word corpus consisting of unscripted, face-to-face spoken interactions from organizational settings. The informants come from a mixture of L1 backgrounds, and since the individual L1 collections result in small samples, the corpus is used in its entirety. VOICE represents spoken communication in which the informants' objective is to inform and to maintain interpersonal relations.

WrELFA is an approximately 1.5-million-word corpus of academic writing divided into three text types in the academic genre. Unedited research papers provide half of the material, the so-called SciELF corpus. The PhD examiner's report genre contains some 400,000 words, and the research blog component some 372,000 words. The collection process targeted the academic user of ELF, and, according to the compilers, the texts have not undergone professional proofreading or checking by an English native speaker (see <http://www.helsinki.fi/englanti/elfa/wrelfa.html>). It represents second-language use in written scientific communication, and 35 L1s are represented in it. Additionally, an undetermined number of blog commenters are included in the blog component, and according to the corpus compilers, their identities cannot be verified. Similarly to the other ELF corpora, native speakers of English are occasionally included in the blog and in the PhD examiners' subcorpora. Since the results in the following section are the first ELF results obtained using this method, I will only use the PhD examiner's statements contained in WrELFA.

To complement these first-generation corpora and to compensate for the fact that "genuine ELF written text databases are still missing" (Mauranen et al. 2015: 402), the author and his associates are currently compiling second-generation written ELF corpora. They offer a larger sample for a smaller set of L1 backgrounds than the first-generation corpora and broaden the stock of ELF genres available. They concentrate on second language use of English in specific geographical settings.

Our pilot work focuses on two Nordic countries, Sweden and Finland, where the role of English has undergone considerable changes in recent decades. The two countries are not undergoing a language shift, but the sociolinguistic situation is that of urban multilingualism in which English is used as an additional resource alongside the main languages, primarily, but not exclusively, by younger generations who live in urban areas and work in white- and pink-collar professions (see Laitinen 2016).

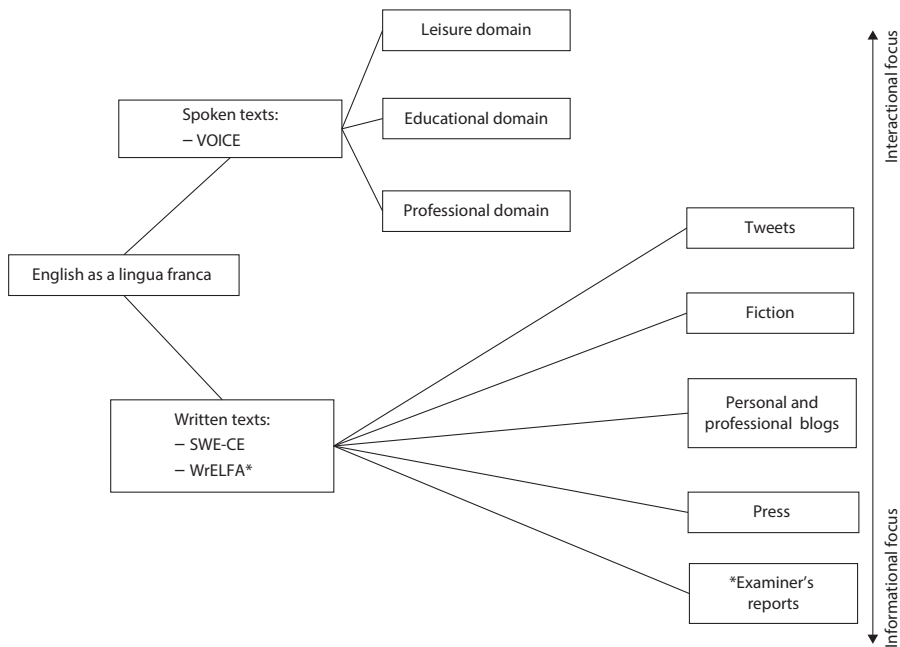
The working titles for the corpora are SWE-CE, the *Corpus of English texts in Sweden*, and FIN-CE, the *Corpus of English texts in Finland*. They are systematically-collected and sufficiently large sources of baseline data that fulfil the requirement for empirical validity. They contain texts from the written mode of communication,

and together with the already-existing spoken ELF corpora they make it possible to investigate a range of genres. The texts have been produced by non-native writers who use English as a second language resource. The majority are taken from non-learner settings. The only exception is fiction, which we are collecting in collaboration with teachers organizing creative writing courses. The rationale is that (fan) fiction is an important arena of ELF writing (Leppänen 2012), but unfortunately such texts do not fulfil our need to identify the authors, and we have to collect material from educational settings.

We know to what extent the materials have been subjected to normative language checking by professional editors, translators and native speakers. Preference is given to texts that are not edited, but it is assumed that the more informationally oriented a text is, the more likely it is to have undergone some degree of language checking and collaborative effort. Furthermore, the informants' use of spell-checkers and other tools which nowadays are available in most web-browsers and mobile devices cannot be ruled out. To what extent such tools have an influence on our data is beyond our control, but it is clear that such tools are part of contemporary writing practice and are equally used by native writers. Published materials edited by native speakers are excluded.

These second-generation ELF corpora cover a range of genres. We draw from Biber's (1988: 104–108) multidimensional analysis of textual variation, and more specifically from dimension 1, i.e. information density and exact content vs. interactional and generalized content, to place texts within the genre matrix. This dimension is used as a heuristic tool and has not yet been validated empirically. Figure 1 visualizes the textual division covered in the study.

When the written and spoken ELF materials are combined, the resulting corpora include some 2.3 million words. The results are based on 1,023,082 words of spoken VOICE. On the written side, I use the Swedish component as the material. The written corpus consists of 332,290 words of tweets (short micro-blog messages). This material has been collected from 50 randomly-selected individuals at [www.curatorsofsweden.com](http://www.curatorsofsweden.com). The site collects tweets sent by people who are citizens of Sweden and who manage the 'official' Twitter account Sweden for one week at a time (for more on our Twitter data collection see Laitinen et al. 2017). The ELF fiction subcorpus is 193,755 words. The professional and personal blogs consist of 263,486 words, and they are considered as a single component (note that we will divide them into professional and personal blogs; see Grieve et al. (2010) on the classification of blogs using linguistic criteria). The news subcorpus consists of 196,232 words, while the examiner's statement component from which the native English writers have been excluded comprises 276,712 words in 236 statements.



**Figure 1.** The genre distribution of the ELF corpora used in this study

To ensure comparability with the previous observations, my study applies the method presented in Section 2 with no major modifications. One minor modification, however, is that the results for the VM category, (7) in the list above, also include the contracted forms *gonna*, *gotta*, *hafta*, and *wanna*, but their frequencies are low. The material that was untagged (WrELFA and SWE-CE) was parts-of-speech tagged using the CLAWS7 tagset. Tagging was also tested for blogs and tweets, but the error rate turned out to be high for these genres and hence it was determined that it would be best to run the material through a PERL script that attached 1,000 randomly selected items in the subcorpora with a tag. The script was kindly provided by Benedikt Szmrecsanyi. These items were manually analyzed for their POS. The results for the fiction, blogs, and press subcorpora are the frequencies generated through automatic POS-tagging. The VOICE results were obtained using the POS-tagged version (<http://www.voice.univie.ac.at>). Some of the results that required broader contextualization were checked using the XML-version of the corpus (<https://www.univie.ac.at/voice/help>). The tagging used in the VOICE corpus is based on a modified set of Hepple tags (Seidlhofer et al. 2014), and a scheme was created to convert the results comparable to those obtained using CLAWS7 (as illustrated in Table 1 below).

**Table 1.** The analytic (A) and synthetic (S) component categories as defined through the POS tags

Feature	CLAWS7	VOICE tagset or search function
A1: Conjunctions, subjunctions and prepositions	CC*, CS*, I*	CC, IN
A2: Articles, determiners and WH-words	APPGE, AT*, D*, RGQ*, RRQ*	DT, PDT, PRE, WRB, WDT, WP
A3: Existential THERE	EX	EX
A4: Pronouns	P*	PP*, indefinite, reflexive and reciprocal pronouns
A5: MORE/MOST	RGR, RGT	RBR, RBS
A6: Infinitive marker TO	TO	TO
A7: modals	VM*, <i>gonna, gotta, hafta, wanna</i>	MD, <i>gonna, gotta, hafta, wanna</i>
A8: negator NOT/N'T	XX	NOT, N'T
A9: auxiliary BE	VBD*[VBG VBM VBN VBR VBZ* + (*)? + V*	lemma: BE + (XX0)?  (*)? + (*)? + V*
A10: auxiliary DO	VD* + (*)? + V*, VD* + XX	lemma: DO + (XX0)?  (*)? + (*)? + V*
A11: auxiliary HAVE	VH* + (*)? + V*, VD* + XX	lemma: HAVE + (XX0)?  (*)? + (*)? + V*
S12: Germanic genitive marker ('s)	GE, MCGE	POS
S13: Comparative and superlative adjectives	JJR, JJT	JJR, JJS
S14: plural nouns	NN2, NNL2, NNO2, NNT2, NNU2, NP2, NPD2, NPM2	NNS
S15: plural reflexive pronouns	PPX2	*SELVES
S16: Inflected verbs	VBDR, VBDZ, VBG, VBM, VBN, VBR, VBZ, VDD, VDG, VDN, VDZ, VHD, VHG, VHN, VHZ, VVD, VVG, VVGK, VVN, VVNK, VVZ	VVD, VBD, VHD, VVG, VBG, VHG, VVN, VBN, VHN, VVZ, VBZ, VHZ, VHS, DOS, VBS, VBP

The numeric results for the indices have been provided by two trained research assistants, and their initial searches have been checked once. As in Szmrecsanyi (2009), the indices are ratios of the number of markers normalized per 1,000 words.

## 4. Results

### 4.1 Grammaticity

Table 2 illustrates the total number of both grammatical markers, i.e. grammaticity. It is the most robust category, showing how transparent a variety is. According to Szmrecsanyi (2009), the higher the score, the more efficient the output is in terms of pragmatic functions, since the relationship between overt marking and negotiating meanings is indirect. Similarly, the lower the index, the more needs to be negotiated using pragmatic means. To acquire understanding of where ELF is situated among the varieties of English, the ELF results (in bold) are compared with the figures drawn from Szmrecsanyi (2009: 329). To make the samples more comparable, the ELF observations at this stage exclude tweets. They will be included in the subsequent tables and figures, but as they constitute a highly distinct genre, they are presented separately (for more on the characteristics of tweets and other e-genres, see Knight, Adolphs & Carter 2014).

**Table 2.** Grammaticity index (GI) of ELF compared with the other varieties (from Szmrecsanyi 2009)

Language variety/form	GI	z score
Hong Kong E	539	-1.93
Singapore E	549	-1.70
<b>ELF</b>	<b>574</b>	<b>-1.13</b>
Philippine E	592	-0.72
Irish E	598	-0.58
New Zealand E	607	-0.38
Standard AmE	607	-0.38
Somerset (southwest)	626	0.05
Jamaican E	627	0.07
Indian E	632	0.19
Standard BrE	643	0.44
East African E	647	0.53
Kent (Southeast)	657	0.76
Lancashire (North)	667	0.98
Glamorgan (Wales)	669	1.03
Shropshire (Midlands)	680	1.28
Sutherland (Highlands)	689	1.49

The GI scores are the arithmetic means and indicate that ELF is roughly one standard deviation below the mean value of this index. ELF falls between two outer circle varieties, a little lower than Philippine English and a little higher than the Southeast

Asian L2 English varieties, Hong Kong and Singapore English. These last two in particular are highlighted by Szmrecsanyi (2009) as contact-induced varieties in which adult language learning plays a significant role.

The results in Table 2 are important on at least two levels. For the first time, we are not confined to a limited set of genres in the ELF setting but can rely on evidence from various discourse situations on the spoken–written continuum. In addition, as opposed to much of the previous ELF evidence, Table 2 makes use of evidence based on aggregated linguistic structures rather than single grammatical, lexical, or phraseological features.

The results in Table 2 are quantitative evidence of structural simplification observed in previous ELF studies. They support some of the previous findings in the ELF literature based on spoken data (see Mauranen 2012: 244), namely that one characteristic of spoken ELF consists of the negotiation of meanings in interaction, which leads to enhanced transparency and structural simplification. Table 2 offers a quantitative view of what this transparency means in corpus data. The results also add evidence that ELF speakers often avoid overt grammatical marking. According to previous studies by Breiteneder (2009) and Jenkins, Cogo & Dewey (2011: 289–290), one characteristic of ELF interaction is the omission of grammatical markers, such as third person *-s* or articles, both of which are included in the indices in this study.

While Section 4.3 focuses on genre differences in the ELF corpora in more detail, I will next focus on how the spoken and written ELF modes differ from each other in terms of grammaticity. Specific attention is also paid to how one genre behaves relative to the spoken–written continuum. This genre, tweets, is written in its form, but it tends to exhibit spoken characteristics. Since the standard corpora used in Szmrecsanyi (2009) do not contain material from this genre, it is kept separate. The results also include the arithmetic mean values of four main variety types of English, drawn from Szmrecsanyi (2009: 329–330). They serve for reference purposes to show how substantial the differences between the spoken and written ELF subcorpora are.

Table 3 illustrates that the tweet subcorpus has the lowest grammaticity index (GI: 536), and it is clearly a specific written genre in which more emphasis needs to be placed on contextual cues and pragmatic inference than in spoken communication. VOICE corpus has a GI of 553. This result illustrates the emergent nature of spoken ELF, in which meanings are negotiated through enhanced explicitness (Mauranen 2012: 245). On the written side, however, the result is markedly different, and the grammaticity score is substantially higher (the mean is 597), giving it greater similarity to the outer circle L2 varieties than spoken ELF.

Table 3. Grammaticity indices of ELF compared with the data from Szmrecsanyi (2009)

Language variety/form	GI	z score
ELF tweets	536	-1.26113
Spoken ELF (VOICE corpus)	553	-0.83599
Southeast Asian Englishes (Singapore, Philippines, Hong Kong)	560	-0.66093
Written ELF (WrELFA and SWE-CE)	597	0.264372
Other L2 (outer circle) varieties	598	0.28938
Transplanted L1	607	0.514454
Low-contact L1 dialects	654	1.689839

The quantitative patterns observed are clear. With regard to the unidimensional grammaticity index, ELF falls between the two L2 variety types of Southeast Asian Englishes and other outer circle varieties. It is clearly not on a par with the transplanted L1 varieties and is well below the average of the traditional low-contact L1 dialects.

The illustrations below show what these quantitative differences mean in actual texts. Note that, for visualization purposes, only two of the analytic markers (determiners and modals) and synthetic ones (plural nouns and inflected verbs) have been included in the illustrations provided here. It goes without saying that any automatically-generated contextual information in tweets (i.e. the time of sending a tweet) and the material not keyed in by an individual author (URL-links, re-tweet mark-up, etc.) are separated by our text-level coding scheme used in the this sub-corpus. They are not included in the results.

- (1) <TIME>May 31, 2015, 1:18 p.m.</TIME> <AT>@47thANNA</AT> Haha.  
 <TIME>May 31, 2015, 1:18 p.m.</TIME> <AT>@MarissaTree</AT> <AT>@niannelynn</AT> When **will** the wedding be?  
 <TIME>May 31, 2015, 1:17 p.m.</TIME> <AT>@HarietaNoPotter</AT> Haha, sorry! True detective!  
 <TIME>May 31, 2015, 1:17 p.m.</TIME> <AT>@va\_ellen</AT> Strangely enough I **haven't** been there.  
 <TIME>May 31, 2015, 1:15 p.m.</TIME> <AT>@Kyroenna</AT> **That's** my guess also ...  
 <TIME>May 31, 2015, 1:13 p.m.</TIME> **The** Vegetable Man **goes** to the beach wearing a zukini.  
 <TIME>May 31, 2015, 1:09 p.m.</TIME> What **could** be the favorite food and drink of <Q>True blood</Q> writer Nic Pizzolatto, I wonder?  
 <TIME>May 31, 2015, 1:05 p.m.</TIME> <AT>@niannelynn</AT> You **are** absolutely right.  
 <TIME>May 31, 2015, 12:59 p.m.</TIME> <AT>@niannelynn</AT> But we hardly know each other?!

<TIME>May 31, 2015, 12:56 p.m.</TIME> <AT>@dmacuk</AT> Well **said**.  
 <TIME>May 31, 2015, 12:56 p.m.</TIME> <AT>@niannelynn</AT> Do you  
 propose?

<TIME>May 31, 2015, 12:50 p.m.</TIME> <AT>@dmacuk</AT> But **the** vote  
**said** no? (SWE-CE, tweets, May 2015) (15 markers = 23%)

- (2) yes to force to force to integr- to force **the** integration and that's **the** that's **the**  
 the main point of difference because i **went** abroad i **got** my education and i **used**  
 it at home for **my** duke for **my** bishop fo- for **my** ho- hometown and today i  
 think we get educa- er we get education and we **don't** know where we **gonna**  
 use it and that's **the** that's **the** big difference

(VOICE, EDsed251) (18 markers = c. 22%)

- (3) As **mentioned** earlier, when **an** employee compiles a quote in **the** office, **the**  
 customer **might** feel that he is not part of **the** process and worry that **the** price is  
 manipulated by **the** company. Examples of this **can** be **found** in forum threads,  
 such as at byggahus.se [3], **discussing** the subject. Delayed price quotes and  
 customers **feeling** cheated is a problem for **the** company. This thesis **will** look  
 at how **the** manual process **can** be **sped** up and **made** more transparent

(SWE-CE, theses, 2014) (29 markers = c. 35%)

Following on from the most robust category, i.e. the grammaticity index, the next section will examine analyticity and syntheticity in more detail and locate ELF among the various world English varieties.

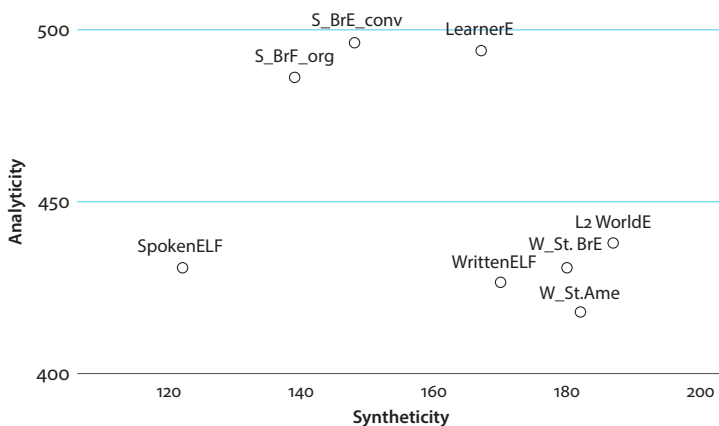
## 4.2 ELF on a two-dimensional plane

The backdrop to this section is the observation that the variety types differ substantially on a two-dimensional analyticity-by-syntheticity plane. Space permits me to illustrate some of the previous findings only briefly, but they are explicitly explained in the sources used in this section. Despite the risk of oversimplifying matters, it is fair to say that the findings can be summarized as follows. On the one hand, Szmrecsanyi (2009) observes that the traditional regional dialects found in the British Isles are more synthetic than the varieties labeled as high-contact varieties. This latter group forms a heterogeneous set of varieties. They exhibit a considerable spread in which indigenized L2 varieties (East African English, Indian English, Jamaican English, Hong Kong English, Singapore English and Philippine English) form a clearly distinct group. This group is different not only from standard BrE and AmE, but also from language-shift Englishes (Irish and Welsh English) and transplanted L1 Englishes (i.e. New Zealand English and spoken AmE). The indigenized L2 varieties can be further divided into Southeast Asian L2 varieties, which are substantially less analytic and synthetic than “non-Southeast Asian L2 varieties”



(Szmrecsanyi 2009: 328). Standard AmE is slightly less synthetic than BrE. On the other hand, Szmrecsanyi & Kortmann (2011: 182) observe that traditional learner essay data in the *International Learner Corpus of English* (ICLE) is less synthetic but clearly more analytic than Standard BrE. This observation forms the basis for the two authors to draw a distinction between learner language and second language varieties on structural grounds.

Figure 2 visualizes the two-dimensional analyticity–syntheticity plane, setting the written and spoken ELF results side-by-side with some of the results presented in Szmrecsanyi (2009) and Szmrecsanyi & Kortmann (2011). Note that the learner English data consist of all the ICLE results combined and the same holds for the indigenized L2 Englishes. The spoken British English data are from of the spoken genres in Szmrecsanyi (2009: 333) and are used as a point of comparison for VOICE. On the ELF side, the spoken data are from VOICE in their entirety. The written ELF results exclude tweets and are based on 930,185 words in WrELFA and SWE-CE.



**Figure 2.** Written and spoken ELF compared with select variety types in Szmrecsanyi (2009) and Szmrecsanyi & Kortmann (2011)

The results visualize how spoken and written ELF could be positioned relative to a select set of English varieties. For spoken ELF, the synthetic value is 122, and the analytic value 431, while the respective values for written ELF are 170 and 427. The differences in the analytic values are not statistically significant (log-likelihood (LL) value 1.58,  $p > 0.05$ ), but they are highly significant for the synthetic values (LL 13.16,  $p < 0.001$ ). This finding is slightly different from our preliminary observations (Laitinen, Levin & Lakaw in press), in which our written sample consisted only of formal academic and news genres. However, they do not change the main observation indicating that considerable differences exist between spoken

and written ELF use. Figure 2 illustrates how the differences, which were already visible in grammaticity (Table 3), are brought about by a smaller share of synthetic markers in the spoken data. The Pearson residuals vary between  $-1.554$  and  $1.496$ , but the effect size in Cramer's phi for these nominal variables is minimal ( $0.073$ ).<sup>1</sup>

The result regarding the differences between spoken and written modes in ELF is similar to that observed in other varieties in Szmrecsanyi (2009). His result shows that all of the major varieties of English exhibit similar decreases in analyticity and increases in syntheticity between their spoken and written modes of production. The ELF evidence is not random but conforms to the general pattern relative to the mode of production. However, it needs to be pointed out that VOICE exhibits lower index values for both analyticity and syntheticity relative to spoken British English, thus clearly highlighting the emergent characteristics of ELF in which meanings are negotiated in interaction.

More importantly for the ELF debate, the differences indicate increased transparency and output economy only in spoken ELF, but not necessarily on the written side. No such tendency is discernible in the written data, and more research needs to be carried out on the structural properties of written ELF.

Another important feature in Figure 2 is that the observations indicate substantial differences between traditional learner data and ELF. They confirm that, on purely structural grounds, language acquisition in foreign language settings should be viewed differently from second language use (cf. Mauranen 2011 on the notion that acquisition and use are connected but dissimilar). The two forms of non-native English are different. Written learner data exhibit close similarities with spoken native varieties, and Granger and Rayson (1998) suggest that such tendencies are discernible in register interferences and the over-representation of speech-like features in learner data. However, the results here show that similar tendencies cannot be detected in my written ELF data. The latter are more synthetic and substantially less analytic than learner English data. The total figures for written ELF are 423 analytic markers and 170 synthetic ones, and for learner English they are 494 and 167 (according to Szmrecsanyi & Kortmann 2011). The result is statistically significant for the analytic markers (LL 5.50,  $p < 0.05$ ) but not for the synthetic ones ( $0.03$ ,  $p > 0.05$ ).

---

1. Pearson residuals are utilized to check whether the observed values in two-dimensional data are larger or smaller than the expected frequencies (cf. Levshina 2015: 120). This method makes it possible to observe the effect of the dependent variable. The values that are smaller than  $-3.841$  or greater than  $3.841$  are considered to be particularly noteworthy, and their effect is more pronounced than those that fall between. Cramer's phi is a post-test used in determining the strengths of association between two variables and is a measure of association ranging between 0 and 1.

The total frequencies of written ELF in Figure 2 suggest that it crops up within the broad group of standard BrE and AmE (data from the Freiburg versions of the Brown corpora in Szmrecsanyi 2009) and the indigenized L2 varieties. On the whole, comparisons of the syntheticity indices between written ELF and written BrE (LL 0.06,  $p > 0.05$ ) and AmE (LL 0.02,  $p > 0.05$ ) show no statistically significant differences. As for the analytic indices, the same holds true: there are no statistically significant differences in the data.

Thus far, I have considered ELF only as spoken and written modes of communication and have compared these two with the other varieties. In the next section, I will focus on ELF, and explore the extent to which the various written genres are structurally different from each other and from the spoken evidence in VOICE. Some comparative evidence from BrE is included.

### 4.3 Genre differences in ELF

The results in the previous sections establish that ELF (i.e. second language use) is structurally different from EFL (i.e. second language acquisition) and similar to other L2 uses of English in terms of both grammaticity and analyticity by syntheticity. These results, based on a large set of aggregate data, not only confirm a similar assumption in the ELF literature (Mauranen 2011), but they also illustrate correspondences between written ELF and the major standard varieties of English. Differences exist between ELF and Standard English, but these are more pronounced on the spoken side than on the written, as illustrated by the results in Table 3 and Figure 2. These results are important, considering the status of ELF in general, since they show that it is a structurally distinct variety type. As pointed out in the introduction, I am only referring to its structural properties here, as we should be careful in assessing the sociolinguistic angle of ELF being a focused variety.

A key question in this section is the systematicity of ELF genres. If the genre differences are systematic, so that both spoken data and the spoken-like written genres and the various written genres (see Figure 1 above) exhibit similar tendencies as in the native varieties, the results should indicate that ELF speakers show at least some degree of the stability required of a focused variety and exhibit awareness of genre characteristics in terms of structural features.

These previous findings form the backdrop to the quantitative observations presented here. Figure 3 shows how the spoken VOICE and the five written ELF genres locate on a three-dimensional plane that integrates analyticity ( $y$ -axis), syntheticity ( $x$ -axis) and grammaticity ( $z$ -axis).

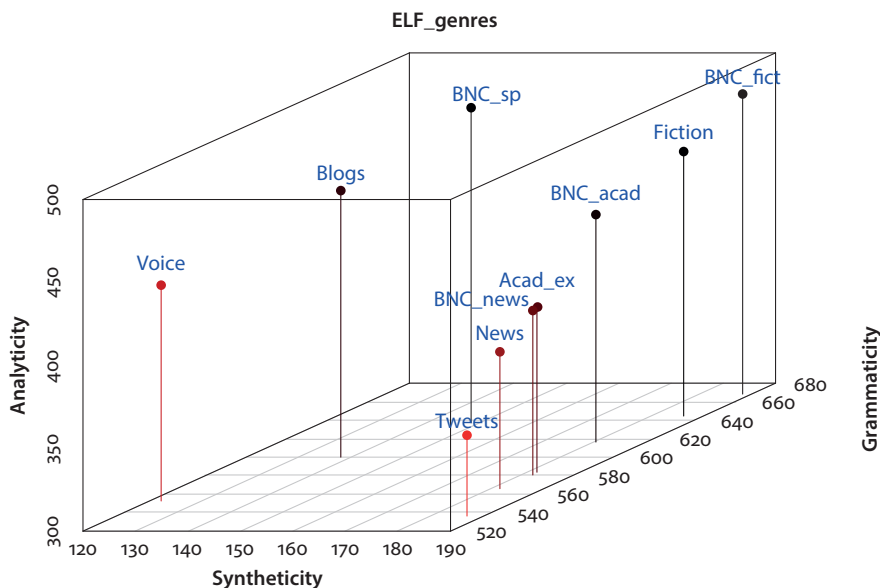


Figure 3. Spoken ELF data and written ELF genres in a three-dimensional genre space

I have included three written genres from Szmrecsanyi (2009), marked “BNC\_ genre” in an abbreviated form (i.e. academic journal articles, news, and fiction). In addition, it includes the spoken British English (BNC\_sp) results, which are the arithmetic mean figures of 16 spoken genres (2009: 333).

The results show, firstly, that spoken VOICE and the most interactive written ELF genre, tweets, are highly dissimilar. Recall that Table 3, above, demonstrates that both of them exhibit low grammaticity values, while Figure 3 shows that spoken material is characterized by higher frequencies of analytic markers, whereas tweets have more synthetic markers. This is also confirmed by the Pearson residuals, which indicate that in this regard the observed frequency of the synthetic markers is higher than expected (2.831). The Pearson residual value for the analytic markers in VOICE is (1.754). In all, Cramer’s phi for nominal variables indicates that the effect size is small (0.142). These observations suggest that spoken ELF with its slightly increased analyticity highlights transparency and its negotiation of meaning through explicit analytic marking, but the same tendency is not true for interactive tweets. In tweets, economy and the compression of information to 140 characters weighs more, but the correlation coefficient remains mild. The tweet component shows characteristics of more formal genres, such as news. This finding is similar to that observed in native tweets in Knight et al. (2014), whose evidence comes from the relative frequencies of broad syntactic categories.

The blog data visualized in Figure 3 stand out not only from tweets but also from the other written ELF genres. They are closest to spoken data and exhibit lower frequencies in syntheticity than do the other written datasets. The Pearson residuals are highest for these synthetic markers, but they are not outside the critical values 3.841 and  $-3.841$ . Similarly, Cramer's phi shows no significant effect ( $-0.052$ ). The finding goes against the observations in Knight et al. (2014), whose results on the native side show that blogs display the characteristics of formal genres.

When the two spoken corpora are compared with each other, the differences in the data are small. There is a slightly higher share of analytic markers in the BNC than in VOICE, and the same holds true for the synthetic markers. However, the Pearson residuals of this variable (analytic vs. synthetic in two datasets) show that the differences between the observed and the expected frequencies suggest no noteworthy differences. Cramer's phi, measuring the effect size, falls close to zero (0.014).

The analytic and synthetic indices for the ELF news genre are 383 (AI) and 182 (SI), and the differences are not statistically significant when we compare them to the BNC\_news frequencies. The Pearson residuals vary between 0.189 and  $-0.187$ , while Cramer's phi shows no effect (0.009). In this genre, ELF and standard native English are therefore very close to each other, which is not surprising. According to Hundt and Mair (1999: 236), news as a genre tends to be agile, and authors (journalists) in native settings tend to be "receptive to" change and innovation. It would be unusual to assume that they would not be in ELF. In addition, it is likely that as language professionals, native and non-native journalists alike must be aware of their language production. The only marginal difference between ELF and native Standard English is the higher analytic score in the latter, but the correlation coefficient remains low.

In the academic genre, the datasets represent highly formal and informative academic writing, but the difference is that the ELF evidence comes from examiners' statements, whereas the BNC material has been taken from academic articles. Despite these differences, the quantitative patterns observed in the material are highly similar, with no statistically significant differences (the Pearson residuals vary between 0.462 and  $-0.449$ , while Cramer's phi shows no effect of correlation 0.022). The only difference is the slightly higher analytic index in the native dataset when compared with ELF, but not at any statistically significant levels.

Finally, the same tendency of remarkable genre similarities continues in fiction. On the ELF side, the analyticity index is 460 and that of syntheticity 186, while in the BNC data the corresponding values are 481 and 188. The differences are not statistically significant (the Pearson residuals vary between 0.167 and  $-0.164$ , while Cramer's phi shows no effect of correlation 0.007), and ELF is highly similar to native data. The finding is noteworthy since our fiction component is closest to learner writing. The texts were collected from creative writing courses in Sweden, but they

exhibit only a little of the traditional learner language characteristics in the indices. One reason for this close similarity could be connected with our informants: people learning to write fiction in their L2 could be aware of their language production capabilities, which results in these close similarities.

## 5. Discussion and implications

In this last section I will first provide a brief overview of my observations. I will then go on to explore some of the implications of these observations and aim at connecting them to the theoretical models of World Englishes in general and to issues related to the study of ELF in particular.

The results presented are the first ELF results obtained using the typological profiling method, and they enable the assessment of the variety status of ELF. They show that on purely structural grounds ELF is another distinct variety type among the English varieties. The quantitative patterns observable in the data are clear: Second language use is structurally different from second language acquisition, and there is a quantitative basis for drawing a distinction between ELF and traditional learner data (EFL) using purely structural criteria, as has been attempted here.

The results in Section 4.1 show that, with regard to grammaticity, ELF is similar to the many indigenized L2 Englishes. There exist substantial differences between spoken and written modes, and new genres, such as tweets, which are characteristic of the globalization of English, behave quite distinctly from the more traditional genres. Section 4.2 shows that, when it comes to ELF and native evidence, spoken ELF is structurally different from spoken native data. This observation can be compared to Mauranen et al.'s (2015: 402) observations. They point out that “even a short fragment of ELF talk heard or seen in transcription is usually enough to tell it is not ENL” [i.e. English as a native language]. My results offer quantitative confirmation of this. However, they also contrast some of the previous findings in ELF and add another angle to them. Mauranen et al. (2015) continue that when it comes to “word lists of individual word and n-grams” there is “notable overall similarity” between ELF and native uses in academic settings. My results show that spoken ELF is lower in terms of both syntheticity and analyticity when compared with native spoken data. It is important to note that Section 4.3 shows that no such differences can be discerned on the written side. When we use analyticity and syntheticity as an index of structural similarities, written ELF is not distinguishable from native data.

The empirical results enable refuting the null hypothesis, and hence they compel us to rethink the traditional tripartite division of Englishes. For instance, studies that have explored closing the paradigm gap between learner varieties and post-colonial indigenized L2 uses have suggested that the EFL–ESL should be seen

as a continuum (cf. Mukherjee & Hundt (eds) 2011). However, the results here show that such a discussion excludes a crucial component, *viz.* a structurally distinct variety type of ELF. One possible explanation for such an exclusion is the fact that ELF is still seen to be limited to certain domains. However, as the corpus materials used here show, it is clear that the underlying determinants of the expansion of English, such as digitization and technologization, have led to a situation in which ELF serves a much broader set of communicative functions than simply certain specific purposes. This is clearly seen in the digital genres covered in WrELFA, such as scientific blogs, which were not included in the results here, and in personal and thematic blogs as well as in tweets. The tweet component serves as a prime example of technologization and digitization since the material for the second-generation ELF corpus used here comes from a government-funded site that recruits ordinary citizens to manage the official Twitter account of Sweden. The great majority of these messages are in English; indeed the entire site is in English. Despite this fact, which derives from a single, specific case, it points to the question of the extent to which the current corpora used in the study of World Englishes can indeed capture the diversification of English and whether many of the corpora that are used actually reflect the real world. The second-generation ELF corpora used in this study offer one corpus design model that enhances representativeness. At the same time, it needs to be acknowledged that we also need more comprehensive geographic coverage of written ELF corpora, since the observations here are based on the Nordic context.

Lastly, one of the key points of ELF in Mauranen's (2012) study is that the emergence of second language use and the global spread of English add extra uncertainty to what we know about language change. She argues that ELF may lead to a situation where "we do not know in which respects the processes observed in earlier research on language change are valid" (2012: 243). One example of an area where rethinking is needed is in dissecting the observation that the spread of English through contact and adult language learning leads to simplification. The results presented here illustrate that some simplification takes place, especially in the most robust category of grammaticity, but there are also areas in which structural simplification is not present. The ELF corpora here display remarkable similarities with native Englishes and post-colonial L2 varieties, and new theoretical approaches are needed to understand such observations.

In a separate study, I have, together with my colleagues, explored the idea of applying variationist sociolinguistics as the theoretical toolbox necessary for understanding ELF (Laitinen et al. 2017). We used the social network model of the diffusion of innovations as the starting point (cf. Milroy & Milroy 1987). Of interest was the idea whether social network structures and an increase in weak ties in multilingual settings in particular might be used to account for some of the macro-level

developments observed above. The social network model is well known in variationist sociolinguistics but is rarely used in studying World Englishes. Sociolinguists have shown that dense and multiplex network structures tend to impose norms on their members and therefore promote language maintenance, whereas loose network structures lead to increased linguistic variation (Milroy & Milroy 1987). The model suggests that people are likely to accommodate to each other linguistically in weak-tie contact situations, and that contacts of this kind lead to the eradication of marked variant forms and therefore tend to result in conditions that are favorable to language change.

One of our suggestions is that ELF speakers, who are multilingual by definition, might have a larger number of weak ties in general than those who do not use English as a second language resource. These multilingual individuals could act as agents of linguistic change. Our results come from a 'big data' network of nearly 200,000 Twitter accounts in the Nordic region, where English is often used as ELF. We made use of two parameters that are automatically generated and available for third-party users in the Twitter stream. Our main finding is that those who tweet in English have a substantially larger number of network ties than those who primarily use the main L1s of the region (Danish, Finnish, Icelandic, Norwegian and Swedish) in their communication. This result supports the idea that the ELF settings and multilingual speakers in general favor innovation and change, and such settings and speakers might therefore offset part of the impact of simplification that normally takes place in adult language acquisition.

## Acknowledgement

I want to thank the two anonymous reviewers for their valuable comments on an earlier version of this article. I have also benefitted greatly from comments made by Professor Anna Mauranen and from the discussions and comments at the Changing English conference in Helsinki in 2015, ISLE-4 in Poznan in 2016 and ICAME38 in Prague in 2017. The usual disclaimers apply.

## References

- Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: CUP.  
<https://doi.org/10.1017/CBO9780511621024>
- Biber, D. & Conrad, S. 2009. *Register, Genre and Style*. Cambridge: CUP.  
<https://doi.org/10.1017/CBO9780511814358>
- Breiteneder, A. 2009. English as a lingua franca in Europe: An empirical perspective. *World Englishes* 28(2): 256–269. <https://doi.org/10.1111/j.1467-971X.2009.01579.x>
- Cogo, A. & Dewey, M. 2012. *Analysing English as a Lingua Franca: A Corpus-driven Investigation*. London: Continuum.



- Granger, S. & Rayson, P. 1998. Automatic profiling of learner texts. In *Learner English on Computer*, S. Granger (ed.), 119–131. London: Longman.
- Greenberg, J. 1960. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26(3): 178–194. <https://doi.org/10.1086/464575>
- Grieve, J., Biber, D., Friginal, E. & Nekrasova, T. 2010. Variation among blog text types: A multi-dimensional analysis. In *Genres on the Web: Corpus Studies and Computational Models*, A. Mehler, S. Sharoff, & M. Santini (eds). New York NY: Springer.
- Hundt, M. & Mair, C. 1999. “Agile” and “uptight” genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4(2), 221–242. <https://doi.org/10.1075/ijcl.4.2.02hun>
- Jenkins, J, Cogo, A. & Dewey, M. 2011. Review of developments in research into English as a lingua franca. *Language Teaching* 44(3): 281–315. <https://doi.org/10.1017/S0261444811000115>
- Kachru, B. B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the World: Teaching and Learning the Language and Literatures*, R. Quirk & H. G. Widdowson (eds). Cambridge: CUP for The British Council.
- Knight, D., Adolphs, S. & Carter, R. 2014. CANELC: Constructing an e-language corpus. *Corpora* 9(1): 29–56. <https://doi.org/10.3366/cor.2014.0050>
- Leppänen, S. 2012. Linguistic and generic hybridity in web writing: The case of fan fiction. In *Language Mixing and Code-Switching in Writing*, M. Sebba, S. Mahootian, & C. Jonsson (eds), 233–254. London: Routledge.
- Laitinen, M. 2016. Ongoing changes in English modals: On the developments in ELF. In *New Approaches in English Linguistics: Building Bridges* [Studies in Language Companion Series 177], O. Timofeeva, S. Chevalier, A-C. Gardner, & A. Honkapohja (eds), 175–196. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.177.07lai>
- Laitinen, M. & Levin, M. 2016. On the globalization of English: Observations of subjective progressives in present-day Englishes. In *World Englishes: New Theoretical and Methodological Considerations* [Varieties of English Around the World G57], E. Seoane & C. Suárez-Gómez (eds), 229–252. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g57.10lai>
- Laitinen, M., Levin, M. & Lakaw, A. In press. Charting new sources of lingua franca data: A multi-genre corpus approach. In *From Data to Evidence in English Language Research*, C. Suhr, T. Nevalainen, & I. Taavitsainen (eds). Leiden: Brill.
- Laitinen, M., Lundberg, J., Levin, M. & Lakaw, A. 2017. On the globalization of English: Making use of complex Twitter data as a diagnostic tool in language choice. *Journal of Universal Computing Science*.
- Laitinen, M., Lundberg, J., Levin, M. & Lakaw, A. 2017. Utilizing multilingual language data in (nearly) real time: the case of the Nordic Tweet Stream. *Journal of Universal Computer Science* 23(11), 1038–1056.
- Levshina, N. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.195>
- Low, E-L. 2016. A features-based description of phonological patterns in English as a lingua franca in Asia: Implications for standards and norms. *Journal of English as a Lingua Franca* 5(2): 309–332. <https://doi.org/10.1515/jelf-2016-0022>
- Mair, C. 2013. The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34: 253–278. <https://doi.org/10.1075/eww.34.3.01mai>

- Mauranen, A. 2011. Learners and users – who do we want corpus data from? In *Taste for Corpora: In honour of Sylviane Granger*, F. Meunier, S. De Cock & G. Gilquin (eds.), 155–171. Amsterdam: John Benjamins.
- Mauranen, A. 2012. *Exploring ELF: Academic English Shaped by Non-Native Speakers*. Cambridge: CUP.
- Mauranen, A., Carey, R. & Ranta, E. 2015. New answers to familiar questions: English as a lingua franca. In *Cambridge Handbook of English Corpus Linguistics*, D. Biber & R. Reppen (eds), 401–417. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139764377.023>
- Milroy, L. 1987. *Language and Social Networks*, 2nd edn. Oxford: Blackwell.
- Milroy, J. & Milroy, L. 1987. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21: 339–384. <https://doi.org/10.1017/S0022226700010306>
- Mukherjee, J. & Hundt, M. (ed.). 2011. *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap* [Studies in Corpus Linguistics 44]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.44>
- Pietikäinen, K. 2017. ELF in social contexts. In *The Routledge Handbook of English as a Lingua Franca*, J. Jenkins, W. Baker, & M. Dewey (eds), 321–332. London: Routledge.
- Ranta, E. 2013. Universals in a Universal Language? Exploring Verb-Syntactic Features in English as a Lingua Franca. *Tampere: Acta Electronica Universitatis Tampereensis* 1366.
- Schneider, E. W. 2007. *Postcolonial English: Varieties Around the World*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511618901>
- Seidlhofer, B. 2011. *Understanding English as a Lingua Franca*. Oxford: OUP.
- Seidlhofer, B., Osimk-Teasdale, R. & Radeka, M. 2014. *Parts-of-Speech Tagging and Lemmatization Manual*, 1st rev. version. Vienna: VOICE Project.
- Szmrecsanyi, B. 2009. Typological parameters of interlingual variability: Grammatical analyticity vs. syntheticity in varieties of English. *Language Variation and Change* 21(3): 319–353. <https://doi.org/10.1017/S0954394509990123>
- Szmrecsanyi, B. 2012. Analyticity and syntheticity in the history of English. In *The Cambridge Handbook of the History of English*, T. Nevalainen & E. Closs Traugott (eds), 654–665. Cambridge: CUP.
- Szmrecsanyi, B. & Kortmann, B. 2011. Typological profiling: Learner Englishes versus indigenized L2 varieties of English. In *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap* [Studies in Corpus Linguistics 44], J. Mukherjee & M. Hundt (eds), 167–187. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.44.09kor>



# Modeling World Englishes from a cross-linguistic perspective

Peter Siemund

University of Hamburg

The increasing diversification of English qua World Englishes contributes to cross-linguistic variation. Still, we tend to consider and model varieties of English from a language-internal perspective. Taking Mair's (2013) World System of Englishes as a starting point, I here explore how to model World Englishes from a cross-linguistic typological perspective, commenting on the tension between normative pressure and cross-linguistic tendencies and generalizations. Variation studies and language typology investigate micro and macro variation, respectively, though the empirical domains in focus and the methodologies employed show considerable overlap. Moreover, the traditional distinction between language and variety becomes increasingly difficult to draw in today's multilingual and highly interconnected world. I examine the commonalities and differences of the two approaches to language variation that have largely been working independently of one another focusing on language universals and the ways that grammatical phenomena from World Englishes match up against them. Varieties of English attest a good deal of typological variation, though they also offer curious features rarely found in other languages.

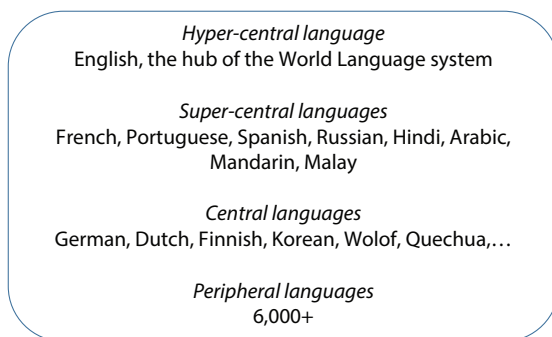
**Keywords:** macro variation, micro variation, typology, variation studies

## 1. World Englishes and cross-linguistic diversity

Just as there is cross-linguistic diversity, there is diversity of Englishes. Due to definitional problems, the exact number of languages in the world is difficult to assess, but there is widespread consensus that some figure between six and seven thousand languages represents a reasonable estimate. Kortmann and Lunkenheimer (2013) offer information on seventy-six varieties of English, though this figure is most certainly the result of practical constraints and hence just as open to debate as the precise number of languages in the world. As Mair (2013: 254) reminds us, different varieties of English tend to be construed in terms of territorial boundaries,

often representing political borders. Such territories may harbor more than one variety that are in principle distinguishable (for example, several Indian or Chinese Englishes). Conversely, one variety may be used across several territories (as, for example, Nigerian Pidgin across West African countries, or Singlish in Singapore, Malaysia, and Indonesia).

Languages, as a concomitant of the societies in which they are spoken, possess social value or prestige, as is well known, although we explicitly ignore this fact in the scientific exploration and analysis of languages. As a matter of fact, there are good reasons for focusing analytic effort on the less prestigious languages, as these tend to be under pressure, often lack documentation, and require more attention. If languages represent social capital and if they possess different social values, a hierarchy of languages emerges in the same way as there is a hierarchy of national prestige or economic power. De Swaan's (2001) World Language System nicely models these observations, as illustrated in Figure 1. The position of a language in this hierarchy is motivated by the number of speakers for whom it is relevant – socially, culturally, or economically. Evidently, English – as the hyper-central language – is relevant for everybody these days, whether they live in Canada, Germany, Kazakhstan, or the Malaysian rainforest. Globally speaking, it possesses the highest social capital. In contrast, a small clan language spoken by one hundred people in the Amazon basin receives its social value primarily from being used in this particular speech community. Accordingly, it counts as a peripheral language in de Swaan's model, most languages being peripheral in this sense. A small number of internationally relevant languages are placed amongst the super-central and central languages. These also represent important social capital for substantial numbers of people.

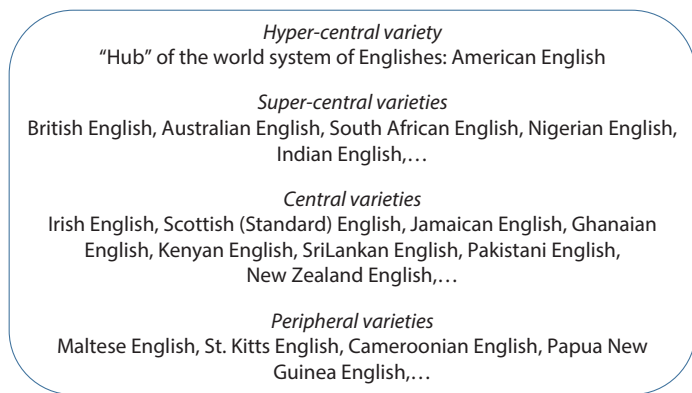


**Figure 1.** The World Language System according to de Swaan (2001: 4–6)

A similar hierarchy of languages can also be established within nation states, where the national language (for example, German in Germany, French in France, Russian in Russia, etc.) can be opposed to locally prestigious foreign languages (say, English,

Spanish, and French in Germany; English and German in Poland; Chinese and Korean in Kazakhstan, etc.), i.e. languages used and taught in the respective education systems, officially acknowledged minority languages (Frisian in Germany, Basque in Spain, etc.), non-official minority languages (e.g. all Native American languages in Canada), and immigrant languages (Turkish in Germany, Mandarin in Australia, Hindi in Dubai, etc.) – the relative level of prestige of these language groups being subject to local variation.

Building on de Swaan's model, Mair (2013) firstly proposes to extend this model to a classification of standard World Englishes, distinguishing a hyper-central variety, a few super-central varieties, a small number of central varieties, and a great majority of peripheral varieties. This extension of de Swaan's model is shown in Figure 2. American, rather than British English is here considered the most important variety of English. Post-colonial Englishes such as Nigerian and Indian English are listed among the super-central varieties, since they are relevant for very large numbers of speakers. We may note that from a global perspective, Irish English, New Zealand English, Jamaican English, Ghanaian English, etc. have the same status as central varieties, which makes good sense, even though the economic power behind the relevant countries is very different.



**Figure 2.** The World System of Standard Englishes according to Mair (2013: 261)

The mutual influence of the varieties in this model and the transmission of features from one variety to another is generally held to be top-down and not bottom-up. For example, American English is substantially more likely to influence the super-central, central, and peripheral varieties than the other way around. In a similar way, the super-central varieties are more important for the central and peripheral varieties than vice versa. Borrowings from American English can be easily detected in British, Australian, Indian, Irish, and Ghanaian English, though the opposite direction of lexical transfer is less likely to occur. The same holds for

norms of pronunciation, orthographical conventions, and perhaps even grammatical features, although influence in the latter domain can primarily be expected to manifest itself in terms of distributional shifts. For example, one could expect that the relative ratios of past tense and present perfect use in British English are slowly moving towards the North American norm. Similarly, British English serves as a model for Irish English, Indian English exerts influence on Sri Lankan English, and Australian English on New Zealand English and the variety of English spoken on Papua New Guinea. These varieties are epicenters in their respective region of influence (Mair 2013: 259; see also Peters 2009).<sup>1</sup>

Mair (2013: 264) further proposes to extend the model shown in Figure 2 so as to include non-standard varieties of English. Crucially, we can also distinguish super-central, central, and peripheral non-standard varieties alongside the corresponding standard varieties. For example, African American Vernacular English and Jamaican Creole count as super-central non-standard varieties, since they exert substantial influence on several standard and non-standard varieties through popular culture (e.g. *hip-hop*, *rasta talk*). The more academic counterpart to these super-central varieties of popular culture can perhaps be found in Lingua Franca English, in which the acceptance of widely encountered non-standard pronunciations (missing *th*), morphological simplifications (missing third person *-s*), lexical recategorizations (*informations*, *evidences*, etc.), and the occurrence of higher ratios of parataxis in comparison to hypotaxis is generally high. Again, some of these non-standard features may find their way into other varieties in the long run, due to the large and increasing number of speakers of English as a Lingua Franca.

## 2. Variation studies and language typology

Both variation studies (dialectology) and language typology have an extensive history. Variation studies originates in dialectological work carried out in the nineteenth and twentieth century primarily in Europe that produced comprehensive dialect atlases. For example, the dialectal landscape of French, English, and German is covered in the *Deutsche Sprachatlas* (1927–1956) coordinated by Georg Wenker, the *Survey of English Dialects* directed by Harold Orton (Orton & Dieth 1962–1971), and the *Atlas linguistique de la France* by Jules Gilliéron (Gilliéron & Edmont 1902–1914). These typically focused on accentual and lexical differences, largely excluding grammatical phenomena. Typological work also brought about comprehensive language atlases, notably the *World Atlas of Language Structures*

---

1. Mair (2013: 259) defines epicenters as “focal points in the pluricentric constellation whose pull is not quite as strong as that of British or American English.”

authored by Dryer and Haspelmath (2013), but these appeared only recently and primarily explore grammatical differences. Nevertheless, the comparative method underlying it can be traced back into the eighteenth and nineteenth century, important contributions emanating from von Humboldt (1836) and Schlegel (1818).

Cross-linguistic typological studies, in their endeavor to identify the patterns and limits of structural variation, strive to work on the basis of balanced samples of languages that represent the 250 or so language families in more or less equal proportions. Balanced samples of 200 to 500 languages can be considered as extremely meaningful. The implicit assumption behind this sampling method is that all languages have the same scientific value and are independent from one another. Evidently, research methodology here steers clear of cultural and social status as well as language contact (see Siemund 2013). These assumptions, of course, represent a perfectly reasonable and understandable scholarly perspective in which scientific interest is disconnected from social status, and various confounding factors are screened out. What this methodological approach usually does not take into consideration, however, is that English, by now, has probably influenced most, if not all other languages in this world and continues to exert strong pressure. It is without doubt safe to assume that there are only few languages without at least a few borrowings from English, even strongly protected languages such as French or languages spoken in remote pockets of the rainforest. Structural influence is much harder to pin down, but certainly not unlikely given the prominence of English and the vast number of bilingual speakers. In a similar way, the so-called ‘super-central’ and ‘central’ languages have been exerting influence in their respective habitats, especially Spanish in Latin America, French in parts of Africa, German in Eastern Europe, Russian in the Russian Federation, as well as Chinese, Arabic, and Malay in their areas of influence. The issue of language contact tends to get neglected in genealogical models of languages, but also in models based on structural parameters (see Onysko 2016 for a classification of World Englishes based on language contact). However, the majority of people in this world is multilingual, thereby living in language contact situations.

In the study of World Englishes, or varieties of English for that matter, we equally abstract away from social and cultural differences as well as issues of language contact, happily comparing American English with Indian English, New Zealand English with Irish English, or Ghanaian English with Singapore English.<sup>2</sup> The International Corpus of English (ICE) frequently serves as data base for such

---

2. In our current understanding, the term “variety of English” also covers English based pidgins and creoles, even though the social conditions leading to these varieties (or languages) are quite different from those found in the context of other varieties. Nevertheless, ongoing discussions in the context of such pidgin and creole languages show that the influence of traditional British dialects on the development of these languages had long been underrated (Mufwene 2001).



comparisons. This, of course, can be very instructive, but the interpretations of the observable differences rarely address the social hierarchy behind the relevant varieties. Mair (2013) – correctly to my mind – points out that our predominant territorial conception of World Englishes is increasingly becoming meaningless, since speakers from different varieties frequently meet due to easy and fast global transport. Moreover, large numbers of multilingual speakers are in constant contact using social media. In today’s world, it is growing to become insignificant where we are located physically.

If it is the case, however, that territorial borders have become highly permeable and speakers can interact with one another quite independently from where they are, the question where one language ends and another begins becomes increasingly difficult to answer. The same holds, *mutatis mutandis*, for varieties of English and of other languages. For example, Chinese diasporas can be encountered all over the world. Western European and North American territories are interspersed with immigrant communities of various origins, making it difficult to find monolingual speakers of the relevant national languages in certain neighborhoods. In this increasingly multilingual and interconnected world, anything can potentially influence everything. Can I really exclude that certain aspects of Chinese grammar enter my English or German through interaction with my Chinese students in English?

If these considerations are correct, they also have consequences for comparative research on languages and their so-called ‘varieties’, since there seems to be no principled difference any longer between comparing languages with languages, varieties of a language with other varieties of this language, or between languages and varieties. Traditionally, these research paradigms used to work side by side, quite independently from another, with functional typology investigating cross-linguistic variation (“macro-variation”) and dialectology, sociolinguistics, and variation studies being dedicated to charting intra-language variation (“micro-variation”; Davydova et al. 2011; Siemund 2009). Nevertheless, typological studies comparing languages of the Germanic, Romance, or Malay families could be easily listed under variation studies, since the languages are typologically very similar (Spanish and Italian, German and Dutch, different forms of Malay). Equally, some traditional dialects of these languages can be markedly different, as for example, Bavarian, Occitan, and Kuching Malay, and could be used for typological comparisons. Most importantly, even though the scope of structural variation observable across language varieties (and World Englishes, for that matter) will – rather trivially – be smaller than that found in the world’s languages, the cognitive principles constraining the observable linguistic structures can be assumed to be the same (see Siemund 2013 for an overview).

What linguists working in variation studies call “variable” and “value” (or “variant”), translates into “categories” and their means of encoding in typology. Across

varieties of English, for example, the variable *-ing* may assume either the values /ɪn/ or /ɪŋ/, the third person singular agreement suffix – another variable – either *-s* or zero as values, and the variable relative marker may take up the values *what*, *that*, *at*, *as*, or *zero*, amongst others. The values are subject to well known conditioning factors such as region, age, ethnic background, sex, level of formality, i.e. topographic and social dimensions, as well as those relating to the situational and linguistic context. Typologists aim to identify grammatical categories and their means of encoding in the sampled languages. For instance, a cross-linguistic survey of the category of number reveals that languages may express number contrasts or leave them opaque. If they are expressed, languages may, *inter alia*, distinguish singular, dual, paucal, plural, etc. Moreover, number contrasts may be expressed in different ways (affixes, clitics, stem change, agreement, etc.; see Corbett 2000: Chapter 5). Crucially, languages may lack common grammatical categories, such as number, or possess rather idiosyncratic categories, as, for example, click sounds.

I here approach World Englishes primarily from a cross-linguistic typological perspective,<sup>3</sup> placing emphasis on the grammatical phenomena themselves. Functional typology, the framework within which cross-linguistic comparisons are primarily pursued, offers a particular model or conception of language that is especially suitable for comparative work. For example, it does not assume a closed syntactic framework, as in Universal Grammar, and tries to abstract away from language-specific categories, even though every language – strictly speaking – possesses its own grammar. The model of functional typology assumes a logically conceivable space of grammatical variation that is, however, constrained by functional pressure (economy) and dependencies between categories. It explores the patterns and limits of variation. The central research question addressed here is whether and to what extent the model of functional typology interacts with Mair's (2013) World System of Englishes. This is important, as researchers working in functional typology tend to pass over sociolinguistic constraints, while sociolinguists generally show little awareness of fundamental architectural constraints on language.

In Section 3 below, I discuss grammatical variation in World Englishes in relation to language universals, especially regarding the influence exerted by the hub-variety and the super-central varieties. Section 4 takes up for discussion the problematic notions of “angloversals” and “vernacular universals”. Section 5 offers explanations and motivations for the grammatical variation encountered, bringing in crucial information from historical migration, language contact, language acquisition, and grammaticalization.

---

3. For a similar approach, see Filppula (1999), Anderwald (2002), Herrmann (2003), Siemund (2004), Filppula et al. (2009), and Trudgill (2011).

### 3. Typological universals and World Englishes

Even though language typology is first and foremost taxonomic, researchers in the field widely share the belief that the logically conceivable space of language variation is heavily constrained. This metaphorical space is supposed to contain focal areas of cognitive domains and means of their encoding that can be consistently encountered in languages of different provenance. In addition, researchers assume that there are logical connections between grammatical domains such that the encoding of one domain can serve as a predictor for the encoding of others. As is well known, researchers talk about language universals in this respect that can be either absolute (i.e. exceptionless) or statistical, on the one hand, and conditional versus unconditional on the other. This crisscrossing of parameters yields the four basic types of universals summarized in Table 1. Conditional universals may be chained, leading to implicational hierarchies.

**Table 1.** Logical types of universal statement (following Greenberg), from Evans and Levinson (2009: 437)

	Absolute (exceptionless)	Statistical (tendencies)
Unconditional (unrestricted)	Type 1. "Unrestricted absolute universals" All languages have property X	Type 2. "Unrestricted tendencies" Most languages have property X
Conditional (restricted)	Type 3. "Exceptionless implicational universals" If a language has property X, it also has property Y	Type 4. "Statistical implicational universals" If a language has property X, it will tend to have property Y

To be sure, this cannot be the place to elaborate on language universals. Suffice it to say, perhaps, that true absolute universals are relatively rare and need to be conceptualized as very fundamental properties of language, such as constituent structure, recursion, the contrast between consonants and vowels, and the like. Instead, I will here focus on some examples of rather well-known implicational hierarchies and explore their relevance for varieties of English.

The distribution of reflexive markers, for example, is known to be governed by the implicational hierarchy in (1), which basically says that dedicated reflexive markers of the first person predict those of the second and third person. English possesses dedicated reflexive markers for all persons (*myself, yourself, himself/herself*), whereas German and Spanish only employ them in the third person (*sich; se*) and use simple personal pronouns to express reflexive relations in the first and second person (*mich, dich; me, te*). Such observations are consistent with the hierarchy in (1).

- (1) third person > second person > first person (Faltz 1985: 120)

Non-standard varieties of English frequently allow the use of simple personal pronouns in positions that require reflexive markers in the standard varieties (essentially object positions co-referential with the subject), as shown in the examples in (2).

- (2) a. Cape Flats English (McCormick 2004: 999)  
*I'm going to buy me biscuits and chocolates.*  
 b. Appalachian English (Christian 1991: 15–17)  
*He was looking to buy him a house for his family.*

However, there are important distributional differences with first person pronouns being vastly more prominent in these reflexive uses than second or third person pronouns. Moreover, and to the extent that this can be empirically tested, all varieties align themselves with the hierarchy in (1). Some archaic varieties preserve the situation of Old English, in which simple personal pronouns were consistently used to mark reflexive relations, as there was no dedicated reflexive marker. The parallels are illustrated in Examples (3) and (4).

- (3) Yorkshire English (Wright 1898–1905: volume iii, 164)  
 a. *He has cut him.* (= himself)  
 b. *He went to bathe him.* (= himself)
- (4) Old English (Aelfric's Grammar 96.11, Zupitza 1966)  
*hine he bewerað mid wæpnum*  
 him he defended with weapons  
 'he defended himself with weapons'

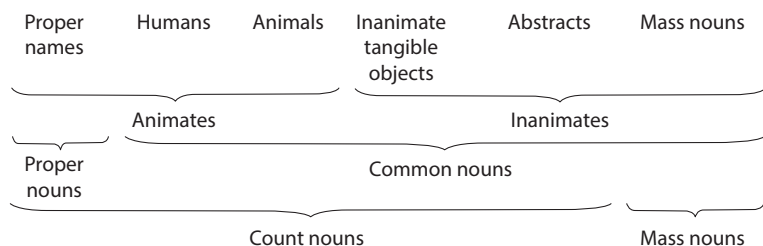
Even though non-standard varieties of English, thus, amply attest alternative typological states,<sup>4</sup> these are unlikely to spread due to the hierarchical organization of World Englishes. Feature percolation works top-down and not bottom-up, as argued in Section 1.

Conditional universals and implicational hierarchies may form the conceptual basis of semantic maps, i.e. graphic representations of related cognitive domains with topographic proximity being tantamount to cognitive relatedness (see Haspelmath 1990; Kemmer 1993; Siemund 2010, 2014). It is widely assumed that grammatical markers cover adjacent areas on such maps (i.e. that adjacent areas get encoded). Besides conditional relations, semantic maps may reflect polysemy

---

4. Other such alternative typological states include or concern, *inter alia*, missing copulas, negative concord, gender distinctions in personal pronouns, articles and the expression of definiteness/indefiniteness, distance contrasts in demonstratives, case marking, tense and aspect, the encoding of modality, and clause types (see Siemund 2013 for an overview). Some of these areas will be discussed in more detail below.

patterns or represent grammaticalization paths. One construct that may either be considered as a semantic map or an implicational hierarchy is the hierarchy or continuum of individuation (or animacy). It can be found in different versions in the literature, but they all rank nominal referents in terms of animacy (animate versus inanimate) and degree of individuation (boundedness). One version can be seen in Figure 3.<sup>5</sup> Descriptions of humans range at the top (left), while abstract and mass nouns are placed at the bottom (right). Nouns describing animals and inanimate objects go in between.



**Figure 3.** Morphosyntactic distinctions along a continuum of ‘individuality’ (Sasse 1993: 659; Siemund 2008: 4)

Typological research has uncovered several grammatical subsystems that are sensitive to this hierarchy, including person and number marking, word order, and case marking (Croft 2004).<sup>6</sup> As far as varieties of English are concerned, it is primarily pronominal gender and case marking that can be related to it.

The standard system of pronominal gender with its transparent split between animate (human) and inanimate referents (*he/she* versus *it*) appears with a seemingly completely different semantic basis in some traditional vernaculars of Southwest England (Somerset, Devon), Newfoundland, and Tasmania (related through migration). There, the neuter form *it* is used for masses, substances, liquids, as well as abstract concepts, while the animate forms *he* and *she* can be used for everything countable, animate or inanimate (Kortmann 2004: 1097; Pawley 2004: 616–628; Siemund 2008; Wagner 2004). The animate forms are used for different semantic domains, depending on the dialect, as can be seen in Example (5).<sup>7</sup>

5. One of the editors points out that this figure confuses noun types and nominal referents.

6. Even though animacy and individuation are not exactly the same conceptually, grammatical categories like person, number, gender, and case tend to reflect them so that they can be placed in the same hierarchy.

7. See Siemund (2008) for an extensive discussion of the use of masculine and feminine pronouns across a wide variety of semantic domains.

- (5) a. Southwest England (Siemund 2008: 43–45)  
 [What's the matter with your hand?]  
*Well, thòld horse mued on, and the body of the butt valled down, and he [the hand] was a jammed in twixt the body o' un and the sharps (bran-pollard).*
- b. Newfoundland English (Wagner 2004: 90, 274)  
*Put the cover an the chest again an' ... locked un up screwed un up, however they had done with un.*
- c. Australian (Tasmanian) English (Pawley 2004: 617, 620, 625)  
*That timber gun, she splits the log open.*

Example (6) illustrates the usage of neuter *it* in relation to liquids and abstract concepts. There is even a minimal pair in (6a).

- (6) a. Southwest England (Siemund 2008: 46)  
*Thick there cask 'ont hold, tidn no good to put it [the liquid] in he [the cask].*
- b. Southwest England (Siemund 2008: 35)  
*I sure you, mum, 'twas a terble awkward job, and I widn do it ageean vor no such money.*

The hierarchy of individuation shown above can help us to understand the relationship between the standard and non-standard systems of pronominal gender, since the split between animate (*he/she*) and inanimate (*it*) pronouns is simply shifted further to the right in the vernacular systems.<sup>8</sup> Interestingly enough, extensions in the use of *he* and *she* into the inanimate domains are also attested in the informal registers of North American and Australian English, but if they occur, they observe the hierarchy of individuation and refer to individuated entities (see Siemund 2008: 108; examples taken from Mathiot 1979):

- (7) [A workman trying to fix a key that keeps getting stuck in the lock said:]  
 ... that's why *she* is so hard to turn.
- (8) [the referent is a new Philco color television set; the speaker is a man]  
 Let's see how *she* goes!
- (9) [the referent in both cases is a cash register; the speaker is a clerk at A M & A]  
 Oh, *he's* the one who used to sit over there. Our other one was taken away. *He* must be sick!
- (10) [the referent is a vacuum cleaner; the speaker is a woman]  
 Oh yeah, sometimes *he* follows behind me like a good little guy. After all *he's* so little, you can't expect *him* to do much!

---

8. As a concomitant of this shift, the dialectal gender markers may be more appropriately considered as encoding a contrast in number (Siemund 2008: Chapter 10).

A somewhat different version of the animacy hierarchy can be found in (11). An important difference to the version shown in Figure 3 lies in the differentiation of the space above proper nouns (i.e. to the left of proper nouns), namely first, second, and third person pronouns.

- (11) 1 > 2 > 3 > proper nouns > human > animate > inanimate  
(adapted from Silverstein 1976: 122)

These distinctions are relevant for issues of case marking, the general typological observation being that if some noun class on this hierarchy shows case marking, all noun classes further to its left will also show case marking. The standard varieties of English offer a perspicuous exception to this generalization, as second person pronouns do not mark the otherwise available contrast between subject and object case (*you*). The Konstanz Universals Archive considers this an extremely exceptional grammatical property, a *rarissimum* so to speak:

Phenomenon: independent personal pronouns for 1st and 3rd (animate) person inflecting for both number and case, but that for 2nd person inflecting for neither category (defectiveness of 2nd person pronouns in number alone being more common)<sup>9</sup>

Non-standard varieties, as illustrated by the Shetland English examples in (12), offer grammatical subsystems that align themselves more closely with cross-linguistic generalizations (*du* versus *dee*), but, again, these do not seem to spread to other varieties due to normative pressure exerted by the World System of Englishes.<sup>10</sup> Shetland English needs to be considered a peripheral variety due to its small speaker base and geographical isolation.

- (12) a. Shetland English (Melchers 2004: 38)  
Du *minds me awful o dee grandfaider*.  
You remind me awful of you grandfather.  
'You remind me awfully of your grandfather.'
- b. Shetland English (Melchers 2004: 43)  
Set *dee doon*.  
Sit you down.  
'Sit down.'

Another widely discussed implicational hierarchy is Keenan and Comrie's (1977) accessibility hierarchy, as shown in (13). It captures the availability of noun phrases

9. <<https://typo.uni-konstanz.de/rara/nav/search.php?PHPSESSID=3tfkhj92l4kdmilghunc846u-ucplc8gh>> (3 June 2017).

10. We may note in this context that North American varieties have re-created the number contrast on second person pronouns (*y'all, you guys*), though not the case distinction.

for processes of relativization (i.e. forming the head of relative clauses) in relation to their syntactic function. It predicts that if noun phrases in a certain syntactic function can be relativized on, this will also be possible for noun phrases in all functions further to its left, though this needs to be assessed for different relativization strategies separately (gapping, non-reduction, pronoun retention, relative pronouns). Keenan and Comrie (1977: 92–93) argue that the hierarchy is reversed for pronoun retention.<sup>11</sup>

- (13) *Accessibility Hierarchy* (Keenan & Comrie 1977: 66)  
 Subject > Direct Object > Indirect Object > Oblique > Genitive > Object of Comparative

Again, the standard varieties of English offer an interesting exception to this hierarchy, since gapping as a relativization strategy is not available for subject referents (*The man ate the apple* -> *The man who ate the apple*; where the relative pronoun cannot be omitted). In non-standard regional Englishes, by contrast, subject gapping is pervasive, as shown in (14).

- (14) a. Southwest of England (Wagner 2004: 166)  
*You know anybody Ø wants some, he'll sell them.*  
 b. Appalachian English (Montgomery 2004: 278)  
*They is people Ø gets lost in these Smoky Mountains.*  
 c. Newfoundland English (Clarke 2004: 315)  
*There's no one Ø pays any attention to that.*  
 d. Australian Vernacular English (Pawley 2004: 637)  
*I knew a girl Ø worked in an office down the street there.*

Merely attesting the availability of subject gapping, however, fails to appreciate the full story, as the phenomenon is heavily constrained contextually and typically occurs in existential constructions (15a), cleft sentences (15b), and possessive constructions (15c) (Tagliamonte et al. 2005: 96). Such observations could serve as a reminder to typologists that mere attestation and distribution are quite different pairs of shoes.

- (15) Tagliamonte et al. (2005: 96)  
 a. *There's no many folk Ø liked going to the pit to work.*  
 b. *It was an earthen floor Ø was in that house.*  
 c. *I have a woman Ø comes in on a Thursday morning.*

---

11. Gapping is the most prominent strategy for subjects. Comrie and Kuteva (2013) provide the following distribution of the four strategies introduced above in a sample of 166 languages: gapping: 125; non-reduction: 24; relative pronoun: 12; pronoun-retention: 5. These figures represent the relativization on subjects.



Again, we may speculate that subject gapping could easily gain greater prominence, since it is expected according to the accessibility hierarchy, but that the internal organization of the World System of Englishes does not permit the necessary upward percolation of this grammatical feature.

Non-standard varieties of English also attest pronoun retention, referred to as “resumptive pronouns” in the relevant literature. These are pronominal copies of the head noun of the relative clause that replace the constituent relativized on. Some examples are shown in (16), but as Herrmann’s (2003) study reveals, their frequency of occurrence is relatively low. Resumptive pronouns are usually not available in standard English

- (16) a. Irish English (Filppula 2004: 85)  
*They jumped banks that time on the race-course that they wouldn’t hunt over them today.*
- b. Scottish English (Miller 2004: 63)  
 They’re the ones that the teacher thinks they’re going to misbehave.

Besides relative pronouns, gapping, as well as pronoun retention, the fourth strategy of relative clause formation mentioned in Comrie and Kuteva (2013), i.e. non-reduction, is also attested in World Englishes, albeit only in the contact variety of South African Indian English (*Which-one I put in the jar, that-one is good.* ‘The ones [i.e., pickles] I put in the jar are the best.’; Mesthrie & Dunne 1990: 37).

The preceding paragraphs have shown that variation across World Englishes can be successfully modelled using implicational hierarchies uncovered through cross-linguistic typological work. Grammatical phenomena in standard and non-standard varieties select different values on these hierarchies and can thus be systematically related to one another. This is not to say, however, that all variation manifested by World Englishes can be modelled in this way. Interestingly, non-standard features sometimes exhibit greater harmony with cross-linguistic generalizations than the corresponding standard features. If these represent the preferred typological states, they arguably do not spread into the standard varieties due to the World System of Englishes.

#### 4. Vernacular universals and angloversals

The number of recognized and documented varieties of English has increased substantially over the past decades so that researchers have started to search for generalizations across them that in status are similar to typological universals – even though these are based on only one language (defined genetically). These are known as “vernacular universals” or “angloversals”.

The idea of vernacular universals in the sense of “a small number of phonological and grammatical processes [that] recur in vernaculars wherever they are spoken” has been primarily advocated by Chambers (2004: 128). Besides certain phonological reduction processes, Chambers (2004: 129) views the following four morphosyntactic phenomena as instances of vernacular universals.

1. conjugation regularization, or leveling of irregular verb forms: *Yesterday John seen the eclipse, Mary heard the good news;*
2. default singulars, or subject-verb nonconcord: *They was the last ones;*
3. multiple negation, or negative concord: *He didn't see nothing;*
4. copula absence, or copula deletion: *She smart, We going as soon as possible.*

These would count as absolute angloversals within the classification of universals introduced in Section 3, but more recent work has revealed that their scope is relatively narrow and practically restricted to the vernacular forms of English spoken in North America so that they may be more aptly referred to as “areoversals” (Szmrecsanyi & Kortmann 2009: 36–37). Curiously enough, though, the status of the phenomena in the 1. and 2. above is very different from that in 3. and 4. The levelling of irregular verb forms and the omission of subject-verb concord are processes of regularization (driven by economy) that can be expected to occur in (mainly spoken) non-standard varieties disconnected from the normative pressure exerted by the standard. They would also be expected in learner data. In contrast, negative concord and copula absence represent fundamental typological parameters that represent general architectural options for languages. Let us look at subject-verb agreement, negative concord, and copula absence in more detail.

There are few grammatical domains that show a similar degree of variation across World Englishes as subject-verb agreement. Non-standard agreement ranges from regular agreement to highly idiosyncratic systems with substantial internal variation. For example, the British East Anglian dialects show zero marking (no agreement), while several western and northern dialects of England use the *-s* suffix across the entire verbal paradigm (Siemund 2013: 200–201). Consider the following examples:

- (17) a. East Anglia (Trudgill 1999: 102)  
*That rain a lot there.*
- b. Reading (Edwards 1993: 223)  
*I gets out of the car and walks down the street for a few yards before I sees them boys coming towards me.*

Such regularization processes are certainly expected from a processing perspective. At the opposite end of the spectrum, we find highly idiosyncratic agreement systems of which the Northern Subject Rule perhaps furnishes the best known

example. In its most categorical form, it occurs in traditional northern English and Scottish dialects and follows the rule given in (18), illustrated in (19). What is most notable is the adjacency condition between finite verb and pronominal subjects, which is extremely rare cross-linguistically (Corbett 2006: Chapter 6).

- (18) *The Northern Subject Rule (A)*: (Pietsch 2005: 5)  
Every agreement verb takes the *-s* form, except when it is directly adjacent to one of the personal pronouns *I, we, you* or *they* as its subject.
- (19) a. Lancashire (Pietsch 2005: 90)  
*They peel 'em and boils 'em.*
- b. Yorkshire (Pietsch 2005: 90)  
*They go in and cuts em down.*

Such clear and categorical instances of the Northern Subject Rule, however, are very difficult to observe, as there is always variation. The alternative formulation in (20) captures the empirical reality more adequately.

- (20) *The Northern Subject Rule (B)*: (Pietsch 2005: 6)
- All third singular subjects (and, where preserved, the old second singular *thou*) always take verbal *-s*.
  - Type-of-Subject Constraint*: All other subjects except the personal pronouns *I, we, you, they* (and, where it exists, *youse*) take verbal *-s* variably.
  - Position-of-Subject Constraint*: Non-adjacency of subject and verb favors verbal *-s*.

According to Pietsch (2005: 6), the formulation of the Northern Subject Rule in (20) can be viewed as the result of competition between two categorical systems, namely the traditional Northern Subject Rule (18) and the agreement pattern found in standard English. Godfrey and Tagliamonte (1999: 106), in contrast, altogether conceptualize the distribution of the *-s* suffix as a statistical problem. This analysis is based on the observation that verbs can take the *-s* suffix variably in practically identical contexts, as shown in the examples in (21).

- (21) Devon English (Godfrey & Tagliamonte 1999: 89)
- Her gives me a hug and a kiss, when I comes in and one when I go.
  - People says 'yeah but look at your weather, you gets it freezing cold in the winter, you get all the rain.'
  - He comes every- three times a week he come.
  - Kiddies come over ... and they'm talking to the animals and that. And the animals looks down, you know. And there's a fantastic thing – animals and kiddies.

The observable variation is apparently driven by an intricate mix of phonological, lexical, semantic, and syntactic conditioning factors. For example, the verb *say*, habitual contexts, full noun phrases, as well as subject non-adjacency favor the appearance of the *-s* suffix. This would mean, however, that the Northern Subject Rule describes only a subset of the relevant factors. Chambers (2004: 133) even posits an implicational hierarchy of plural subject types governing the appearance of singular agreement on the verb (“default singulars”). The hierarchy is given in (22) and says that if a variety shows singular agreement with some subject type on this hierarchy, it will also show singular agreement with all subject types further to its left.

(22) there > you > we > NP<sub>PL</sub> > they

Considering the examples in (23), there is thus an increasing likelihood to encounter singular agreement from (23a) to (23e), with examples like (23e) being acceptable in the standard varieties (*There’s too many McDonald’s in Helsinki*). Having singular agreement in all the relevant contexts can be regarded as a good diagnostic of a basilect variety.

- (23) a. they: *They was all born in Georgia, mama and my daddy both.*  
 b. NP<sub>PL</sub>: *All the student teachers was comin’ out to Wellborn.*  
 c. we: *We was in an ideal place for it.*  
 d. you: *You was a majorette?*  
 e. there: *There was about twenty-somethin’ boys and just four girls.*

What usually tends to be forgotten in the discussion of subject-verb agreement is that the system found in the standard varieties is highly exceptional from a cross-linguistic point of view, too. Again, the Konstanz Universals Archive considers this a cross-linguistic *rarissimum*, as illustrated by the following description taken from the archive:

Phenomenon: verb inflection with non-zero exponent for 3rd person (subject or object agreement/cross-reference), but zero for all other persons<sup>12</sup>

It does not appear implausible to argue that such an exceptional agreement system can only survive since it is part of the grammatical inventory of the standard varieties. The more regular systems found in various dialects fail to get promoted due to the internal logic of the World System of Englishes, while highly complex systems like the Northern Subject Rule are driven out of use.

---

12. <<https://typo.uni-konstanz.de/rara/nav/search.php?PHPSESSID=fac335gq8gmk9lirs3r9l6u-jb7js3pk0>> (3 June 2017).

A similar argumentative logic can be invoked for negative concord and copula absence. The term “negative concord” describes the agreement in polarity between clausal negation and indefinite expressions, as in the example from Spanish below:

- (24) Spanish (Christoph Gabriel)  
 Nadie dice nunca nada a nadie.  
 nobody says never nothing to nobody  
 ‘Nobody says never anything to anybody.’

This pattern is quite common cross-linguistically. In a sample of 206 languages, Haspelmath (2013) identifies no fewer than 170 languages with negative concord. What is frequently denounced as incorrect non-standard usage in English, represents the majority pattern cross-linguistically. Normative pressure from above helps to explain why negative concord remains a non-standard feature, even though it clearly serves as a marker of high prestige pop culture (Mick Jagger: *I can't get no satisfaction*; Pink Floyd: *We don't need no education*).<sup>13</sup> Similarly, the expression of predicate nominals (*John is a doctor*) without copula represents a widely attested option in the languages of the world, even though it is not the majority pattern. Stassen (2013) investigated a sample of 386 languages and found 175 languages allowing zero copulas (e.g. Russian, Maori). Accordingly, English vernacular copula deletion, as for example in African American Vernacular English, represents a perfectly legitimate option that, however, remains a non-standard feature due to the hierarchical organization of the World System of Englishes (see Mair 2013: 257–265).<sup>14</sup>

Recent comparative work on World Englishes has also revealed implicational connections between non-standard features, though many of them are of the statistical type. The following examples of biconditional implicational connections (correlations) are nonetheless perfect in the sense that they do not have exceptions (see Szmrecsanyi & Kortmann 2009: 41). However, they can only be observed in specific types of varieties, like L1 varieties (traditional vernaculars), L2 varieties (many post-colonial Englishes), and English-based pidgins and creoles. Each biconditional angloversal is followed by exemplification, taken from the electronic World Atlas of Varieties of English (Kortmann & Lunkenheimer 2013).

---

13. Originally, of course, pop culture artists wanted to draw on the connotations associated with vernacular usage.

14. According to Mair (2013), the hierarchical organization manifests itself in terms of the following factors: demographic weight, institutional support, prestige, direction of translations, media presence, selection of languages for the purpose of language learning.

- (25) L1 varieties: If a variety permits non-coordinated subject pronoun forms in object function, it also permits non-coordinated object pronoun forms in subject function, and vice versa.<sup>15</sup>
- (26) a. *Us say 'er's ['she is'] dry.*  
 b. *Well, if I didn't know they, they knowed I.*
- (27) L2 varieties: If a variety permits *ain't* as the negated form of *be*, it also permits *ain't* as the negated form of *have*, and vice versa.
- (28) a. *They're all in there, ain't they?*  
 b. *I ain't had a look at them yet.*
- (29) Pidgins and creoles: If a variety has lack of inversion in main clause *yes/no*-questions, it also has lack of inversion / lack of auxiliaries in *wh*-questions, and vice versa.
- (30) a. *You get the point?*  
 b. *What he wants?*

In the same manner, the following group of examples lists three exceptionless one-way implicational angloversals, again, differentiated according to variety type (see Szmrecsanyi & Kortmann 2009: 42). Illustration is taken from Kortmann and Lunkenheimer (2013).

- (31) L1 varieties: If a variety permits *ain't* as the negated form of *be*, it also has multiple negation / negative concord.
- (32) a. *They're all in there, ain't they?*  
 b. *He won't do no harm.*
- (33) L2 varieties: If a variety has a regularized reflexives-paradigm, it also has generic *he/his* for all genders.
- (34) a. *hissself, theirselves/theirself*  
 b. *My mother, he's a primary school teacher.*
- (35) Pidgins and creoles: If a variety has invariant *don't* for all persons in the present tense, it also has multiple negation / negative concord.
- (36) a. *He don't like me.*  
 b. *He won't do no harm.*

---

15. The function of this so-called “pronoun exchange” has been a matter of much debate, but current consensus sees the effect of using subject forms in object position as that of adding emphasis. They are used as emphatic forms, so to speak. I am not aware of other languages using a similar contrast in form for a similar function.

Merely stating the above implicational universals, as is well known, is not the same as interpreting them. While one may be willing to see some functional connection between *ain't* and negative concord, this is less obvious with reflexives and generic gender, or invariant *don't* and negative concord. As far as I can see, finding functional explanations for the observed connections is one of the major residues in the study of angloversals, though some of the connections may be interpretable as diachronic extensions (see Szmrecsanyi & Kortmann 2009: 43 for some ideas). But whatever the correct interpretation, the predictive power of angloversals will never reach that of typological universals.

## 5. Explanations and motivations

The preceding sections surveyed a number of morphosyntactic phenomena from varieties of English in terms of typological data points, relating them to major typological parameters as well as typological universals. I find these perspectives very illuminating, but they clearly pass over many crucial issues necessary for their understanding, such as historical migration, language contact, second language acquisition, grammaticalization, and globalization. Some of them are related to the World System of Englishes, as discussed in Section 1. This section will try to bring in some of these additional perspectives.

### 5.1 Historical migration

Since its beginnings, English has been an export product taken to new territories and shaped by many external influences. Today's diversity of varieties of English is the product of historical migration and many similarities and differences between them can be explained in such terms.<sup>16</sup> As this represents a rather broad strand of research, I will here just mention a couple of salient facts.

The major export routes to North America, Australia, New Zealand, Africa, India, and Southeast Asia notwithstanding, it is especially the correspondence of specific features in specific locales that has attracted research. For example, many similarities between traditional Newfoundland English and Irish English as well as the dialects of Southwest England can be explained by historical migration. Similarly, the presence of African American Vernacular English in Nova Scotia, Sierra Leone, and Liberia must be attributed to migration. Scottish English and its distinct features were exported to Northern Ireland, the Appalachian Mountains,

---

16. Strevens (1980) offers a model of World Englishes based on historical migration.

pockets in the Canadian Rockies, Tasmania, and Otago. The influence of the colonists' dialects on the development of pidgins and creoles is a matter of ongoing debate. Quite typically, we are not talking about the categorical presence or absence of certain grammatical features, but quantitative differences relating to their usage and distribution (Poplack & Tagliamonte 2001).

For example, the use of *them* as a demonstrative pronoun is attested in many traditional vernaculars (Appalachian English, Ozark English, Newfoundland English, Irish English, Australian Vernacular English, etc.), and can be traced back to English dialects. Conspicuously, it is also found in African American Vernacular English and several English-based creoles. Similarly, *was/were*-generalization (*You were hungry but he were thirsty; You was hungry but he was thirsty*) is a feature of traditional vernaculars as well as some creoles that originates in dialects of England (Kortmann & Lunkenheimer 2013). Negative concord is a feature of traditional vernaculars and creoles, but very difficult to find in post-colonial Englishes (Siemund 2013: 6).

## 5.2 Globalization and social media use

There are good reasons to believe that globalization, international migration, and the extensive use of social media influence the development of World Englishes. After all, the currently observable set of World Englishes is the product of globalization processes that started a long time ago. The main difference is that these globalization processes have intensified immensely due to fast international transport and electronic communication. Notions like 'super-diversity' (Vertovec 2007) try to capture the current situation. Non-standard features of English keep travelling, but in today's world primarily through globalization and the use of social media.

Ongoing corpus-based empirical work on World Englishes is, at least in principle, able to monitor changes in real time, as it were, even though currently available corpora like the International Corpus of English or GloWbE lack a diachronic dimension (see Heller et al. 2017a, b for recent applications). The methodologies employed allow the measurement of distributional shifts and differences as well as the relevant factor weights with great precision. What they cannot determine, though, are the paths and promoters of ongoing language change since such large-scale corpora necessarily sample collective rather than individual speech. The problem of actuation and diffusion remains unresolved.

In another strand of research on World Englishes, the focus of attention is placed on individual language repertoires and speech contributions. For example, Siemund et al. (2014) and Leimgruber et al. (2018) explore the language repertoires and the language use patterns (including social media) in a sample of 450



Singaporean students drawn from three different educational institutions (university, polytechnic, Institute of Technical Education). These studies show that language repertoires correlate with educational institutions, ethnic as well as social background, though there is a limited number of dominant repertoires. They strongly suggest that these diverse language repertoires represent an important confounding factor that should be taken into consideration in corpus work on Singapore English. The focus on individual language repertoires is extended to diachronic work on Colloquial Singapore English in Siemund and Li (2017) in which the speech productions of speakers born between 1899 and 1958 were analyzed. These authors demonstrate substantial individual variation, though also surprising collective stability. Again, these differences tend to be passed over in current corpus work, but they are important to understand the local linguistic textures.

Due to globalization and extensive social media use, some countries are currently witnessing a shift in the status of English from a foreign to a second language. This shift is practically complete in Scandinavian countries and the Netherlands, but ongoing in Germany and other European countries. Exposure to, use of, and schooling in English begin very early in an individual's career, with many young individuals not only speaking German as their background language, but also one or more immigrant languages (Turkish, Polish, Russian, Farsi, etc.). This has important consequences for the acquisition and use of English, as it develops in multilingual speakers and is influenced and perhaps even fostered by the available background languages. These issues are explored in Hopp et al. (2018) for lexical and grammatical development, Lorenz (2018) for tense and aspect, and Siemund et al. (2018) for demonstrative pronouns.

### 5.3 Language contact and second language acquisition

Notwithstanding the fact standard English is a contact language par excellence that was shaped by influence from Celtic, Scandinavian, French, and several other languages, it is especially the outer circle varieties (in the sense of Kachru 1985) located in the former colonies of the British Empire that are strongly characterized by influence from the relevant indigenous languages. Nigerian English, Indian English, and Singapore English are special precisely because of heavy influence from Yoruba, Igbo, Hindi, Cantonese, Malay, and several other languages. They are contact varieties spoken in different degrees of remoteness from the (mainly) British standard. Typically, the so-called "New Englishes" peacefully coexist with the relevant local languages, forming a means for interethnic communication in territories characterized by many smaller ethnic groups and languages. In some areas, notably Singapore, we can observe massive language shifts from the so-called

“mother tongues” to English. Home language use of English in Singapore, for instance, went up from practically zero to around forty per cent in the course of a few decades (see Siemund et al. 2014). There are historical precursors to such shift varieties in Great Britain, since today’s Irish English represents the outcome of nearly an entire population changing their native tongue within a couple of centuries. This case is less salient, since it happened some centuries ago, but the underlying language learning and nativization processes can be assumed to be quite similar.

The earliest attested contact-induced second language acquisition processes in the history of English date back to the Viking period in the eighth and ninth centuries, though in research on varieties of English, it is primarily the post-colonial Englishes spoken in India, Africa, and Southeast Asia that are discussed in this context. Irish English is also strongly characterized by language contact, but the relevant contact period is less recent and practically over.

When English was taken to the former colonies in the aforementioned territories, it was primarily disseminated through the education systems and by making administration practically English only. The relevant autochthonous languages remained spoken at home, so that English largely came to be learnt as a second language, often through school teachers who themselves had learnt English as a second language. As is well known, second language acquisition is error prone and characterized by substrate influence (transfer) from previously learnt languages. Fossilization is a typical outcome. Since the number of learners of English as a second language is high in the former colonies, while the number of English mother tongue speakers is low, the resulting second language varieties came to be accepted as local norms of English. This is known as “nativization” (see Schneider 2007), which typically proceeds by embracing the new local norms as solidarity codes and imbuing them with social prestige. New varieties emerge, referred to by labels such as “Indian English”, “Colloquial Singapore English”, or “Nigerian English”.

Transfer or cross-linguistic influence from the autochthonous substrate languages can equip English with typological features that it originally did not possess. The *after*-perfect of Irish English (*I’m after having dinner*) and the *already*-perfect of Colloquial Singapore English (*I eat lunch already*) may serve to illustrate this point (see Siemund 2013). The non-reduced relative clauses found in South African Indian English, as discussed in Section 3, also belong into this category. One can clearly find many additional examples. Second language acquisition frequently leads to the simplification of grammatical features that learners find difficult to master. This explains the widespread replacement of dental fricatives by dental stops in contact Englishes. Asian Englishes typically shed the already parsimonious inflectional morphology of standard English, which can be interpreted as negative transfer in second language acquisition.

## 5.4 Grammaticalization

Post-colonial Englishes – and of course also pidgin and creole Englishes – furnish interesting cases of grammaticalization, or, at least, incipient grammaticalization. I here mention grammaticalization as a separate point, because it systematically enriches the inventory of grammatical markers and constructions in these varieties. Some cases of cross-linguistic influence may be reconstructable in terms of contact-induced grammaticalization in the sense of Heine and Kuteva (2003).

In Colloquial Singapore English, definite and indefinite articles are frequently omitted (analyzable as negative transfer from Chinese), but crucially, demonstrative pronouns can be found in the function of definite articles. This phenomenon is quite typical of Englishes characterized by intense language contact and may be viewed as incipient grammaticalization (Aboriginal English: *That door bin close* ‘the door closed’; Malcolm 2008: 431). In a similar way, the numeral *one* is used in the function of an indefinite article (Aboriginal English: *They seen one [‘a’] green snake tangled round a tree*; Malcolm 2008: 431). Here, varieties of English replicate well-known paths of grammaticalization (see Heine and Kuteva 2002 for an extensive collection), although the distinction between internal grammaticalization and contact-induced grammaticalization needs to be decided on a case-by-case basis.

Numerous additional examples can be found, just to be named and briefly illustrated in the following: body part reflexives in Nigerian Pidgin (*Di man bit im bodi* ‘The man beat himself’; Faraclas 1996: 103), *already* as a perfect marker in Cape Flats English (*Were you there already?* ‘Have you been there before?’; McCormick 2008: 523), *go* as a future tense marker in Gullah (*Uh ain ga go nowhere* ‘I won’t go anywhere’; Mufwene 2008: 563), locative expressions (*stap* < *stop*, *stay*) encoding progressive aspect in Tok Pisin (*Ol i wokabout i stap* ‘they are walking’; Smith 2008: 500). The nominalizer cum relative marker *one* in Colloquial Singapore English (calqued on Chinese 的 *de*), as in *The cake John always buy one very nice* ‘The cake that John always buys is very nice’ (Alsagoff & Lick 1998: 128), may furnish a case of contact-induced grammaticalization, as it is even used as a discourse marker: *I always use microwave one* ‘I ALWAYS use a microwave!’ (Bao 2009: 340).<sup>17</sup>

Standard and non-standard varieties of English show fundamental differences in the grammaticalization of negation. Whereas the standard varieties are

---

17. The distinction between contact-induced grammaticalization (Heine & Kuteva 2003) and systemic transfer (Bao 2005) is a difficult one. The matter is further complicated by the fact that some outcomes of these processes are already attested in earlier forms of English (see Ziegeler 2014 on replica grammaticalization). As far as I can see, the resulting empirical and conceptual intricacies easily furnish a new research project.

formally completely symmetric in the encoding of positive and negative clauses, non-standard varieties typically offer asymmetric paradigms. Consider the data in Table 2.

**Table 2.** Asymmetrical paradigms (adapted from Anderwald 2002: 199)

Positive	<i>am</i>	<i>is</i>	<i>are</i>	<i>has</i>	<i>have</i>	<i>do</i>	<i>does</i>	<i>was</i>	<i>were</i>
Negative	<i>ain't</i>					<i>don't</i>		<i>wasn't/weren't</i>	

It is obvious from Table 2 that the negative paradigm has fewer forms than the positive one. The invariant form *ain't* is only one example of paradigm simplification under negation. Others include what is known as “third person singular *don't*” where positive *do* and *does* fall together in negative *don't* and the leveling of negative past tense *be* to either *wasn't* or *weren't* (Anderwald 2002: 198–201). Miestamo (2000: 78) argues that such asymmetric negative paradigms are in fact functionally symmetric, because non-occurring situations are typically less precise with respect to basic parameters such as *who*, *when*, *where*, and *why*: “In the asymmetric paradigms there is a ‘vertical’ analogy (or iconicity): the ontology of non-facts is less differentiated than the ontology of facts, and linguistic structure reflects this distinction.” The suspicion that here arises is that the standard English symmetric system of negation is a product of standardization processes (“more logical”) and kept in place as a result of the hierarchical organization of the World System of Englishes.

## 6. Summary and conclusion

By now, World Englishes have been approached from a multitude of perspectives. While issues of historical development and colonial expansion figured prominently in the earlier proposals, more recent approaches focus on the nativization, appropriation, and diversification of English in new areas as well as matters of contact-induced change due to interaction with local languages and its use by multilingual speakers. In the present contribution, I explored the interaction of typological constraints and the social hierarchies behind the World System of Englishes, arguing that normative pressure from the hyper-central and super-central varieties may easily override otherwise pervasive constraints on the architecture of language and lead to the preservation of grammatical features that appear quite exceptional from a cross-linguistic perspective. Languages are always in the process of change and at each point in their development, they are bound to generate idiosyncrasies due to internal change or external forces. Normative pressure tends to preserve such

idiosyncrasies, though global migration, multilingualism, and the extensive use of social media have created a strong undercurrent that works against normative pressure from above and brings non-standard features to the fore. Crucially, these often turn out to be in greater harmony with cross-linguistic generalizations than the corresponding standard features.

Given the scope of the project of modelling World Englishes from cross-linguistic perspective, there necessarily remain many loose ends that offer highly promising avenues for future research. I here restrict myself to mentioning three such areas: First, the status of angloversals and vernacular universals in relation to typological universals remains somewhat unclear, requiring more theoretical work especially concerning their functional basis. Second, the tension between contact-induced grammaticalization, replica grammaticalization, and systemic transfer practically represents uncharted territory asking for more attention. And third, we need to explore in more detail the relationship between World Englishes and multilingual development, including the use of English as a Lingua Franca by multilingual speakers. Based on our current assumptions one could, for instance, hypothesize that multilingual speakers – in comparison to monolingual speakers – are the “better” English as a Lingua Franca speakers, but this idea remains to be explored in future work.

## Acknowledgement

I would like to thank the reviewers and editors for their constructive feedback on earlier versions of this paper. All remaining errors and inconsistencies are my own.

## References

- Alsagoff, L. & Lick, Ho Chee. 1998. The relative clause in colloquial Singapore English. *World Englishes* 17(2): 127–138. <https://doi.org/10.1111/1467-971X.00087>
- Anderwald, L. 2002. *Negation in Non-Standard British English: Gaps, Regularizations and Asymmetries*. London: Routledge.
- Bao, Z. 2005. The aspectual system of Singapore English and the systemic substratist explanation. *Journal of Linguistics* 41(2): 237–267. <https://doi.org/10.1017/S0022226705003269>
- Bao, Z. 2009. *One* in Singapore English. *Studies in Language* 33: 338–65. <https://doi.org/10.1075/sl.33.2.05bao>
- Burridge, K. & Kortmann, K. (eds). 2009. *Varieties of English*, Vol. 3: The Pacific and Australasia. Berlin: Mouton de Gruyter.
- Chambers, J. K. 2004. Dynamic typology and vernacular universals. In *Dialectology Meets Typology. Dialect Grammar from a Cross-Linguistic Perspective*, B. Kortmann (ed.), 127–145. Berlin: Mouton de Gruyter.

- Christian, D. 1991. The personal dative in Appalachian speech. In *Dialects of English: Studies in Grammatical Variation*, P. Trudgill & J. K. Chambers (eds), 11–17. London: Longman.
- Clarke, S. 2004. Newfoundland English: Morphology and syntax. In Kortmann, Burridge, Mesthrie, Schneider, & Upton (eds), 303–318.
- Comrie, B. & Kuteva T. 2013. Relativization on subjects. In Dryer & Haspelmath (eds), Chapter 122. <<http://wals.info/chapter/122>>
- Corbett, G. 2000. *Number*. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139164344>
- Corbett, G. 2006. *Agreement*. Cambridge: CUP.
- Croft, W. 2004. *Typology and Universals*. Cambridge: CUP.
- Davydova, J., Hilbert, M., Pietsch L. & Siemund, P. 2011. Comparing varieties of English: Problems and perspectives. In *Linguistic Universals and Language Variation*, P. Siemund (ed.), 291–323. Berlin: Mouton de Gruyter.
- Deutscher Sprachatlas. 1927–1956. *Based on Georg Wenker's Sprachatlas des deutschen Reichs, started by Ferdinand Wrede and continued by Walther Mitzka and Bernhard Martin*. Marburg: Elwert.
- Dryer M. & Haspelmath, M. (eds). 2013. *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. <<http://wals.info/>>
- Edwards, V. 1993. The grammar of southern British English. In *Real English: The Grammar of English Dialects in the British Isles*, J. Milroy & L. Milroy (eds), 214–242. London: Longman.
- Evans, N. & Levinson, S. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5): 429–492. <https://doi.org/10.1017/S0140525X0999094X>
- Faltz, L. M. 1985. *Reflexivization: A Study in Universal Syntax*. New York NY: Garland.
- Faraclas, N. 1996. *Nigerian Pidgin*. London: Routledge. <https://doi.org/10.4324/9780203192801>
- Filppula, M. 1999. *A Grammar of Irish English: Language in Hibernian Style*. London: Routledge.
- Filppula, M. 2004. Irish English: Morphology and syntax. In Kortmann, Burridge, Mesthrie, Schneider & Upton (eds), 73–101.
- Filppula, M., Klemola, J. & Paulasto, H. (eds). 2009. *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*. London: Routledge.
- Gilliéron, J. & Edmont, E. 1902–1914. *Atlas linguistique de la France*, 10 Vols. Paris: Ed. du Cths.
- Godfrey, E. & Tagliamonte, S. 1999. Another piece of the verbal -s story: Evidence from Devon in Southwest England. *Language Variation and Change* 11(1): 87–121. <https://doi.org/10.1017/S0954394599111050>
- Haspelmath, M. 1990. The grammaticization of passive morphology. *Studies in Language* 14: 25–72. <https://doi.org/10.1075/sl.14.1.03has>
- Haspelmath, M. 2013. Negative indefinite pronouns and predicate negation. In Dryer & Haspelmath (eds), Chapter 115. <<http://wals.info/chapter/115>>
- Heine, B. & Kuteva, T. 2002. *World Lexicon of Grammaticalization*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511613463>
- Heine, B. & Kuteva, T. 2003. On contact-induced grammaticalization. *Studies in Language* 27: 529–572. <https://doi.org/10.1075/sl.27.3.04hei>
- Heller, B., Bernaisch, T. & Gries, S. T. 2017a. Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes. *ICAME Journal* 41(1): 33–66. <https://doi.org/10.1515/icame-2017-0005>
- Heller, B., Szmrecsanyi, B. & Grafmiller, J. 2017b. Stability and fluidity in syntactic variation World-Wide: The genitive alternation across varieties of English. *Journal of English Linguistics* 45(1): 3–27. <https://doi.org/10.1177/0075424216685405>

- Herrmann, T. 2003. Relative Clauses in Dialects of English: A Typological Approach. PhD dissertation, University of Freiburg.
- Hopp, H., Kieseier, T., Vogelbacher, M. & Thoma, D. 2018. L1 effects in the early L3 acquisition of vocabulary and grammar. In *Foreign Language Education in Multilingual Classrooms* [Hamburg Studies on Linguistic Diversity 7], A. Bonnet & P. Siemund (eds). Amsterdam: John Benjamins.
- von Humboldt, W. 1836. *Über die Verschiedenheit des menschlichen Sprachbaus und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Königliche Akademie der Wissenschaften.
- Kachru, B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the World. Teaching and Learning the Language and Literatures*, R. Quirk & H. Widdowson (eds), 11–30. Cambridge: CUP and the British Council.
- Keenan, E. & Comrie, B. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8: 63–99.
- Kemmer, S. 1993. *The Middle Voice* [Typological Studies in Language 23]. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.23>
- Kortmann, B. 2004. Synopsis: Morphological and syntactic variation in the British Isles. In Kortmann, Burridge, Mesthrie, Schneider, & Upton (eds), 1089–1103.
- Kortmann, B., Burridge, K., Mesthrie, R., Schneider, E., Upton, C. (eds). 2004. *A Handbook of Varieties of English*, Vol. 2: Morphology and Syntax. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110197181>
- Kortmann, B. & Lunkenheimer, K. (eds). 2013. *The Electronic World Atlas of Varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://ewave-atlas.org>>
- Leimgruber, J., Siemund, P. & Terassa, L. 2018. Singaporean students' language repertoires and attitudes revisited. *World Englishes*.
- Lorenz, E. 2018. "One day a father and his son going fishing on the Lake." – A study on the use of the progressive aspect of monolingual and bilingual learners of English. In *Foreign Language Education in Multilingual Classrooms* [Hamburg Studies on Linguistic Diversity 7], A. Bonnet & P. Siemund (eds), Amsterdam: John Benjamins.
- Mair, C. 2013. The World System of Englishes. Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34(3): 253–278. <https://doi.org/10.1075/eww.34.3.01mai>
- Malcolm, I. G. 2008. Australian Creoles and Aboriginal English: Morphology and syntax. In Burridge & Kortmann (eds), 415–443.
- Mathiot, M. (assisted by M. Roberts). 1979. Sex roles as revealed through referential gender in American English. In *Ethnolinguistics: Boas, Sapir and Whorf Revisited*, M. Mathiot (ed.), 1–47. The Hague: Mouton.
- McCormick, K. 2004. Cape Flats English: Morphology and syntax. In Kortmann, Burridge, Mesthrie, Schneider, & Upton (eds), 993–1005.
- McCormick, K. 2008. Cape Flats English: Morphology and Syntax. In *Varieties of English*, Vol. 4: Africa, South and Southeast Asia, R. Mesthrie (ed.), 521–534. Berlin: Mouton de Gruyter.
- Melchers, G. 2004. English spoken in Orkney and Shetland: Morphology, syntax and lexicon. In Kortmann, Burridge, Mesthrie, Schneider, & Upton (eds), 34–46.
- Mesthrie, R. & Dunne, T. T. 1990. Syntactic variation in language shift: The relative clause in South African Indian English. *Language Variation and Change* 2(1): 31–56. <https://doi.org/10.1017/S0954394500000259>



- Miestamo, M. 2000. Towards a typology of standard negation. *Nordic Journal of Linguistics* 23(1): 65–88. <https://doi.org/10.1080/033258600750045787>
- Miestamo, M. 2007. Negation – an overview of typological research. *Language and Linguistics Compass* 1(5): 552–570. <https://doi.org/10.1111/j.1749-818X.2007.00026.x>
- Miller, J. 2004. Scottish English: Morphology and syntax. In Kortmann, Burridge, Mesthrie, Schneider, & Upton (eds), 47–72.
- Montgomery, M. 2004. Appalachian English: Morphology and syntax. In Kortmann, Burridge, Mesthrie, Schneider, & Upton (eds), 245–280.
- Mufwene, S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511612862>
- Mufwene, S. 2008. Gullah: Morphology and syntax. In *Varieties of English*, Vol. 2: The Americas and the Caribbean, E. W. Schneider (ed.), 551–571. Berlin: Mouton de Gruyter.
- Onysko, A. 2016. Modeling world Englishes from the perspective of language contact. *World Englishes* 35(2): 196–220. <https://doi.org/10.1111/weng.12191>
- Orton, H. & Dieth, E. (eds). 1962–1971. *Survey of English Dialects*, 13 Vols. Leeds: E.J. Arnold & Son.
- Pawley, A. 2004. *Australian vernacular English: Some grammatical characteristics*. In Kortmann, Burridge, Mesthrie, Schneider, & Upton (eds), 611–642.
- Peters, P. 2009. Australian English as a regional epicenter. In *World Englishes – Problems, Properties and Prospects* [Varieties of English Around the World G40], T. Hoffmann & L. Siebers (eds), 107–124. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g40.09pet>
- Pietsch, L. 2005. *Variable Grammars: Verbal Agreement in Northern Dialects of English*. Tübingen: Niemeyer. <https://doi.org/10.1515/9783110944556>
- Poplack, S. & Tagliamonte, S. 2001. *African American English in the Diaspora*. Oxford: Blackwell.
- Sasse, H-J. 1993. Syntactic categories and subcategories. In *Syntax. Ein internationales Handbuch zeitgenössischer Forschung / An International Handbook of Contemporary Research*. [Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication 9], J. Jacobs, A. von Stechow, W. Sternefeld, & T. Vennemann (eds), 646–686. Berlin: Mouton de Gruyter.
- von Schlegel, A. W. 1818. *Observations sur la langue et la littérature provençales*. Paris: Librairie grecque-latine-allemande.
- Schneider, E. W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511618901>
- Siemund, P. 2004. English. In *Variationstypologie*, T. Roelcke (ed.), 1–29. Berlin: Walter de Gruyter.
- Siemund, P. 2008. *Pronominal Gender in English: A Study of English Varieties from a Cross-Linguistic Perspective*. London: Routledge.
- Siemund, P. 2009. Linguistic universals and vernacular data. In *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*, M. Filppula, J. Klemola, & H. Paulasto (eds), 321–346. London: Routledge.
- Siemund, P. 2010. Grammaticalization, lexicalization and intensification. English *itself* as a marker of middle situation types. *Linguistics* 48(4): 797–836. <https://doi.org/10.1515/ling.2010.025>
- Siemund, P. 2013. *Varieties of English. A Typological Approach*. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139028240>
- Siemund, P. 2014. The emergence of English reflexive verbs: An analysis based on the Oxford English Dictionary. *English Language and Linguistics* 18(1): 49–73. <https://doi.org/10.1017/S1360674313000270>



- Siemund, P., Schulz, M. E. & Schweinberger, M. 2014. Studying the linguistic ecology of Singapore: A comparison of college and university students. *World Englishes* 33(3): 340–362. <https://doi.org/10.1111/weng.12094>
- Siemund, P. & Li, L. 2017. Towards a diachronic reconstruction of Colloquial Singapore English. In *Negation and Contact. With Special Focus on Singapore English* [Studies in Language Companion Series 183], D. Ziegeler & B. Zhiming (eds), 11–32. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.183.02sie>
- Siemund, P., Schröter, S. & Rahbari, S. 2018. Learning English demonstrative pronouns on bilingual substrate: Evidence from German heritage speakers of Russian, Turkish, and Vietnamese. In *Foreign Language Education in Multilingual Classrooms* [Hamburg Studies on Linguistic Diversity 7], A. Bonnet & P. Siemund (eds). Amsterdam: John Benjamins.
- Silverstein, M. 1976. Hierarchy of features and ergativity. In *Grammatical Categories in Australian Languages* [Linguistic Series 22], R. M. W. Dixon (ed.), 112–171. Canberra: Australian Institute of Aboriginal Studies.
- Smith, G. 2008. Tok Pisin: Morphology and Syntax. In Burrige & Kortmann (eds), 488–513.
- Stassen, L. 2013. Zero copula for predicate nominals. In Dryer & Haspelmath (eds), Chapter 120. <<http://wals.info/chapter/120>>
- Stevens, P. 1980. *Teaching English as an International Language*. Oxford: Pergamon Press.
- de Swaan, A. 2001. *The World Language System. A Political Sociology and Political Economy of Language*. Cambridge: Polity.
- Szmrecsanyi, B. & Kortmann, B. 2009. Vernacular universals and angloversals in a typological perspective. In *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*, M. Filppula, J. Klemola, & H. Paulasto (eds), 33–53. London: Routledge.
- Tagliamonte, S., Jennifer S. & Lawrence, H. 2005. No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change* 17(1): 75–112.
- Trudgill, P. 1999. *The Dialects of England*. Oxford: Blackwell.
- Trudgill, P. 2011. *Sociolinguistic Typology. Social Determinants of Linguistic Complexity*. Oxford: OUP.
- Vertovec, S. 2007. Super-diversity and its implications. *Ethnic and Racial Studies* 30: 1024–54. <https://doi.org/10.1080/01419870701599465>
- Wagner, S. 2004. English dialects in the Southwest: Morphology and syntax. In Kortmann, Burrige, Mesthrie, Schneider, & Upton (eds), 154–174.
- Wright, J. 1898–1905. *The English Dialect Dictionary*, 6 Vols. Oxford: Henry Frowde.
- Ziegeler, D. 2014. Replica grammaticalization as recapitulation: The other side of contact. *Diachronica* 31(1): 106–141. <https://doi.org/10.1075/dia.31.1.04zie>
- Zupitza, J. (ed.). 1966. *Aelfrics Grammatik und Glossar*. Berlin: Weidmannsche Verlagsbuchhandlung.

# “I’m an Anglophile, but ...”

## A corpus-assisted discourse study of language ideologies in the Netherlands

Alison Edwards

Leiden University Centre for Linguistics

This study, at the interface of language ideologies, corpus linguistics and discourse analysis, explores perceptions of English in the Netherlands through recurrent discourse patterns in a corpus of language-attitudinal commentary by 724 Dutch informants. The informants position English as a purely utilitarian tool for international communication. Yet, a key ideological narrative revolves around the perceived “unnecessary” use and “overuse” of English *within* Dutch society to appear cosmopolitan, clever or “cool”. This suggests many people are (or are believed to be) mobilising the language for local interpersonal relations and identity construction. These functions go beyond mere instrumentalism, suggesting English may be being used as not just a “foreign”/“international” language but rather an additional local language for creative self-expression and identity performance.

**Keywords:** language ideologies, corpus-assisted discourse study, English as a foreign language, English as a second language, Netherlands

### 1. Introduction

As English proliferates around the world, so too do the theoretical models seeking to account for its spread and its variegated manifestations, such as Kachru’s (1985) Three Circles model and Schneider’s (2007) Dynamic Model, to name just two of the best known. Such models have been interrogated and evaluated in many studies on morphosyntactic, lexical and pragmatic variation in different English varieties worldwide. It has been suggested, however, that subjecting such macro-level models to ever more fine-grained analyses of linguistic structure can present problems (Hundt & Mukherjee 2011: 213) and that they may therefore be better suited to the study of sociocultural phenomena such as attitudes, identity and ideology. In

this context, this study seeks to interpret language ideologies in the Netherlands, a country in which English has become increasingly entrenched since the second half of the 20th century, through the lens of a relatively new model, Schneider's (2014) Transnational Attraction (TA) framework.

The study is situated at the interface between language ideologies, corpus research and discourse analysis. Studies of language ideology aim to identify explicit and implicit beliefs about language(s) in society (Silverstein 1979; Woolard 1998). Responding to recent calls to apply corpus-linguistic methods to language-ideological analyses to improve their systematicity (Moschonas & Spitzmüller 2010; Vessey 2015), I adopt a corpus-assisted discourse studies (CADS) (Partington 2004) approach, which brings together corpus linguistics and discourse analysis. Data are drawn from a small-scale corpus of open responses to a questionnaire on attitudes to English in the Netherlands (Edwards 2016).

The findings are analysed in the context of Schneider's (2014) TA framework, which sees the use of English in countries such as the Netherlands as an economic resource "driven predominantly by utilitarian considerations". He hints, however, that the drivers behind English usage may eventually come to "transcend such economic motivations" (Schneider 2014: 28). This is confirmed by the results. The informants construct English as a language that – unlike Dutch – allows them to connect with the wider world; a utilitarian stance that echoes "official" and canonical academic discourses on Expanding Circle societies in the World Englishes (WEs) paradigm. However, a key ideological narrative revolves around the perception of constant "unnecessary" use and "overuse" of English *within* Dutch society. This suggests that many people are (or are believed to be) mobilising the language for the purposes of local interpersonal relations and identity construction; functions that go well beyond mere instrumentalism and suggest English may be being used not just as a "foreign" or "international" language but rather as an additional local language for creative self-expression and identity performance.

I begin by outlining Schneider's (2014) TA framework, followed by a brief overview of the role and status of English in the Netherlands today. Next, I contextualise the study against the backdrop of existing research using the CADS and language ideologies approaches. Following a description of the corpus data and methods, the results are presented by way of a largely qualitative analysis of ideologies implicitly and explicitly expressed in discourses surrounding the node terms *English* and *the Dutch/Dutch people*. Finally, I interpret the findings in light of the TA model and connect them to broader sociolinguistic/sociocultural trends in the study of English as a global language.

## 2. Transnational attraction

Braj Kachru’s (1985) Three Circles model divided countries into Inner, Outer and Expanding Circles, which are conventionally mapped in WEs onto the tripartite classification of English as a native (ENL), second (ESL) and foreign (EFL) language. Both the Inner and Outer Circles are home to a range of postcolonial Englishes, posited in Schneider’s (2007) Dynamic Model of the Evolution of Postcolonial Englishes as emerging from a uniform process: as the social distance between the colonisers and the colonised gradually decreases, a new nation is formed and with it a new language variety (e.g. Australian English, Singaporean English). Although widely adopted and applied, the Dynamic Model has been criticised for neglecting “the ways in which the language exists in other parts of the world (i.e. the Expanding Circle)” (Seargeant 2012: 155). English in Expanding Circle countries is traditionally seen as being restricted to certain higher education and scientific fields and having a limited role in public and personal life. Yet as Kachru and Nelson (2006: 28) write, “[t]his, however, may be changing”.

With the language continually spreading to new contexts in today’s globalised, connected world, Schneider (2014) revisited the Dynamic Model to investigate whether the mechanisms posited as underlying the evolution of English in postcolonial contexts also accounted for the processes in the Expanding Circle. Despite “some evolutionary similarities (such as the strong demand for and the increasing nativization of English)”, he concluded that the Dynamic Model is “not really, or only to a rather limited extent, a suitable framework” (Schneider 2014: 28) for exploring the spread of English in non-postcolonial contexts – settings where English is becoming entrenched due to global economic (rather than postcolonial) forces.

To account for such contexts, he proposed Transnational Attraction, a conceptual framework that, unlike the Dynamic Model, taps into the recent social-scientific turn towards processes that transcend national borders and takes account of global flows and forces, resulting in “the appropriation of (components of) English(es) for whatever communicative purposes at hand, unbounded by distinctions of norms, nations or varieties” (Schneider 2014: 28). In his view, and significantly for the present paper, “[t]his process is driven predominantly by utilitarian considerations, that is, with users viewing ‘English as an economic resource’ (Kachru 2005: 91), a symbol of modernity and a stepping stone toward prosperity.”

In other words, TA emphasises the *instrumental* motivations behind the uptake and use of English. In this way, it dovetails at least to some degree with traditional discourses in WEs scholarship that construct the Expanding Circle as a space where English serves as a “mere” tool, an economic commodity that allows users to connect with the global labour market and global cultural phenomena. However, recent research with a more sociocultural and ethnographic orientation emphasises the

ways in which English as a global linguistic resource is made local (Pennycook 2010). In the area of linguistic landscapes, for example, Laitinen (2015) points out that interpreting the use of English in public signage in global contexts as intended either purely for vehicular communication or as a marker of internationalism is too simplistic. Instead, such language is reterritorialized as situated social practice that serves “distinctively local needs” (Higgins 2009: 18), creating new semiotic opportunities for social actors to perform identity and strategically construct the local (see further Edwards & Seargeant 2017).

Clearly, this position runs counter to essentialised discourses about English as either simply a foreign language or a tool for international communication – something Schneider (2014: 28) himself acknowledges: “Importantly, however, there are also signs that English will be able to *transcend such economic motivations*” (italics added); instead it may “become a multicultural resource” with “new roles”. It is this growing cognizance of English coming to serve in new contexts as a locally embedded resource for meaning-making and identity construction that I wish to explore.

### 3. English in the Netherlands

The Netherlands presents an interesting case study of the changing situation in the Expanding Circle as it has, along with other western and in particular northern European countries, seen a vast increase in the presence of English in recent decades. Dutch is the primary official language, with Frisian officially recognised as a regional language. English, however, is frequently described as transitioning from EFL to ESL status, which would see its traditional role as facilitating communication *internationally* being supplemented by an integral role *intranationally*.

Historically one of the most open economies in the world and a small country located at the crossroads of three major language areas (German, French and English) (Ammon & McConnell 2002), the Netherlands has a long tradition of foreign language learning. In the second half of the 20th century English increased in popularity as “the language of the [WWII] liberators” (Ridder 1995: 44) and of Anglo-American pop culture. Today it is indisputably the first foreign language in education, with some 25% of students at the highest secondary-school level (VWO, ≈ grammar schools) following bilingual Dutch-English streams (Dronkers 2013). Further, most Dutch universities now bill themselves as officially bilingual in order to attract foreign students and connect with global academia. As a result the Netherlands has developed into an English education “destination”, offering the highest absolute number of English-medium programmes in continental Europe (Maiworm & Wächter 2014). Films and television programmes are subtitled rather than dubbed, bilingual wordplay is common in advertising and English quotes and

passages in the print media are often left untranslated. According to Nortier (2011), virtually all Dutch citizens under the age of 50 have some mastery of English; this was confirmed by the latest Eurobarometer survey, in which 90% of the population reported being able to hold a conversation in English (European Commission 2012).

All this calls into question the traditional EFL status of the Netherlands. It has been suggested that English is coming to play an essential part in people’s linguistic repertoire even when not strictly necessary to accommodate foreigners; consider its reported use in professional and academic conferences despite all attendees understanding Dutch (Nortier 2011). At the same time, educational curricula remain largely exonormative and Anglocentric. British English is the most popular target model, with young people also open to American English, while ‘Dutch English’ rarely serves as a target model (Edwards 2016; van der Haagen 1998). Public discourses, such as those expressed in newspaper opinion pieces, tend to reflect traditional standard language ideologies and a fairly puristic stance. Early purists were concerned about the influence of Latin and French; later, the popular Society of our Language (Genootschap Onze Taal) was founded in part to combat the perceived infiltration of German. The society now espouses a more balanced outlook, much to the chagrin of those who wish to hold back the “flood” of English (van Oostendorp 2012) in terms of both loanwords into Dutch and encroachment into Dutch-language domains. This led to the founding of two splinter groups, the Foundation for the Defence of Dutch (Stichting Taalverdediging) and the Dutch Language Foundation (Stichting Nederlands). One of the latter’s key initiatives has been to propose Dutch alternatives for English loanwords, with varying success; as Smaakman (2006: 43) writes, “[t]he Dutch do not seem entirely dedicated to this fight”. Still, there remains a sense that foreign words should only be used for “good reason”, such as the absence of a Dutch equivalent (Bakker 1987). English loanwords indeed appear to be more successful if they have some instrumental advantage, such as being easy to pronounce or shorter than the Dutch equivalent (Zenner, Speelman & Geeraerts 2012). This may be changing, however, with young people found to use English words for non-instrumental, emotive reasons relatively more often than older people (e.g. because it “sounds better” or to signal group membership, Edwards 2016).

Further, the Dutch are frequently depicted in public discourse as having little pride in their own language, viewing it as a small, not particularly useful language (and one that is readily given up in an emigrant situation; de Bot & Clyne 1994). Such discourse has been challenged by some academics, who found no empirical support for this apparent lack of regard for Dutch (de Bot & Weltens 1997; Edwards 2016). This raises the question as to what extent the views expressed in the media are actually shared by people on the ground, referred to as the “spokesman problem” (de Bot & Weltens 1997: 145; van Meurs, Korzilius, Planken & Fairley 2007: 202). By

investigating public attitudes by means of a large-scale study involving all segments of the population, I aim to explore questions such as: is English viewed as a threat or an enrichment? What do people see as its “proper” place in society? Is it considered “foreign”, as traditionally conceived in the academic canon, or has it also come to be viewed as an inextricable part of the local “language world” (Bolton 2013)?

#### 4. Language ideologies and CADS

Exploring language ideologies in the Netherlands would seem to be a promising way to approach this question. Language ideologies have been defined as “sets of beliefs about language articulated by users as a rationalization or justification of perceived language structure and use” (Silverstein 1979: 193). As “shared bodies of commonsense notions” (Rumsey 1990: 346) presented as self-evident truth, they can reveal established ideas about language(s): about the perceived legitimacy of standard languages, about “good” and “bad” language use, the notion of languages as discrete entities, the ideal of the mother tongue speaker as the repository of authenticity, and the (de)legitimation of certain speakers and social groups (Woolard 1998: 16–17). This study seeks to identify people’s beliefs about the “proper” role of English in Dutch society: whether it is viewed as “merely” a foreign/international language (EFL/EIL) for use with outsiders, or rather as a second/additional language (ESL/EAL) that plays an increasingly essential role as a local indexical resource.

A fruitful approach to studying language ideologies that has emerged over the last decade or so is that of corpus-assisted discourse studies (CADS) (Baker et al. 2008; Partington 2004). In its broadest sense, this approach views discourse as language-in-use (Blommaert 2005: 2) and discourse analysis as “the study of the way language is used in a variety of sociocultural contexts” (Kirkpatrick & McLellan 2012: 654). According to its proponents, a type of discourse analysis involving corpus-linguistic techniques can help to avoid some of the pitfalls of traditional discourse analysis:

The principles of representativeness, sampling and balance which underlie corpus building help to guard against cherry-picking, while corpus-driven techniques like keywords help us to avoid over-focussing on atypical aspects of our texts. Corpus techniques can thus reassure readers that our analysts are actually presenting a systematic analysis, rather than writing a covert polemic.

(Baker & McEnery 2015: 4)

At the same time, discourse analysis can benefit corpus linguistics by bringing the sociocultural context to the fore, emphasising that language in use is always situated and reflects the behaviour and attitudes of social actors. Hence the recent



calls for more research using the CADS approach to explore language ideologies (Moschonas & Spitzmüller 2010; Vessey 2015)

In fact, there is already a burgeoning collection of research applying CADS techniques to language ideologies. To date, such research has typically focused on political and press registers (Baker & McEnery 2015: 7). For example, Fitzsimmons-Doolan (2014) uses a corpus of language policy texts to explore the ideological stance of the US Department of Education. Earlier, Blommaert and Verschueren (1998) used a corpus of newspaper reports from different European countries to shed light on the role of language in European nationalist ideologies. Newspaper texts have also been used to identify language ideologies such as Anglonormativity and hyperlinguistic awareness in the Canadian press (Freaker 2012) and to explore shared beliefs about English in Norway (e.g. “English is a much richer language than Norwegian”, “English takes over and ruins other languages”, Graedler 2014).

As a newspaper’s success depends on its journalists being able to reflect readers’ views and beliefs, the analysis of press registers allows us to “access how events and entities [e.g. languages] are conceptualised by particular groups” (Partington, Duguid & Taylor 2013: 135). Others have sought to access language ideologies as directly expressed by members of the general public or specific populations. For example, Velasquez (2013) uses a spoken corpus of interviews with Spanish-English bilinguals in El Paso, Texas, to investigate how individual discourse maps onto broader ideologies about language maintenance and loss. In Bleichenbacher (2006), written comments posted in an online tourist guestbook of the city of Kosice, Slovakia, were analysed to reveal a predominantly “monolingual mindset” among the commenters.

Like that in Bleichenbacher (2006), the present analysis is based on a small-scale comments corpus, whereby written commentary is taken to be a discursive site for the (re)production of language ideologies. The next section turns to the collection and composition of this corpus.

## 5. Data and methods

### 5.1 Data collection

The data used in this study were gathered by means of a questionnaire among Dutch informants on the role of and attitudes towards English in the Netherlands (Edwards 2016). The personal questionnaire collected information and responses (on a four-point scale) to attitudinal statements in the areas of learning and using English, perceived competence, the respective status of English and Dutch, and models and varieties of English. In a final, optional question, respondents were asked “Do you have any comments you would like to make about English in the



Netherlands?” It is the responses to this open-ended question that were compiled into the “comments corpus” used in this study.

Both the questionnaire and informants’ responses were written in Dutch (although see discussion). The corpus analyses used the original Dutch-language responses; those concordances reproduced here for illustration purposes were translated into English by the author (with the idiosyncrasies of the original comments left intact where possible).

The questionnaire was disseminated online via Google Forms using a snowball sampling procedure. A range of online networks were targeted, including social media platforms, language organisations, student associations and senior citizens’ groups. The questionnaire was live for approximately six months in 2013.

## 5.2 Corpus

A total of 2,086 questionnaire responses were received. Of these, 724 (35%) included a substantive response to the open question that was included in the analysis. Non-substantive responses, such as “Good luck with your study”, were excluded.

There were some sociodemographic differences between respondents who elected to leave a comment and those who did not. First, the mean age of commenters was markedly higher (49.3) than respondents who did not leave a comment (41.7) (one-way ANOVA  $F(1, 2079) = 51.74, p < .001$ ). Half of the comments were made by male and half by female informants, but as more women than men filled in the questionnaire (55.8% vs 43.8%), men were proportionally more likely to leave a comment (45.0% of women who filled in the questionnaire left a comment, compared to 65.3% of men) ( $\chi^2 = 16.07, df = 1, p < .001$ ). The difference in self-reported proficiency rates of people who left a comment and those who did not was small but statistically significant (3.5 vs 3.6 on a 4-point scale, where 1 = *no English at all* and 4 = *fluent*) ( $F(1, 1910) = 4.15, p = .04$ ). Finally, commenters held slightly – but again, significantly – less positive attitudes towards English than non-commenters (2.8 vs 3.0, where 1 = the least and 4 = the most positive attitudes) ( $F(1, 1916) = 5.92, p = .02$ ).<sup>1</sup>

As computer-mediated communication, the responses can be considered instances of rather informal writing. They ranged in length from a single word (*Status!*) to 341 words ( $M = 45$  words). The corpus as a whole comprises 32,513 tokens. Naturally, this makes it tiny by the standards of today’s mega-corpora. However,

---

1. Scores made up of an average of each participant’s response to four questions: *English is very important to me personally*; *I like using English*; *I always use English when I have an opportunity to do so*; *Sometimes I resent the fact that I have to use English* (reversed). For details, see Edwards and Fuchs (in press).

this is not unusual in CADS approaches, which frequently make use of small, specialised corpora and emphasise intimacy with the data (Partington et al. 2013: 12).

### 5.3 Analyses

First, Wordsmith Tools (version 6; Scott 2015) was used to identify the most frequent content words.<sup>2</sup> These were, perhaps unsurprisingly, *Engels* (English,  $n = 885$ ), *Nederlands* (Dutch,  $n = 384$ ) and *taal* (language,  $n = 289$ ). Also striking was the prevalence of nouns and pronouns designating groups of people, such as *Nederlanders* (Dutch people,  $n = 221$ ) and *mensen* (people,  $n = 192$ ), followed by *we* (we,  $n = 93$ ), *iedereen* (everyone,  $n = 27$ ), *jongeren* (young people), *kinderen* (children) and *studenten* (students) (all  $n = 20$ ). This contribution reports on two studies focusing on several of these high-frequency lexical items, which yielded a manageable amount of data for closer analysis: first, ideological representations of *English*, and second, referential discourse strategies using the terms (*Dutch*) *people* (see subsections below). The assumption was that analysing such recurrent discourse patterns provides access to informants’ “world of beliefs” (Blommaert & Verschueren 1998: 191), enabling us to tease out their underlying ideas about language(s) and how they are used in society.

### 5.4 Ideological representations of English

For this analysis, collocates of the word *English* with mutual information (MI) scores higher than 3.0 (Baker 2006: 101) were extracted using the Concord tool. The procedure was introduced in Fitzsimmons-Doolan (2014) as a way of tapping ideological representations, in her case of the nouns *language*, *literacy* and *English* in a corpus of US language policy texts. Specifically, collocates of the node word *English* (i.e. words that co-occur with *English* more frequently than would be expected by chance) were extracted using a three-word span to the left or to the right of the node in order to identify premodifiers (e.g. *unnecessary English*) and subject predicative adjectives (e.g. *English is useful*). To generate an adequate amount of data, the minimum frequencies were set at 5 for premodifiers and 2 for predicative adjectives (Fitzsimmons-Doolan 2014). Adjectival uses of *English* (e.g. *English accent*) were excluded. Analysed in the context of their wider co-texts, these collocates help to shed initial light on language ideologies by revealing attitudes and beliefs about the English language that are presented affectively or as statements of common sense.

---

2. ‘Content words’ include nouns, verbs, adjectives and most adverbs, as opposed to ‘function words’, which express grammatical relationships but carry little lexical meaning.

## 5.5 Referential discourse strategies: What “the Dutch” think and do

Again using the Concord tool, all concordances of the terms *Nederlander(s)* (Dutch people/the Dutch/Dutch person (as in “the average Dutch person”); henceforth simply “Dutch people”) and *men(sen)* (people/one (as in *men denkt*, i.e. “people think”); henceforth “people”) with 400 characters of co-text were extracted. These reference markers can be seen as being strategically deployed by participants to construct an ideological narrative about themselves and other social actors (Velazquez 2013). As shown in Table 1, *Dutch people* was used 212 times in the corpus by 183 participants (out of 724); i.e. 25% of commentators. *People* were referred to 180 times by 150 or 21% of commentators.

**Table 1.** Frequency data for the node words *Dutch people* and *people*

Dutch	Translation	Freq.	No. of texts	% of commentators
<i>Nederlander(s)</i>	Dutch people	212	183	25.3
<i>men(sen)</i>	people	180	150	20.7

Next, the concordances were coded thematically using a grounded theory approach (Strauss & Corbin 1990). This entails establishing a coding scheme inductively by reading through the concordances and assigning each at least one thematic label, and adding to/adjusting these labels when subsequent concordances cannot be classified under a label already established (see e.g. Gnutzmann, Jakisch & Rabe 2015 for a related example of this approach). The same qualitative codes turned out to be applicable to both the *Dutch people* and *people* data (examples are given in the next paragraph and all codes are listed later in Table 3). Moreover, the majority of the latter cases (159/180, 88%) referred not to people in general but specifically to Dutch people. Therefore, the other 21 cases were excluded, leaving a total of 371 concordances in which the participants comment on what, in their view, they and their compatriots typically think and do (e.g. *Many Dutch people think they speak better English than the English. Fact. Comment 138, female, age 37*).

Eleven different codes emerged from the data, examples including “people overestimate their own competence in English” and “people don’t take enough pride in the Dutch language/culture”. There were a total of 409 code assignments for the 371 concordances; most concordances were given a single code, but in certain cases a secondary code was added as well.

When discussing the findings, I refer where relevant to the respondents’ socio-demographic data: their age, sex and self-reported English proficiency levels. To identify potential associations between particular thematic codes and sociodemographic variables, (i) one-way analysis of variance (ANOVA) was used to compare the mean age and proficiency levels of informants whose comments were assigned a

certain code to all other survey respondents; and (ii) chi-square tests (or Fisher’s exact tests in the event of low expected frequencies) were used to identify differences in sex distributions of informants whose comments were assigned a particular code versus all other survey respondents. All statistical tests were performed in RStudio version 0.99.491 (RStudio Team 2015) and only significant differences ( $p < .05$ ) are reported. It should be noted, however, that statistics are not the focus here. The analyses are largely exploratory and qualitative in nature, with quantitative results used to identify the most salient patterns and guide the discussion.

## 6. Results

### 6.1 Ideological representations of English

Table 2 lists the collocates of *English*, grouped into premodifiers ( $n = 7$ ) and subject predicative adjectives ( $n = 8$ ). Examination of the relevant concordances and their wider co-texts points to various shared discourse patterns and ideological narratives captured in the corpus.

**Table 2.** Adjectival collocates of the node word *English* (MI > 3.0)

Dutch	Translation	MI	Raw freq.
<b>Premodifiers</b>			
<i>Brits</i>	British	5.21	8
<i>Steenkolen</i>	coal	4.73	7
<i>Amerikaans</i>	American	4.41	21
<i>onnodig</i>	unnecessary	3.94	29
<i>goed</i>	good	3.19	179
<i>slecht</i>	bad	3.13	35
<i>zoveel</i>	so much	3.08	22
<b>Predicative adjectives</b>			
<i>nuttig</i>	useful	4.48	7
<i>prachtig</i>	beautiful/wonderful	3.99	11
<i>verplicht</i>	compulsory	3.89	8
<i>beter</i>	better	3.89	71
<i>boven</i>	above/higher	3.89	5
<i>(ten) koste (van)</i>	detrimental (to)	3.89	5
<i>mooi</i>	beautiful/nice	3.48	19
<i>slechter</i>	worse	3.04	9

Let us begin with perceptions of the instrumental value of English as a tool for intercultural contact. Predicative adjectives include, in addition to “useful”, affective

judgements such as “nice” and “beautiful”, in some cases even highlighting the especial suitability of English as a world language.

- (1) Personally I think the **English language is beautiful**, often it provides more possibilities for expression than Dutch. It’s not for nothing that it’s a global language. (comment 596, female, age 21)
- (2) I think **English is wonderful**, it enables us to communicate with many nationalities and as such is an enrichment for us here in the Netherlands, but elsewhere too. (comment 698, female, age 59)

This perceived association between the English language and contact with the wider world dovetails with academic discourses on English in the Expanding Circle/EFL societies as predominantly a useful tool for international communication. In line with this, the majority of informants clearly believe English has – or ought to have – a certain (restricted) place and role, whereby it does not infiltrate the Dutch language or affect the way compatriots communicate intranationally. Concern about English encroaching beyond its sanctioned role as a lingua franca can be seen in the use of premodifiers such as “unnecessary” and “so much”, as well as concordances referring to the disadvantages of the predominance of English and its perceived detrimental effect on mastery of both Dutch and other foreign languages.

- (3) Use of **English can be useful**, but lately it’s used too often in situations where Dutch would also suffice. For example what’s wrong with *uitverkoop* instead of “sale”? (comment 28, male, age 44)
- (4) Sometimes I do find that **unnecessary English** is used. Personally I don’t have anything against that, because I love the English language, but I think it makes communication with lots of other people (often an older generation) more difficult and I see that as an objection. (comment 225, female, age 46)
- (5) It drives me hopping mad. I already know couples who speak English with one another, it’s a matter of time until it becomes normal for highly educated people to raise your children in it. Then Dutch will disappear like the regional languages are now. [...] Also, the advance of **English has been detrimental** to our ability to speak other languages. Previously, every Dutch person could speak fluent German, these days I know almost nobody my age who still uses German in Germany. (comment 83, male, age 29)

The comparative adjectives “better” and “higher” were used to attribute the perceived threat to the national language to a belief among some people that English has a higher status than Dutch:

- (6) I don’t think that English is a threat to Dutch, it’s the Dutch who are a threat to Dutch. Apparently it’s thought that **English is somehow better** or more beautiful than Dutch. [...] Own language first, I say! (comment 384, female, age 41)

- (7) I find what’s sometimes happening in the Netherlands a real shame, because I get the feeling that for too many people the **English language is higher** than Dutch. If they also have terrible pronunciation and make grammatical mistakes that are just really not okay, I sort of have the feeling, you traded your Dutch in for this much worse level of speaking? Do you actually realise what you’re doing? (comment 517, female, age 26)

The latter comment reveals a tension between the narratives that, on the one hand, people should not use English so often and, on the other hand, they should be better at using it. That is, in addition to the discourse on “unnecessary” English, another central discourse revealed by the collocations revolved around highly normative views of “good” English, whereby Received Pronunciation is seen as the ideal, US influence is undesirable, transfer from Dutch is ridiculed and English-language competence is presented as woefully lacking. These normative discourses were marked with attributive adjectives such as “good” and “bad”, “British” and “American”, and “coal” (from *Steenkolen-Engels*, a derogatory term roughly meaning “broken English” or “Dunglish”).

- (8) I’m sometimes ashamed for England and the **coal English** that Dutch people manage to produce. Particularly the accent is painful. (comment 6, female, age 23)
- (9) There is much too much **bad English** used in the Netherlands and added to that lots of half-American and incomprehensible computer language. (comment 686, female, age 77)
- (10) I usually find it awful to hear **American English**. Particularly the way words like “fast” and “last” are pronounced [...] hurts my ears. To stay closer to what we see as the “original” English, my wife and I watch the BBC journal at 10pm (Dutch time) almost every evening. (comment 258, male, age 65)

Only four comments identified in this analysis went against the grain by questioning or contesting the standard language ideology and suggesting that a locally flavoured English variety may have its merits.

- (11) I think **Dutch English is wonderful!** It should be appreciated a bit more! (comment 578, female, age 28)
- (12) Although I personally tend to want to speak as **good English** as possible, I’m not really bothered by the bit of “Dutch sauce” that we add to it, I think that never entirely goes away if you don’t speak English constantly. It also has its charm, and anyway there are enough differences even between the versions that are spoken in English-speaking countries. So maybe “Dutch English” is just another variant. (comment 716, male, age 55)

This initial exploration of how the English language is referenced in the corpus by means of attributive and predicative adjectives has revealed some key ideological discourses, namely a traditional view that English is to be used only when “necessary” and a highly normative view of “good” and “bad” English. The next section turns to how “the Dutch” are referenced, which, as we shall see, extends these ideological narratives into how language(s) are understood in general.

## 6.2 Referential discourse strategies: What “the Dutch” think and do

Table 3 presents a thematic classification of all concordances involving the node words *Dutch people* and *people*, reflecting the sets of propositions informants put forth concerning the behaviours and beliefs of (segments of) the Dutch population. The total adds up to more than 100% because, as mentioned previously, some concordances were given several codes (409 code assignments for 371 concordances). Looking at the first four rows, we can see that some 75% of code assignments reflect and extend some of the key discourses identified in the previous section: a normative (and critical) stance towards English competence, and a belief that “unnecessary” overuse of English will have a negative impact on Dutch.

**Table 3.** Thematic classification of the concordances of *Nederlander(s)* and *men(sen)*

Code	Theme	No. of code assignments	% of code assignments
1	English competence is lacking	83	22.4%
2	People overestimate their own English competence	75	20.2%
3	People (too) readily use English, e.g. to seem interesting/cool/cosmopolitan	74	19.9%
4	English is threatening/impacting Dutch, or causing disadvantage/domain loss	52	14.0%
5	Other*	34	9.2%
6	English competence is generally good	24	6.5%
7	People don't take (enough) pride in the Dutch language/culture	19	5.1%
8	English is useful/is to be used for international contacts	18	4.9%
9	English is not a threat but an enrichment, languages evolve	13	3.5%
10	Other languages are important to learn too/are no longer being learnt	9	2.4%
11	English doesn't sound authentic when used by the Dutch	8	2.2%
	<b>Total</b>	<b>409</b>	<b>110.2%</b>

\* Includes any theme with a less than 2% share, e.g. communication is more important than perfection, more instruction in English required.

Broadly speaking, many informants seemed to see their compatriots as willingly, if blindly, trading in Dutch for English (whether in reference to codeswitching or to the use of English in traditionally Dutch-language domains; code 3). These informants tended to be slightly older than the other questionnaire respondents (mean age 46 cf. overall mean 44,  $F(1,2079) = 3.70, p = .05$ ). Paradoxically, however, in positioning themselves as standing alone against this perceived rush towards English, such commenters in fact consistently reproduced one of the most dominant discursive routines (“everyone else thinks/does  $x$ , but I think/do  $y$ ”). *Dutch people* and formulae such as *some people* were often used as markers of distance, distinguishing informants’ own views and practices from what they posited as the beliefs and behaviours of their compatriots.

- (13) I’m an Anglophile, but I find it really unfortunate that **Dutch people** use so many English words when there’s a perfectly good Dutch word for them.  
(comment 319, female, age 71)
- (14) I’m sometimes annoyed to death by the frequent use of English words in the media. Often there’s a beautiful Dutch word for them, but still an English word is chosen. **Dutch people** who use “flabbergasted” all the time, I have to throw up in my mouth a little.  
(comment 25, female, age 35)
- (15) Obviously command of English is of vital importance. But **some people** take things too far. Absurd situations arise when **Dutch people** speak English with **Dutch people**. There are even institutes that forbid their students from speaking Dutch in the cafeteria.  
(comment 326, male, age 63)
- (16) I’m just back from [a film festival in The Hague] and there you’re even greeted in English [...]. That we’re not welcomed in Dutch (by **Dutch people**), I find it rude. English seems to be more important than Dutch here in our own country. For me that’s going too far!  
(comment 337, female, age 62)
- (17) I know **people** who in my opinion vastly overestimate English and would have no problem relegating Dutch to a dialect. The use of English is useful for companies, internationalisation and the like, but that that’s where the money is (sigh) is no reason to not speak your mother tongue.  
(comment 57, male, age 21)

Implicit in commentary on the perceived overuse of English is the view that English not kept in its proper place poses a threat to Dutch and Dutch speakers (code 4). Again, some also highlighted the threat posed to the learning of other languages; these comments (code 10) tended to be made by older informants, presumably those whose education placed more emphasis on French and German than do modern curricula (mean age 55 cf. 44,  $F(1, 2079) = 3.82, p = .05$ ).



- (18) I have the feeling young generations place more value on English than Dutch. I fear for this reason that Dutch will die out. In fifty years Dutch will only be spoken by **older people** who don't have a good command of English. Speaking English is good in certain circumstances. However, everyone in the world should always master their own mother tongue. Only then is there room for a "foreign language", like English. (comment 642, male, age 20)
- (19) I get the sense that there are **many people** who are fed up with the frequent use of English. [...] **Dutch people** act so international and perhaps they actually were in the past; knowledge of foreign languages was above average compared to other countries. But now that's no longer the case. They've made themselves into provincials, a province of the Anglo-Saxon world. (comment 678, male, age 53)

Commenters put forth several assertions as to what, in their view, prompts the perceived overuse of English. Key among these were that people wish to be seen as interesting, cool or cosmopolitan (code 3) or do not take enough pride in the Dutch language or culture (code 7).

- (20) English is used improperly, incl. a lot in advertising, to suggest progressiveness and other fashionable nonsense, which mainly hits its mark among **people** with little feel for language. (comment 699, male, age 70)
- (21) A certain group of **Dutch people** seem to find it interesting to pepper their Dutch with English terms. They then appear, in their own eyes, more interesting than if they were only to use Dutch words. (comment 696, male, age 41)
- (22) English, or rather American, is not the threat to Dutch but **Dutch people** who so love to rub up against a dominant culture and come out with English all the time. The Dutch government is guilty of this too. Pride in Dutch is seen as a pernicious trait. (comment 556, male, age 65)
- (23) I find the zeal with which many **Dutch people** try to belong to the Anglo-Saxon culture almost childish, the fact that many **Dutch people** only want to look west when it comes to language and culture almost pathological. (comment 617, male, age 42)

Besides the narrative on overuse of English, the thematic classification revealed a highly normative discourse on the apparently poor English-language competence of the Dutch (code 1). Commenters criticising English levels in the Netherlands were on average older than other survey respondents (mean age 48 vs 44,  $F(1, 2079) = 5.48$ ,  $p = .02$ ). A highly routinized discourse asserting that Dutch people tend to overestimate their English-language skills (code 2) was discernible, with comments in this vein made more commonly by respondents with marginally higher proficiency levels than others (mean 3.7 vs 3.6 on a 4-point scale,  $F(1, 2076) = 7.08$ ,  $p = .008$ ).

Together, claims about the supposed lack of English competence and inflated views of that competence accounted for over 40% of the assigned codes.

- (24) In general the average Dutch person can speak quite a bit of English, but in 70% of **people** it sounds lame, while they think it’s fantastic. That gets on my nerves. (comment 580, female, age 20)
- (25) It’s sad how well **people** think they speak English and how uncritically they view their own performance in that area. Besides the grammar or word choice that also/especially applies to pronunciation. (comment 284, female, age 70)

Coexisting with these criticisms of proficiency levels was a somewhat contradictory narrative, albeit a much less prevalent one: even as people perceived as using *Steenkolen-Engels* were criticised, those seen as fully adopting “native” English were accused of a lack of authenticity (code 11).

- (26) It strikes me that when I hear **Dutch people** speak perfectly accentless English, or in any case English that sounds very “native”, I don’t like it. Whereas it’s obviously a great achievement, and I’d like to be able to do it myself. But it creates distance somehow. As though he or she is no longer one of us. Apparently I’m happy for **Dutch people** to switch languages, but not accent. I also personally recognise many situations where I speak English, but more or less deliberately retain my Dutch accent. Often among friends or colleagues whom I trust. (comment 420, male, age 54)
- (27) When selecting an accent in English you also have to make a normative choice about the class you want to belong to. Even Received Pronunciation is associated with a certain class. [...] That’s why I’ve chosen to retain my Dutch accent in English instead of going for Oxford English, for example. **People** also have prejudices/preconceptions about **people** with Denglish, but at least I fulfil them. (comment 173, male, age 30)

Two other counter-discourses are worth mentioning. First, a minority of the coded concordances highlighted the generally good English proficiency levels in the Netherlands (code 6). Although these were heavily outweighed by comments critical of those levels, they often revealed some measure of “discursive consciousness” (Kroskrity 1998: 117): awareness of the striking discord between the frequency and intensity of public critiques of English proficiency in the Netherlands and the reality of having among the highest reported competence levels in continental Europe (European Commission 2012).

- (28) In England people are really impressed by the level of English of **most Dutch people** I wouldn’t change too much about that. Certainly compared to other European countries. (comment 426, male, age 48)

- (29) **People** should complain less about the Dutch accent of e.g. politicians who speak English. (comment 18, male, age 34)

Finally, another minority discourse pushed back against the dominant view of English as a threat, contesting the perception that Dutch is as imminent risk and/or positing English as an enrichment (code 9).

- (30) **People** shouldn't panic so much about the "Anglicisation" of Dutch. It's an enrichment, just as French once was. (comment 15, male, age 35)
- (31) Many **people** complain that Dutch is becoming increasingly Anglicised, but a hundred years ago **people** already worried about that. The influence of English is now greater, due to the new media (internet, games), but it certainly won't wipe out Dutch in the short term. (comment 250, female, age 35)

## 7. Discussion and conclusion

The findings echo those in Blommaert and Verschueren's (1998: 206) study of language ideologies and European nationalism: "The way in which language is presented [reveals] a decidedly unsophisticated folk view". This view has its roots in the European Enlightenment and later Romantic perceptions, and is often associated with the philosophies of Johann Gottfried Herder (Gal 2006: 15). It is characterised by traditional ideas about standard languages, which are assumed to be objectively definable and ideally kept "pure", and the notion that "there are discrete, clearly bounded languages with better and worse varieties and that the function of language is to denote things in the world" (Johnstone 2016: 429).

Such beliefs have, as classic language ideologies, "become so well-established that their origin is often forgotten by speakers; the beliefs accordingly become naturalized, perceived as common sense, and are socially reproduced" (Vessey 2015: 2). Indeed, despite a few contestations, they were reproduced remarkably consistently in the present data. The analyses revealed two main ideological threads: first, the notion that languages have a "proper" place and (communicative) purpose, and their use to any other effect is "unnecessary"; and second, the normative idea that linguistic forms and practices can be judged, self-evidently, as "good" or "bad".

These public discourses reflect those "official" discourses to which learners have conventionally been exposed through English language teaching, whereby falling short of the "native" ideal renders one's usage "deficient". Moreover, they also reflect canonical academic discourses on English as a global language. In WEs, English in the Expanding Circle has traditionally been constructed as a "foreign" or "international" language, a mere tool for transnational communication. Although many researchers now recognise that in some Expanding Circle settings English

is an important additional means of expression as well as an economic commodity (Pennycook 2003), this shift has clearly not yet trickled down into the public imagination. While only a small proportion of the comments explicitly described English as a useful language for international communication, the implicit view that it is appropriate in that context only was clear. However, there was also a discernible tension between perception and practice. Despite the explicit and implicit insistence on positioning English as purely utilitarian, the prevalence of the narratives on “unnecessary” use and “overuse” testifies to apparently highly salient “superfluous” uses of the language that far exceed the merely instrumental. In future research it may be interesting to combine the direct method of eliciting language beliefs used here with a more indirect approach to explore whether people’s beliefs about the “unnecessary” use and “overuse” of English tally with their actual linguistic behaviour.

This perceived “unnecessary” use and “overuse” may be just that – a perception, and one that is particularly prevalent among more conservative members of society. As previously noted, the corpus population was on average older and somewhat more negative towards English than the rest of the survey respondents. Indeed, previous research has shown that English codeswitches in Dutch newspapers, for instance, are popularly perceived as more frequent than they really are (van der Sijs 2012). However, given the undeniable increase of English in various domains in the Netherlands, the informants can be considered at least in part to be responding to actual changes in society.

An interesting illustration of the “unnecessary” use of English can be seen in the fact that some 48 informants wrote their comments in English rather than Dutch. The researcher does have a particularly English-sounding name, but as the questionnaire and all accompanying information was in Dutch, it will have been clear that English was not necessary for the purposes of communication. Instead such a choice appears to be connected, consciously or otherwise, with the perception and presentation of self. The analyses showed that many commentators were critical of compatriots who seemingly used English to make themselves appear cool, clever or cosmopolitan. Such mobilisation of the language was apparently felt to be unsanctioned; to betray an attempt to project an image of oneself that one was not licensed to hold. This suggests two things. First, a sizeable chunk of the population is (perceived to be) actively engaged in constructing and performing an “English-knowing” identity. Second, many of their peers pit themselves in opposition to this, as people who know how to use English “properly” (recall comments such as ‘It’s sad how well people think they speak English’, where *people* is used to create distance). For this latter group, therefore – somewhat paradoxically – English may also be seen as playing a role in their identity construction as it is for the subjects of their criticism (see also Edwards & Seargeant *fc.*).

In this way, English appears to play an important role in social positioning and in the negotiation and enactment of identity in the Netherlands, as has been found in various western and northern European countries (Erling 2004; Hilgendorf 2001; Hult 2003; Preisler 1999; Leppänen et al. 2011; Taavitsainen & Pahta 2008). This indexical role is one it is not afforded in traditional WEs discourse. Schneider's (2014: 28) Transnational Attraction framework, too, still appears to emphasise instrumental motivations underlying English use. Certainly this is how it is interpreted in D'Angelo (2016: 137, 174), where TA is equated with a sort of internationalism or "world mindedness", characterised by an interest in "global involvement" and a desire to connect with "other countries' people and cultures". TA does tap into recent poststructural and postnational trends in the humanities and social sciences: it replaces the traditional WEs focus on bounded varieties with an emphasis on language as practice or activity rather than system and on global flows and processes that transcend national borders (Blommaert & Rampton 2011; Heller 2008; Pennycook 2010). English enters countries like the Netherlands for reasons that are often (but not always) economic in nature; for example, universities wish to tap into the globalised higher education market. But once entrenched, it does not remain foreign. Rather, it is reterritorialized so as to serve the purposes of local meaning-making and identity construction. Schneider (2014: 28) acknowledges this by suggesting that the forces behind TA may ultimately "transcend [...] economic motivations", leading English to develop "new roles" in Expanding Circle societies. In this way, the TA framework can contribute to an emerging rapprochement between WEs, with its traditional focus on native and postcolonial varieties, and more ethnographic and sociocultural approaches such as the sociolinguistics of globalisation (Blommaert 2010), which seek to explore English wherever and in whatever form it is found (see further Bolton 2013; James 2016; Mair 2016). The further elaboration of the TA model – and other modelling efforts in the field of World Englishes – can benefit from the incorporation of a language ideologies approach (largely neglected in the field so far) as such an approach helps to shed light on people's attitudes to English and the ways in which they construct their English-using identities.

The corpus used in this study, although small, was sufficient to give rise to a fairly robust picture of informants' language ideologies and to highlight some emerging indexical, rather than purely instrumental, functions of English in an Expanding Circle setting. Naturally, the study makes no claims to conclusiveness or "truth". Corpus-assisted discourse analyses, like all discourse analyses, are not objective but subject to the analyst's gaze, and a different focus would undoubtedly have yielded different thematic codes and interpretations. Still, the combination of the CADS and language ideologies approaches has proven fruitful, providing a snapshot of Dutch attitudes to English at a time of transition, when the increasing presence of the language is being embraced by some and vigorously protested by more conservative elements in society.

## References

- Ammon, U. & McConnell, G. D. 2002. *English as an Academic Language in Europe*. Frankfurt: Peter Lang.
- Baker, P. 2006. *Using Corpora in Discourse Analysis*. London/New York: Continuum.
- Baker, P., Gabrielatos, C., Khosravimik, M., Krzyżanowski, M., McEnery, T. & Wodak, R. 2008. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society* 19(3): 273–306. <https://doi.org/10.1177/0957926508088962>
- Baker, P. & McEnery, T. 2015. *Corpora and Discourse Studies: Integrating Discourse and Corpora*. Houndmills: Palgrave Macmillan. <https://doi.org/10.1057/9781137431738>
- Bakker, J. J. 1987. De Woordenaar: Tien redenen. *Onze Taal* 56(6): 73.
- Bleichenbacher, L. 2006. Dobry den Košice – üdvözlöm Kassát – Hello Kosice: Language choice in a Slovak internet guestbook. In *Language Ideologies, Policies and Practices*, C. Ma-Molinero & P. Stevenson (eds), 220–36. Houndmills: Palgrave Macmillan. [https://doi.org/10.1057/9780230523883\\_16](https://doi.org/10.1057/9780230523883_16)
- Blommaert, J. 2005. *Discourse: A Critical Introduction*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511610295>
- Blommaert, J. 2010. *The Sociolinguistics of Globalization*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511845307>
- Blommaert, J. & Rampton, B. 2011. Language and superdiversity. *Diversities* 13(2): 1–21.
- Blommaert, J. & Verschueren, J. 1998. The role of language in European nationalist ideologies. In *Language Ideologies: Practice and Theory*, B. Schieffelin, K. Woolard, & P. Kroskrity (eds), 189–210. Oxford: OUP.
- Bolton, K. 2013. World Englishes, globalisation, and language worlds. In *Of Butterflies and Birds, of Dialects and Genres: Essays in Honour of Philip Shaw*, N-L. Johannesson, G. Melchers, & B. Björkman (eds), 227–251. Stockholm: Acta Universitatis Stockholmiensis.
- D’Angelo, J. 2016. A broader concept of World Englishes for educational contexts: Applying the “WE enterprise” to Japanese higher education curricula. PhD dissertation, North-West University.
- de Bot, K. & Clyne, M. 1994. A 16-year longitudinal study of language attrition in Dutch immigrants in Australia. *Journal of Multilingual and Multicultural Development* 15: 17–28. <https://doi.org/10.1080/01434632.1994.9994554>
- de Bot, K. & Weltens, B. 1997. Multilingualism in the Netherlands? In *Perspectives on Foreign-Language Policy: Studies in Honour of Theo Van Els*, T. Bongaerts & K. de Bot (eds), 143–56. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.81.12bot>
- Dronkers, J. 2013. Sluipenderwijs wordt Nederland tweetalig. *De Volkskrant*, 17 July 2013, <[www.volkskrant.nl/vk/nl/3184/opinie/article/detail/3476812/2013/07/17/Sluipenderwijs-wordt-Nederland-tweetalig.dhtml](http://www.volkskrant.nl/vk/nl/3184/opinie/article/detail/3476812/2013/07/17/Sluipenderwijs-wordt-Nederland-tweetalig.dhtml)> (11 January 2017).
- Edwards, A. 2016. *English in the Netherlands: Functions, Forms and Attitudes* [Varieties of English Around the World G56]. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g56>
- Edwards, A. & Fuchs, R. In press. Varieties of English in the Netherlands and Germany. In *English in the German-speaking world*, R. Hickey (ed.). Cambridge: CUP.
- Edwards, A. & Seargeant, P. fc.. Beyond English as a second or foreign language: Local uses and the cultural politics of identification. In *Cambridge Handbook of World Englishes*, D. Schreier, M. Hundt, & E. Schneider (eds). Cambridge: CUP.

- Erling, E. J. 2004. Globalization, English and the German University classroom. PhD dissertation, Freie Universität Berlin.
- European Commission. 2012. Europeans and their languages: Special Eurobarometer 386. Brussels. <[http://ec.europa.eu/public\\_opinion/archives/ebs/ebs\\_386\\_en.pdf](http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf)> (11 January 2017).
- Fitzsimmons-Doolan, S. 2014. Using lexical variables to identify language ideologies in a policy corpus. *Corpora* 9(1): 57–82. <https://doi.org/10.3366/cor.2014.0051>
- Freake, R. 2012. A cross-linguistic corpus-assisted discourse study of language ideologies in Canadian newspapers. *Proceedings of the 2011 Corpus Linguistics Conference*, Birmingham University. <[www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-17.pdf](http://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-17.pdf)> (11 January 2017).
- Gal, S. 2006. Migration, minorities and multilingualism: Language ideologies in Europe. In *Language Ideologies, Policies and Practices*, C. Mar-Molinero & P. Stevenson (eds), 13–27. Houndmills: Palgrave Macmillan. [https://doi.org/10.1057/9780230523883\\_2](https://doi.org/10.1057/9780230523883_2)
- Gnutzmann, C., Jakisch, J. & Rabe, F. 2015. Communicating across Europe: What German students think about multilingualism, language norms and English as a lingua franca. In *Attitudes Towards English in Europe*, A. Linn, N. Bermel & G. Ferguson (eds.), 78–88. Berlin: Walter de Gruyter.
- Graedler, A. L. 2014. Attitudes towards English in Norway: A corpus-based study of attitudinal expressions in newspaper discourse. *Multilingua* 33(3–4): 291–312. <https://doi.org/10.1515/multi-2014-0014>
- Heller, M. 2008. Language and the nation-state: Challenges to sociolinguistic theory and practice. *Journal of Sociolinguistics* 12(4): 504–524. <https://doi.org/10.1111/j.1467-9841.2008.00373.x>
- Higgins, C. 2009. *English as a Local Language: Post-Colonial Identities and Multilingual Practices*. Bristol: Multilingual Matters.
- Hilgendorf, S. K. 2001. Language contact, convergence, and attitudes: The case of English in Germany. PhD dissertation, University of Illinois at Urbana-Champaign.
- Hult, F. M. 2003. English on the streets of Sweden: An ecolinguistic view of two cities and a language policy. *Working Papers in Educational Linguistics* 19: 43–63.
- Hundt, M. & Mukherjee, J. 2011. Discussion forum: New Englishes and Learner Englishes – quo vadis? In *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap* [Studies in Corpus Linguistics 44], J. Mukherjee & M. Hundt (eds), 209–217. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.44.11muk>
- James, A. 2016. From code-mixing to mode-mixing in the European context. *World Englishes* 35(2): 258–275. <https://doi.org/10.1111/weng.12194>
- Johnstone, B. 2016. Language theory in contemporary sociolinguistics: Beyond Dell Hymes? In *Sociolinguistics: Theoretical Debates*, N. Coupland (ed.), 417–432. Cambridge: CUP. <https://doi.org/10.1017/CBO9781107449787.020>
- Kachru, B. B. 1985. Standards, codification and sociolinguistic realism. In *English in the World*, R. Quirk & H. Widdowson (eds), 11–30. Cambridge: CUP.
- Kachru, Y. & Nelson, C. L. 2006. *World Englishes in Asian Contexts*. Hong Kong: Hong Kong University Press.
- Kirkpatrick, A. & McLellan, J. 2012. World Englishes and/or English as a lingua franca. In *The Routledge Handbook of Discourse Analysis*, J. P. Gee & M. Handford (eds), 654–699. New York NY: Routledge.
- Kroskrity, P. 1998. Arizona Tewa Kiva speech as a manifestation of a dominant language ideology. In *Language Ideologies: Practice and Theory*, B. Schieffelin, K. Woolard, & P. Kroskrity (eds), 103–122. Oxford: OUP.



- Laitinen, M. 2015. English “on the move”: What’s beyond modernity and internationalism? In *Places in Change: Dimensions of Sociolinguistic Landscapes in Europe*, M. Laitinen & A. Zabrodska (eds), 105–124. Frankfurt: Peter Lang. <https://doi.org/10.3726/978-3-653-05170-4>
- Leppänen, S., Pitkänen-Huhta, A., Nikula, T., Kytölä, S., Törmäkangas, T. et al. 2011. *National survey on the English language in Finland* [Studies in Variation, Contacts and Change in English 5]. <[www.helsinki.fi/varieng/series/volumes/05](http://www.helsinki.fi/varieng/series/volumes/05)> (17 February 2014).
- Mair, C. 2016. Englishes beyond and between the three circles: World Englishes research in the age of globalisation. In *World Englishes: New Theoretical and Methodological Considerations* [Varieties of English Around the World G57], E. Seoane & C. Suárez-Gómez (eds), 17–36. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g57.02mai>
- Maiworm, F. & Wächter, B. 2014. The big picture. In *English-Taught Programmes in European Higher Education: The State of Play in 2014*, B. Wächter & F. Maiworm (eds), 25–62. Bonn: Lemmens.
- Moschonas, S. & Spitzmüller, J. 2010. Prescriptivism in and about the media. In *Language Ideologies and Media Discourse*, S. Johnson & T. M. Milani (eds), 17–40. London: Continuum.
- Nortier, J. 2011. The more languages, the more English? A Dutch perspective. In *English in Europe Today* [AILA Applied Linguistics Series 8], A. Houwer & A. Wilton (eds), 113–132. Amsterdam: John Benjamins. <https://doi.org/10.1075/aals.8.08nor>
- Partington, A. 2004. Corpora and discourse, a most congruous beast. In *Corpora and Discourse*, A. Partington, J. Morley, & L. Haarman (eds), 11–20. Bern: Peter Lang.
- Partington, A., Duguid, A. & Taylor, C. 2013. *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)* [Studies in Corpus Linguistics 55]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.55>
- Pennycook, A. 2003. Global Englishes, Rip Slyme, and performativity. *Journal of Sociolinguistics* 7(1992): 513–533. <https://doi.org/10.1111/j.1467-9841.2003.00240.x>
- Pennycook, A. 2010. *Language as a Local Practice*. Abingdon: Routledge.
- Preisler, B. 1999. Functions and forms of English in a European EFL country. In *Standard English: The Widening Debate*, T. Bex & R. Watts (eds), 239–267. Routledge: London.
- Ridder, S. 1995. English in Dutch. *English Today* 11(4): 44–50. <https://doi.org/10.1017/S0266078400008622>
- RStudio Team. 2015. *RStudio: Integrated Development for R*. Boston MA: RStudio.
- Rumsey, A. 1990. Wording, meaning and linguistic ideology. *American Anthropologist* 92(2): 346–361. <https://doi.org/10.1525/aa.1990.92.2.02a00060>
- Schneider, E. W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511618901>
- Schneider, E. W. 2014. New reflections on the evolutionary dynamics of world Englishes. *World Englishes* 33(1): 9–32. <https://doi.org/10.1111/weng.12069>
- Scott, M. 2015. *WordSmith Tools*, Version 6. Liverpool: Lexical Analysis Software.
- Seargeant, P. 2012. *Exploring World Englishes: Language in a Global Context*. London: Routledge.
- Silverstein, M. 1979. Language structure and linguistic ideology. In *The Elements: A Parasession on Linguistic Units and Levels*, P. Clyne, W. Hanks, & C. Hofbauer (eds), 193–248. Chicago IL: Chicago Linguistic Society.
- Smaakman, D. 2006. Standard Dutch in the Netherlands: A sociolinguistic and phonetic description. PhD dissertation, Radboud University Nijmegen.
- Strauss, A. & Corbin, J. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park CA: Sage.
- Taavitsainen, I. & Pahta, P. 2008. From global language use to local meanings: English in Finnish public discourse. *English Today* 24(2): 25–38. <https://doi.org/10.1017/S0266078408000266>



- van der Haagen, M. 1998. *Caught Between Norms: The English Pronunciation of Dutch Learners*. The Hague: HAG.
- van der Sijs, N. 2012. Engelse leenwoorden revisited. *Onze Taal* 5: 132–134.
- van Meurs, F., Korzilius, H., Planken, B. & Fairley, S. 2007. The effect of English job titles in job advertisements on Dutch respondents. *World Englishes* 26(2): 189–205. <https://doi.org/10.1111/j.1467-971X.2007.00501.x>
- van Oostendorp, M. 2012. Bilingualism versus multilingualism in the Netherlands. *Language Problems and Language Planning* 36(3): 252–272. <https://doi.org/10.1075/lplp.36.3.04van>
- Velazquez, I. 2013. Individual discourse, language ideology and Spanish transmission in El Paso, Texas. *Critical Discourse Studies* 10(3): 245–262. <https://doi.org/10.1080/17405904.2013.789975>
- Vessey, R. 2015. Corpus approaches to language ideology. *Applied Linguistics* 38(3): 277–296. <https://doi.org/10.1093/applin/amv023>
- Woolard, K. 1998. Introduction: Language ideology as a field of inquiry. In *Language Ideologies: Practice and Theory*, B. Schieffelin, K. Woolard, & P. Kroskrity (eds), 3–50. Oxford: OUP.
- Zenner, E., Speelman, D. & Geeraerts, D. 2012. Cognitive sociolinguistics meets loanword research: Measuring variation in the success of anglicisms in Dutch. *Cognitive Linguistics* 23(4): 749–792. <https://doi.org/10.1515/cog-2012-0023>

# American and/or British influence on L2 Englishes – Does context tip the scale(s)?

Gaëtanelle Gilquin

Université catholique de Louvain

Taking Mair's (2013) World System of Englishes as a starting point, this chapter seeks to investigate whether American English is a more important source of influence than British English for the other varieties of English, including English as an institutionalised second language and English as a foreign language. The study is based on twenty pairs of items that are distinctive between American and British English and whose frequency is calculated in data from the *Global Web-based English Corpus* (GloWbE) and the *EF-Cambridge Open Language Database* (EFCAMDAT). The results reveal a global influence of American English, as predicted by Mair's model, but also show that varieties are not necessarily homogeneous in this respect and that more local contextual factors may have an impact on the degree of American and/or British influence.

**Keywords:** New Englishes, Learner Englishes, World System of Englishes, Americanization, context of acquisition

## 1. Introduction

In 1870, Richard Grant White, an American critic, wrote the following in the preface to his book entitled *Words and Their Uses, Past and Present: A Study of the English Language*:

[I]n my remarks on what I have unavoidably called, by way of distinction, British English and "American" English, and in my criticism of the style of some eminent British authors, no insinuation of a superiority in the use of their mother tongue by men of English race in "America" is intended, no right to set up an independent standard is implied. Of the latter, indeed, there is no fear. When that new "American" thing, so eagerly sought, and hitherto so vainly, does appear, if it ever do (sic) appear, it will not be a language, or even a literature. (p. 8)

Today, a century and a half later, we can say that American English has not only given rise to a national literature that is recognised as such, but that it has also become, if not a language of its own, at least an independent standard and a well-established variety of English, on a par with British English. In fact, it seems that the student may even have surpassed the teacher, so to say, as it has been suggested that American English, rather than British English, should now be seen as the most prevalent model among speakers of English. Thus, when describing the ‘World System of Englishes’, in which the different varieties of English are organised hierarchically, Mair (2013: 260) notes that “[a]t the risk of causing offence to British readers, the hub of the ‘World System of Englishes’ is Standard American English”. Being the hub of the system, Standard American English is considered to be the variety that is relevant to all other varieties and is “a potential factor in the[ir] development” (Mair 2013: 261). Without necessarily formalising it by means of a model, several scholars have made a similar claim and assigned American English a leading role in the current development of English throughout the world (see Section 2).

This chapter starts from Mair’s (2013) World System of Englishes and seeks to test the purported influential role of American English on the other varieties of English, using corpus data as evidence and considering British English as a possible competitor. While Mair (2013), in his model, only includes varieties of English belonging to the Inner and Outer Circles, here I will also study the potential influence of American English on varieties from the Expanding Circle.<sup>1</sup> In doing so, I follow Schneider’s (2014) suggestion to apply models designed for the Inner and Outer Circles to the Expanding Circle (in his case, the Dynamic Model, see Schneider 2003, 2007) and I build on a recent trend which consists in bringing closer together research on Outer Circle varieties and research on Expanding Circle varieties (see Section 2).

The chapter is organised as follows. Section 2 considers the place of American and British English as potential standards. It also introduces Mair’s (2013) model in more detail and presents the three hypotheses that will be tested in this study. Section 3, on data and methodology, describes the corpora used in the analysis, viz. GloWbE (*Global Web-based English Corpus*) and EFCAMDAT (*EF-Cambridge Open Language Database*), explains the process of selection of the linguistic features, and introduces the concepts of ‘Americanness rate’ and ‘Britishness rate’. The analysis itself, in Section 4, is divided into two parts. The first one investigates the potential influence of American and British English on the varieties associated with each of

---

1. See Kachru (1985) on the distinction between the Inner Circle, which includes countries where English is a native language, the Outer Circle, corresponding to former British or American colonies where English is used as an institutionalised second-language variety for both intra- and international communication, and the Expanding Circle, where English is learned as a foreign language and used for international communication only.

the three Circles, that is, English as a native language (for the Inner Circle), English as an institutionalised second language (for the Outer Circle) and English as a foreign language (for the Expanding Circle). In the second part, a more local type of influence is considered, by distinguishing between the different countries (or, in the case of the Expanding Circle, continents) that are represented in GloWbE and EFCAMDAT. Section 5 zooms in on the varieties of English as a foreign language and shows how the context of acquisition and use of these varieties may account for some of the results obtained in the analysis. The conclusions of this study are found in Section 6.

## 2. The place of British and American English

Historically, British English (BrE) is the “mother variety” (Simo Bobda 1998: 18), the variety from which all the others originally developed. As such, it initially enjoyed linguistic supremacy and was seen to carry considerable prestige, especially through its Received Pronunciation. American English (AmE), by contrast, was considered for a long time as the “underdog”, a kind of “colonial substandard” (Kahane 1992: 212). In the decades following World War II, as a result of “the simultaneous rise of the US as a military and technological power and the decline of the British Empire” (Simo Bobda 1998: 14), the tide started to turn. AmE gained credibility and respectability, and from the 1990s onwards claims about the (forthcoming) superiority of AmE became more widespread. Thus, Simo Bobda (1998: 14) notes that since World War II, “American English has continuously spread its tentacles all over the globe”. Clark (1998: 18) suggests that “American English – penetrating in the wake of Coke, Levis and McDonalds to the outermost ends of the earth – is well on the way to becoming the global standard”. As for Kahane (1992: 211), he claims that “[o]urs is the day of American English”. Some scholars are more moderate and consider that both BrE and AmE can be recognised as valid models (see, e.g., Modiano 1999: 5; Grzegza 2005). This view may be due to the fact that different factors point to different potential models. Thus, Algeo (2006: 1) observes that “American has more native speakers than British and is rapidly becoming the dominant form of English in non-native countries other perhaps than those of Western Europe. Much European established academic bias favors British as a model; but evolving popular culture is biased toward American”. This quote highlights some of the relevant factors in the choice of a model, namely the number of native speakers, the role of popular culture and that of education. While the first two factors predict an American influence, the last one makes a British preference possible. The quote also suggests that different models may be selected in different parts of the world. The European bias towards BrE, in particular, has been underlined by several scholars (see, e.g., Ranta 2010 on the Finnish situation).

Mair's (2013) World System of Englishes relies on the assumption that it is (Standard) AmE that has become the most central variety of English, the 'hyper-central variety', around which all the other varieties are organised hierarchically. BrE, by contrast, is relegated to a lower level, that of the 'super-central varieties', together with Australian English, Indian English, Nigerian English, South African English, African American Vernacular English and "a very small number of others" (Mair 2013: 261). The next level, that of the 'central varieties', is occupied by varieties like Irish English, Canadian English or Kenyan English. Finally, the 'peripheral varieties' include Maltese English and Cameroonian English, among others. As explained by Mair, linguistic traffic is more likely to go 'downwards' than 'upwards'. Lexical borrowings, for example, are expected to come from AmE and spread into the other varieties, rather than the other way round; Irish English is more likely to borrow words from BrE than BrE from Irish English since the latter belongs to a lower hierarchical level than BrE. Mair's model, and its division into several hierarchical layers, mainly relies on sociolinguistic considerations, such as "demographic weight and institutional support" (2013: 258). The purely linguistic considerations, on the other hand, are limited to "anecdotal evidence" (2013: 263) as well as an illustrative corpus-based study of Nigerian Pidgin. The first aim of the present chapter is to empirically test the linguistic consequences of one of the claims made by Mair, namely that AmE, rather than BrE, is the hyper-central variety. Following his claim, the first hypothesis that will be tested in this chapter is that AmE exerts more influence on all the other varieties of English than BrE does, and that these varieties are therefore more likely to display AmE features than BrE features.<sup>2</sup>

As appears from the varieties mentioned above, Mair includes both native and non-native Englishes in his model (and both 'standard' and 'non-standard' varieties). Among the native (ENL) varieties, he considers both national varieties (e.g. British English) and dialects (e.g. Scottish English). Among non-native varieties, his examination focuses on those that belong to Kachru's (1985) Outer Circle, that is, the institutionalised second-language varieties of English (ESL). He tentatively includes domain-specific uses of English as a Lingua Franca (ELF; e.g. business English) among the super-central varieties, but he neglects English as a foreign language (EFL), as represented in Kachru's (1985) Expanding Circle, despite its important place among non-native World Englishes. Schneider (2014), when revisiting another model meant to represent the use of English around the world, the Dynamic Model of the evolution of Postcolonial Englishes (Schneider 2003, 2007),

---

2. The possible influence of AmE on certain Outer Circle varieties has already been documented in some corpus studies (e.g. Hackert 2015; Deshors & Gries 2016; Deshors & Götz forthcoming), but to my knowledge this study is the first one that considers a combination of Outer Circle and Expanding Circle varieties.

underlines the increasing need to pay attention to EFL. Citing Xu (2010: 296), he refers to the “dawning of the age of expanding circle Englishes”, and goes on to investigate whether the Dynamic Model could be applied to EFL varieties. While his conclusion is that the Dynamic Model is of limited use to describe Englishes in the Expanding Circle, his paper is a plea for the inclusion of Expanding Circle varieties in theories of World Englishes. Following his recommendation, the present chapter pursues the second objective of applying Mair’s (2013) World System of Englishes to EFL data, by examining whether AmE can be said to act as a hub for the Expanding Circle too. At the same time, the inclusion of Expanding Circle varieties means that the present chapter aims to contribute to the collaborative effort to bridge the paradigm gap between contact linguistics (and its focus on institutionalised second-language varieties) and second language acquisition research (and its focus on learner varieties). Although this trend was initiated as early as 1986, in a seminal paper by Sridhar & Sridhar, it is only recently that scholars have started to use corpus data to investigate the possible links between Outer and Expanding Circle varieties (see, e.g., Mukherjee & Hundt 2011; Davydova 2012; Deshors 2014; Edwards & Laporte 2015 or Gilquin 2015). These studies have shown that the different contexts of acquisition and use of the two types of varieties can result in different linguistic patterns, but that similarities also occur, thus paving the way for a rapprochement between these varieties and between the fields of research that are associated with them.

The second hypothesis, within this framework, is that, despite a common predominance of the ‘American hub’ (see first hypothesis), the different contexts of acquisition and use in ESL and EFL countries will result in different degrees of influence of AmE. More precisely, it is expected that this influence will be stronger in ESL countries than in EFL countries. In ESL countries, English is an official or semi-official language that is used for intranational communication in settings like the administration or the media. People in these countries therefore receive English input in their everyday life, in contexts that are likely to be subject to the forces of globalisation which according to Mair (2013) are associated with the dominance of AmE. Through national TV channels or newspapers in English or through English contact with people from other ethnic groups, for example, speakers in ESL countries are part of the “mediascape” and “ethnoscape” (Appadurai 1996) that characterise our globalising world. EFL learners, on the other hand, receive limited input in English in their everyday life, since English has no official status in the Expanding Circle. Of course, most of them are still subject to the forces of globalisation associated with AmE, especially in today’s society, where the Internet has made it possible to be connected with the rest of the world without ever leaving one’s computer (in this respect, the “mediascape”, in particular, is likely to be an important setting from which learners can receive English input). However, since

EFL learners' main exposure to English is through classroom instruction, they should be less subject to the forces of globalisation than is the case in ESL countries, and more subject to the forces of education, which tend to be more conservative and more oriented towards BrE models.<sup>3</sup> As pointed out by Trudgill & Hannah (2017: 5), “[t]raditionally, schools and universities in Europe – and in many other parts of the world – have taught the variety of English which is often referred to as ‘British English’”. In this respect too, things have been changing and AmE has become a strong competitor of BrE in teaching (cf. Modiano 2007: 525–526). Trudgill & Hannah (2017: 6) themselves note that AmE is “widely taught to students of EFL and ESL”, especially in North America, Latin America and “other areas of the world”. Yet, generally speaking, it can be said that the BrE norm “is still upheld in educational institutions” (Schneider 2007: 172). In the study of the influence of AmE vs BrE on World Englishes, EFL varieties, which are highly dependent, for their development, on education, are therefore expected to show a less distinct influence of AmE as compared to ESL varieties.

It was suggested at the beginning of this section that different factors may point to different linguistic influences (see Algeo’s quotation). Following up on this idea, a third aim of this chapter is to explore the possible impact of certain factors on the presence of American vs British linguistic features. The approach here will be more local: instead of considering general differences in context between the Kachruvian Circles, a distinction will be drawn between the countries (or continents) that are included in these Circles. Three factors whose potential relevance has been underlined in the literature will be examined. The first one is the historical background of the country, and more particularly the colonial relations that may exist between the United States (US) / United Kingdom (UK) and the countries under study. Talking about Malaysia, for example, Jayapalan & Pillai (2011) note that “[a]s would be expected of a former British colony, there is a tendency to adopt a British model of pronunciation”. Lim (2012: 279) distinguishes Philippine English from other South-East Asian varieties like Singapore or Hong Kong English on the basis of a historical argument as well. She claims that in the Philippines, a former American colony, “the exonormative standard has been and still is American English”, unlike the other countries in South-East Asia, which are all former British colonies. The second factor is economic and will look into imports

---

3. Note that the situation is bound to be different for *users* of English in EFL countries (see Mauranen 2011 on the distinction between users and learners), as these are not (or no longer) exposed to ‘classroom English’ and, on the other hand, are probably more exposed to English for Occupational Purposes, which corresponds to the “domain-specific ELF uses” that Mair (2013: 264) includes among the super-central varieties of his model.

and exports between the US/UK and each of the countries. Such an economic factor is for instance put forward by Braine (2005: xvii) to justify the choice of AmE as a model in Brazil: “[b]ecause the United States is the most powerful trading partner and Brazilians need American English for business communication, the American model is prevailing now”. The third factor that will be examined is the geographical proximity between the US/UK and the different countries. This factor is mentioned by Kachru (1983: 60) to explain the use of AmE as a model in countries like Mexico or Cuba. It is also found in this quotation from Szpyra-Kozłowska (2015: 9) about EFL: “The choice of an English variety (...) for EFL learners has been dictated by geographical proximity as well as by the economic, political and cultural influences of an English-speaking country”. Interestingly, this quotation includes other factors, some of which will be considered here (economic influences) and others not, for practical reasons (cf. difficulty of quantifying influences of a political or cultural type). The specific historical, economic and geographical situation of the countries under investigation will make it possible to test the third hypothesis of this study, namely that the influence of AmE or BrE will vary according to the local context of these countries.

The three hypotheses that will be tested in this study can be summarised as follows:

- Hypothesis 1: AmE is expected to exert more influence than BrE on World Englishes;
- Hypothesis 2: AmE is expected to exert more influence on ESL varieties than on EFL varieties;
- Hypothesis 3: The degree of influence of AmE and BrE is expected to vary according to the local context.

They will be tested on the basis of twenty pairs of items distinctive of AmE vs BrE. The first and second hypotheses will be considered in Section 4.1, while the third hypothesis will be explored in Section 4.2.

### 3. Data and methodology

#### 3.1 The corpora

This study is based on the use of two large corpora, namely the *Global Web-based English Corpus* (GloWbE) and the *EF-Cambridge Open Language Database* (EFCAMDAT). GloWbE is a 1.9 billion-word corpus made up of Internet materials from twenty English-speaking countries (see Davies & Fuchs 2015a). Of these twenty countries, six



belong to the Inner Circle, while the remaining fourteen belong to the Outer Circle.<sup>4</sup> The American and British subcorpora were used to check the distinctiveness of certain linguistic items (Section 3.2). The selected items were then extracted from all the other GloWbE subcorpora through the online interface to the corpus <<http://corpus.byu.edu/glowbe>>. As for EFCAMDAT, it contains essays written by learners of English within the frame of the online school of EF Education First (Geertzen et al. 2014). It is currently made up of almost 39 million words but will continue to grow as more data are added. The data were produced by learners from different proficiency levels and 141 countries. Of these, countries from the Inner and Outer Circles were excluded, so as to keep only those data that were presumably produced by EFL learners.<sup>5</sup> The remaining data represent the production of 513,886 learners from 107 countries.<sup>6</sup> The word count of the different (sub)corpora can be found in Table 1, where ENL corresponds to the Inner Circle varieties from GloWbE with the exceptions of AmE and BrE, ESL to the Outer Circle varieties from GloWbE and EFL to the Expanding Circle varieties from EFCAMDAT.

Table 1. Size of the (sub)corpora used in the study

Variety	AmE	BrE	ENL	ESL	EFL
Corpus	GloWbE	GloWbE	GloWbE	GloWbE	EFCAMDAT
Size	386,809,355	387,615,074	465,393,257	645,815,287	32,653,692

4. The countries belonging to the Inner Circle are: Australia, Canada, Ireland, New Zealand, United Kingdom and United States; those belonging to the Outer Circle are: Bangladesh, Ghana, Hong Kong, India, Jamaica, Kenya, Malaysia, Nigeria, Pakistan, Philippines, Singapore, South Africa, Sri Lanka and Tanzania.

5. In EFCAMDAT, little information is available about learners' linguistic profiles. It was therefore assumed that learners living in a country from the Expanding Circle were EFL learners.

6. The countries, grouped by continent as indicated in EFCAMDAT, are: Africa (Algeria, Angola, Burkina Faso, Chad, Comoros, Côte d'Ivoire, Djibouti, Egypt, Gabon, Madagascar, Mali, Mauritania, Morocco, Mozambique, Niger, Senegal), Asia (Afghanistan, Armenia, Azerbaijan, Bahrain, Cambodia, Cyprus, Georgia, Indonesia, Iran, Iraq, Japan, Jordan, Kazakhstan, Kuwait, Kyrgyzstan, Laos, Lebanon, Macau, Mainland China, Mongolia, Myanmar, Nepal, Oman, Palestine, Saudi Arabia, South Korea, Syria, Taiwan, Thailand, Turkey, Turkmenistan, United Arab Emirates, Uzbekistan, Vietnam, Yemen), Europe (Albania, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Guadeloupe, Hungary, Italy, Latvia, Lithuania, Macedonia, Moldova, Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine), North America (Aruba, Costa Rica, Cuba, Dominican Republic, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama), Oceania (French Polynesia, New Caledonia, Wallis and Futuna) and South America (Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, Venezuela).

While other corpora have typically been used to study ESL and EFL – the *International Corpus of English* (ICE) and the *International Corpus of Learner English* (ICLE) immediately spring to mind – GloWbE and EFCAMDAT present some advantages in comparison with these other corpora that are particularly important for the present purposes. The first one has to do with size. While each ICE subcorpus is made up of about one million words, in GloWbE the smallest subcorpus, the Tanzanian one, contains over 35 million words. The EFCAMDAT subcorpora are smaller in comparison (the largest one, the Brazilian component, is about 10 million words long), but with its grand total of over 33 million words, EFCAMDAT is almost ten times as large as the second version of ICLE. This, it should be noted, also comes with its downside: it is easy to extract huge quantities of data, but the sheer quantity of data means that manual disambiguation is simply impossible and that certain types of searches should therefore be avoided (see Section 3.2). A second advantage of GloWbE and EFCAMDAT is that they include relatively recent data. While no exact dates are provided in the documentation, we know that their collection started only a few years ago (the reference publications, mentioned in the preceding paragraph, date back to 2015 and 2014, respectively) and that their contents are unlikely to be very old. In ICE and ICLE, on the other hand, next to relatively recent subcorpora, certain components were collected at the beginning of these projects, that is, around 1990. As regards the geographical origin of the data, both GloWbE and EFCAMDAT are more varied than their counterparts ICE and ICLE (more ICE and ICLE subcorpora are in preparation, though). EFCAMDAT is also more varied than ICLE in terms of proficiency levels (ICLE includes data produced by higher-intermediate to advanced learners, whereas EFCAMDAT covers the whole range of proficiency levels, from beginners to advanced learners). Finally, both GloWbE and EFCAMDAT can be queried through a freely available online interface which also includes facilities for POS-tag-based searches.

However, both corpora also present some disadvantages (see Mair et al. 2015 on the limitations of GloWbE). The first one is that the metadata are relatively limited. In particular, very little is known about the authors of the texts included in the two corpora. For EFCAMDAT, this means, for example, that the learners' nationalities are known, but not their mother tongues, although the latter are more important than the former linguistically speaking. For GloWbE, this lack of metadata does not even make it possible to “claim that the speakers are actually speakers of the dialect in question” (Davies & Fuchs 2015b: 46). The contributors to ICE and ICLE, by contrast, have been selected more carefully and rich metadata about them can be found, at least in the case of ICLE. Another disadvantage of GloWbE and EFCAMDAT has to do with the text types that are represented. Both corpora are quite limited in this respect, including web-based materials and

written assignments, respectively.<sup>7</sup> In comparison, ICLE is equally limited, being made up of argumentative and some literary essays, but ICE includes a large variety of written and spoken genres. In addition, it must be underlined that the (rather informal) web-based materials included in GloWbE and the (more formal) written assignments included in EFCAMDAT are stylistically quite different, which may have an impact on the ESL-EFL comparison – a limitation that should be borne in mind when considering the results of the analysis. For the present purposes, the advantages of using GloWbE and EFCAMDAT were seen to outweigh the possible limitations of the two corpora (such as the unspecificity of the contributors or the impossibility of manually handling the large quantities of data),<sup>8</sup> as well as the advantages of other, more controlled corpora (e.g. the wide stylistic range of texts included in ICE or the rich metadata found in ICLE).

### 3.2 Selection of linguistic features

In an attempt to assess the influence of AmE as opposed to BrE on World Englishes, the first step of the analysis was to identify linguistic features that make it possible to distinguish between the two native varieties. These features were taken from Algeo's (2006) list of differences between AmE and BrE. Algeo's list is mainly intuition-based, but the claims are supported by corpus data and/or scholarly works when relevant. In my selection, I disregarded any claims that were wholly intuition-based (e.g. *go-slow* vs *slowdown* on p. 71). I also set aside items that were described as rare (e.g. *cellar wine* on p. 12), those that showed the same tendency in AmE and BrE but with different proportions (cf. *burned* vs *burnt*: "although both national varieties prefer the regular form, the American preference for it is significantly stronger", p. 13), as well as those that did not have an equivalent in the other variety or whose equivalent was not made explicit in the entry (e.g. *motorway* on p. 80). Finally, some items had to be ignored for practical reasons as they would not have been easily retrievable from the corpora or would have implied manual weeding out which was not possible given the amount of data involved (e.g. personal object with *pressure/pressurize* on p. 12; functional uses of tenses on p. 24ff.).

---

7. Although the EFCAMDAT website refers to "samples of spoken and written language production", the manual available online only describes written data, and the output of corpus queries seems to be limited to writing.

8. In fact, even the automatic processing of the data may prove difficult at times. Thus, a program like *WordSmith Tools* seems to have some difficulties dealing with the huge files of GloWbE, for example to do a keyword analysis based on word clusters.

The selected pairs of items were then tested as to their ability to actually distinguish between AmE and BrE. This was done by comparing the frequency of the items in the American and British components of GloWbE.<sup>9</sup> In order for their distinctive nature to be confirmed, the items had to display reverse preferences and the difference in frequency had to be statistically significant (with  $p < 0.0001$ ). Thus, with a relative frequency of 7.35 per million words (pmw) in AmE and 32.69 pmw in BrE, *HAVE got* displays preferences that are opposite to those of *HAVE gotten*, which has a relative frequency of 15.73 pmw in AmE and 4.23 pmw in BrE. In addition, a chi-square test reveals that the difference between the absolute frequency of *HAVE got* in AmE and BrE is statistically significant ( $X^2 = 2,563.83$ ;  $p < 0.0001$ ), as is the difference for *HAVE gotten* ( $X^2 = 6,209.05$ ;  $p < 0.0001$ ). The last condition for the items to be included in the study was that they had to be sufficiently frequent in EFCAMDAT, so that reliable claims could be made about the EFL data as well. The minimum threshold was an absolute frequency of 50 for at least one of the two items. A pair like *then and there / there and then*, for example, despite its significant distinctiveness between the American and British components of GloWbE, was excluded because of its very low frequency in EFCAMDAT (3 and 6 occurrences, respectively).

Table 2 shows the twenty selected pairs of items, classified according to Algeo's (2006) categories.<sup>10</sup> Note that while it was not possible to select items from each of the categories, as they did not necessarily meet all of the criteria described above, the selected items can be said to represent a relatively wide range of word classes and phenomena, not being limited to the traditional lexical pairs illustrated by *movie/film* and *apartment/flat*.<sup>11</sup> When extracting the items from the corpora, lemmas were searched for (e.g. *HAVE* in *HAVE got(ten)*) and POS tags were added (e.g. *FLAT* as a noun) whenever necessary. Table 3 gives an overview of the global (absolute and relative) frequencies of the selected items in the different varieties (AmE, BrE, native English to the exclusion of AmE/BrE, ESL and EFL).

---

9. GloWbE was chosen as a basis to test the distinctiveness of the items, rather than general datasets like the *British National Corpus* (BNC) or the *Corpus of Contemporary American English* (COCA), because of the comparability of the GloWbE components (in terms of register, time of collection, etc.)

10. Originally, the pairs *movie/film* and *apartment/flat* belonged to the 'preposition' category and took the form of *in the movies / on (the) film(s)* and *into an apartment / onto a flat*, respectively.

11. Although Algeo (2006) includes pairs of items that differ in spelling (e.g. *catalog vs catalogue*), such cases were excluded from the selection because spelling differences have already been investigated in the literature (see Larsson 2012 on EFL varieties and Reuter 2016 on ESL varieties) and because Mukherjee in Mair et al. (2015) suggests that the distribution of AmE vs BrE spelling variants in GloWbE may be biased by the composition of the corpus.

Table 2. Selected distinctive pairs of items<sup>12</sup>

Category	AmE	BrE
Verbs	<i>HAVE gotten</i>	<i>HAVE got</i>
Determiners	<i>a half</i>	<i>half a(n)</i>
Nouns	<i>math</i>	<i>maths</i>
	<i>MOVIE</i>	<i>FILM</i>
	<i>APARTMENT</i>	<i>FLAT</i>
Adjectives	<i>volunteer work</i>	<i>voluntary work</i>
	<i>all year long</i>	<i>all year round</i>
	<i>free time</i>	<i>spare time</i>
Adverbs	<i>right away</i>	<i>straight away</i>
Prepositions	<i>in college</i>	<i>at (the) college</i>
	<i>right now</i>	<i>at the moment</i>
	<i>in school</i>	<i>at school</i>
	<i>on (the) WEEKEND</i>	<i>at (the) WEEKEND</i>
	<i>toward</i>	<i>towards</i>
Complementation	<i>CHAT with</i>	<i>CHAT to</i>
	<i>different than</i>	<i>different to</i>
Expanded constructions	<i>MAKE a deal</i>	<i>DO a deal</i>
	<i>GIVE it a try</i>	<i>GIVE it a go</i>
	<i>TAKE a vacation</i>	<i>HAVE a holiday</i>
	<i>TAKE a shower</i>	<i>HAVE a shower</i>

Table 3. Absolute and relative frequencies pmw of the selected items

Variety	AmE	BrE	ENL	ESL	EFL
Abs. freq.	405,937	418,940	437,680	607,225	45,591
Rel. freq.	1,049.45	1,080.81	940.45	940.25	1,396.20

### 3.3 Americanness and Britishness

In order to measure the degree of influence of AmE and BrE on the other varieties of English, the rate of ‘Americanness’ and ‘Britishness’ was calculated (see Reuter 2016). The calculation is based on the relative frequency of the two members of the pairs of distinctive features listed in Table 2. The rate of Americanness corresponds to the ratio of the relative frequency of the AmE item out of the combined relative frequency of the AmE and BrE items, while the rate of Britishness corresponds

12. Capital letters indicate that the lemma was searched for. Both the GloWbE and EFCAMDAT interfaces make the search for lemmas possible.

to the ratio of the relative frequency of the BrE item out of the combined relative frequency of the AmE and BrE items. For example, *HAVE gotten* has a relative frequency of 85.44 instances pmw in the ESL data, whereas *HAVE got* has a relative frequency of 229.47. Their combined relative frequency thus equals 314.91. The Americanness rate is calculated by dividing the relative frequency of *HAVE gotten* (the AmE item), i.e. 85.44, by the combined relative frequency, i.e. 314.91. The result, multiplied by 100, yields a percentage of 27.13%. The Britishness rate is calculated by dividing the relative frequency of *HAVE got* (the BrE item), i.e. 229.47, by the combined relative frequency, i.e. 314.91, which yields a percentage of 72.87%. Since the rates of Americanness and Britishness always vary in relation to each other (together, they have to make up a total of 100%), only the Americanness rate will be mentioned when giving the results. This choice reflects the point of departure of this chapter, which is Mair's (2013) claim that AmE is the hub of the World System of Englishes.

#### 4. Assessing the influence of AmE and BrE

##### 4.1 ENL/ESL/EFL

The first step of the analysis consisted in determining the frequency of the AmE and BrE items in countries from the Inner Circle – excluding the US and the UK – (ENL), countries from the Outer Circle (ESL) and countries from the Expanding Circle (EFL). Table 4 lists the twenty pairs of items, together with their Americanness rates in these three types of varieties. What the average results show is that, except in the ENL countries, where the Americanness rate is slightly below 50%, the influence of AmE is more marked than that of BrE, with an Americanness rate of 58.35% in the ESL countries and 63.18% in the EFL countries. These preferences can be illustrated by the following examples, representing ENL, ESL and EFL respectively:

- (1) *So if living in the States [is] on your bucket list, I encourage you to give it a go* [BrE]. (GloWbE-Ireland)
- (2) *I strongly encourage all of you who are thinking of being doctors to give it a try* [AmE]. (GloWbE-Singapore)
- (3) *I know it is quite unsettling and totally challenging to change a job, but I really want to encourage you to give it a try* [AmE]. (EFCAMDAT-Netherlands)

If we bring together the results for the three types of varieties, we obtain an average Americanness rate of 56.09%, which represents a slight tendency towards an American hub, as predicted by Mair (2013). If, on the other hand, we take

into account the types of countries that are explicitly and unequivocally included in Mair's model, that is, ENL and ESL countries, we obtain an even lower Americanness rate of 52.55%, which points to an American influence that is still dominant, but far from overwhelming in comparison with BrE. Of the three types of varieties, it is EFL that displays the highest average Americanness rate, which suggests that Mair's model also applies to EFL countries, at least as far as the influence of AmE is concerned.

**Table 4.** Americanness rate per type of variety (the rates below 50% are shown in bold)

AmE/BrE pair	ENL	ESL	EFL
<i>HAVE gotten/HAVE got</i>	<b>28.09</b>	<b>27.13</b>	<b>6.19</b>
<i>a half/half a(n)</i>	55.85	50.72	65.54
<i>math/maths</i>	<b>49.27</b>	67.28	79.44
MOVIE/FILM	<b>36.25</b>	<b>44.93</b>	92.39
APARTMENT/FLAT	66.67	55.50	92.75
<i>volunteer work/voluntary work</i>	53.77	64.84	78.03
<i>all year long/all year round</i>	<b>19.27</b>	<b>17.81</b>	<b>21.74</b>
<i>free time/spare time</i>	52.65	62.80	84.28
<i>right away/straight away</i>	51.38	75.94	95.36
<i>in college/at (the) college</i>	67.13	68.57	69.60
<i>right now/at the moment</i>	58.04	66.91	65.38
<i>in school/at school</i>	<b>41.43</b>	62.28	<b>48.09</b>
<i>on (the) WEEKEND/at (the) WEEKEND</i>	58.25	57.49	72.83
<i>toward/towards</i>	<b>23.51</b>	<b>18.96</b>	<b>22.95</b>
<i>CHAT with/CHAT to</i>	55.60	72.17	96.91
<i>different than/different to</i>	<b>32.52</b>	<b>49.56</b>	<b>19.17</b>
<i>MAKE a deal/DO a deal</i>	60.34	80.61	96.55
<i>GIVE it a try/GIVE it a go</i>	<b>40.11</b>	76.73	98.44
<i>TAKE a vacation/HAVE a holiday</i>	<b>42.47</b>	68.23	<b>35.87</b>
<i>TAKE a shower/HAVE a shower</i>	<b>42.34</b>	78.57	<b>22.13</b>
AVERAGE	<b>46.75</b>	58.35	63.18

It also appears from Table 4 that not all pairs of items display the same tendency. In ENL countries, in accordance with the average Americanness rate close to 50%, half of the pairs show a stronger American influence, while the remaining ten items show a stronger British influence (i.e. an Americanness rate below 50%; see numbers in bold in the table). In ESL countries, only five pairs (that is, a quarter of the items) show a stronger British influence, whereas in EFL countries this is the case for seven pairs of items. That the number of pairs with a stronger British influence is larger in EFL than in ESL, despite an average Americanness rate that is higher in EFL than in ESL, suggests that the disparity between the different pairs of items is wider in EFL than in ESL. This is confirmed by Figure 1, a boxplot of

Americanness rate per type of variety (ENL, ESL, EFL). The central boxes correspond to the interquartile range, that is, the middle 50% of the data, while the bold-typed horizontal line in each box is the median. The lower and upper whiskers indicate the minimum and maximum observations respectively, and each of them represents 25% of the data. It appears from the boxplot that, whereas ENL and ESL display the same sort of dispersion (except for the presence of two outliers in ESL, indicated by circles), EFL shows much more variation, covering almost the whole scale, from an Americanness rate of 6.19% to one of 98.44%. This difference suggests that ENL and ESL writers are more coherent than EFL writers in their use of AmE. They are less influenced by AmE overall, but for all the linguistic items studied here, they are influenced to more or less the same extent. EFL learners, by contrast, are influenced by AmE to various degrees, depending on the linguistic item that is examined (compare examples (4) and (5), both taken from the same component of EFCAMDAT). It would therefore perhaps be more appropriate to refer to a preference for certain individual AmE words or constructions in EFL, as opposed to the general, across-the-board influence of AmE that is visible in ESL (and, to a lesser extent, in ENL).

- (4) *This was totally **different to** [BrE] what they promised us.* (EFCAMDAT-Mexico)  
 (5) *I like going to the market and **chat with** [AmE] my friends.*  
 (EFCAMDAT-Mexico)

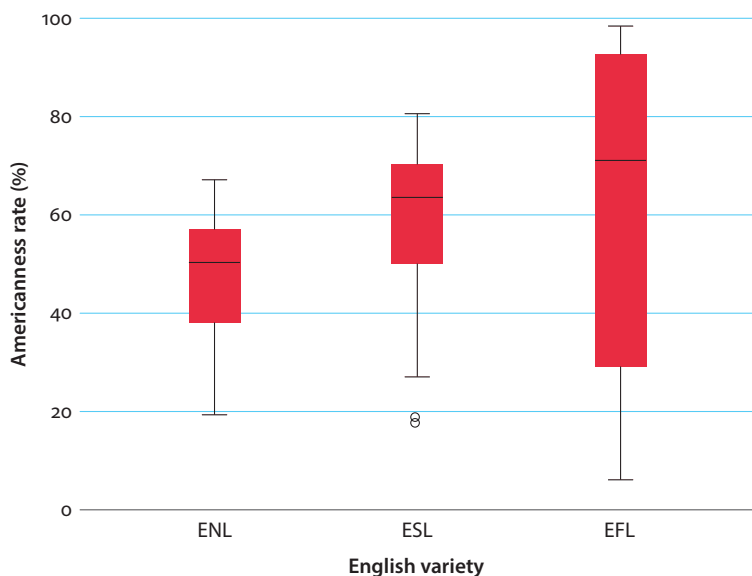


Figure 1. Boxplot of Americanness rate per type of variety



On average, the pairs of items that have the highest Americanness rate across types of varieties are *MAKE a deal/DO a deal* (79.17%), *CHAT with/to* (74.89%) and *right/straight away* (74.23%). Those that have the lowest Americanness rate are *all year long/round* (19.61%), *HAVE gotten/got* (20.47%) and *toward/towards* (21.81%). However, these average figures hide some variation between the three types of varieties. While ten pairs of words/constructions show a shared preference for the AmE item (e.g. *in college*, *MAKE a deal*) and four show a shared preference for the BrE item (e.g. *all year round*), the remaining six pairs display different tendencies. Among these, three show the same preference for the BrE item in ENL and EFL (*at school*, *HAVE a holiday*, *HAVE a shower*), two for the AmE item in ESL and EFL (*math*, *GIVE it a try*), and in the last pair the BrE item *film* is favoured in both ENL and ESL. It thus turns out that EFL is sometimes more similar to ENL and sometimes more similar to ESL, which might indicate a lack of stability of this type of variety. This is also apparent if we consider the categories, as distinguished by Algeo (2006), to which the items belong. Expanded constructions are a case in point. While they are mostly influenced by BrE in ENL (with the exception of *MAKE a deal/DO a deal*, which has an Americanness rate of 60%) and by AmE in ESL (with a minimum Americanness rate of 68% and three out of the four constructions representing the top three in terms of Americanness rate), in EFL two of the expanded constructions show a preference for the BrE option (*HAVE a shower* and *HAVE a holiday*) and the other two constructions are more distinctively American (*MAKE a deal* and *GIVE it a try*).

The above results reveal a mixed picture as regards the influence of AmE. All three types of varieties show *some* influence of AmE in that (i) for each pair of items, a certain proportion of the uses are distinctively American, and (ii) certain pairs of items display a stronger preference for the AmE item than for the BrE item. However, the results also underline the variation, not only among the pairs of items, but also among the types of varieties. More precisely, the ENL varieties appear to be less strongly influenced by AmE than the ESL and EFL varieties, which both have an average Americanness rate higher than 50%. Of these two, it is the EFL varieties that have the highest Americanness rate. The first hypothesis put forward in Section 2 (“AmE is expected to exert more influence than BrE on World Englishes”) is thus only partly confirmed. Overall, AmE is slightly more influential than BrE (the average Americanness rate is 52.55% if we take into account the types of varieties clearly included in Mair’s (2013) World System of Englishes, and 56.09% if we take all the types of varieties into account). If we distinguish between the three types of varieties, however, only the non-native varieties (ESL and EFL) show a predominantly American influence; for the ENL varieties, it is the British influence that predominates. As for the second hypothesis (“AmE is expected to exert more influence on ESL varieties than on EFL varieties”), it is not confirmed.

While ESL and EFL display different degrees of AmE influence, this difference is not as expected, since it is actually the EFL varieties that have the highest Americanness rate overall. This result is qualified by the finding that EFL is characterised by a great deal of variation between the different pairs of items, some of which show an Americanness rate that is lower than the rates found in the other two types of varieties (cf. *HAVE gotten/HAVE got* and its 6.19% in EFL, as against 27–28% in ESL and ENL). More generally, the strong AmE influence in EFL suggests that these varieties can have their place in Mair's (2013) model, which could thus be expanded to include the Expanding Circle.

In what precedes, a broad distinction has been drawn between ENL, ESL and EFL varieties, corresponding to the Inner, Outer and Expanding Circles. The situation within each of these Circles with regard to AmE influence, however, may not be homogeneous, as Kachru's (1985) model groups together countries that are geographically spread out and may differ in their (economic, political, historical, cultural, etc.) relations with the US. The next section will consider the possible variation within ENL, ESL and EFL and will try to link any differences to the historical, geographical and economic context of the countries or continents under study.

## 4.2 Countries and continents

In order to approach the possible influence of AmE and BrE more locally, a distinction was drawn between the different varieties that are included within ENL, ESL and EFL. In GloWbE, ENL is made up of four national varieties and ESL of fourteen (see Table 5). As for EFCAMDAT, it includes data produced by learners from 107 different countries. As distinguishing between each of these 107 countries would have resulted in too small sets of data and potentially insignificant differences, it was decided to consider the higher level of nationalities within the EFCAMDAT interface, namely that of continents.<sup>13</sup> Six continents are distinguished in the EFCAMDAT interface: Africa, Asia, Europe, North America, Oceania and South America. Oceania was however excluded because a mere three pairs of items among the twenty investigated were represented in this subcorpus. The results for Africa are based on fifteen pairs of items, those for North America on nineteen pairs of items and those for the other continents on the full set of items. Only the average Americanness rates for the twenty pairs of items will be presented here.

---

13. While working with continents rather than individual countries necessarily results in a certain degree of approximation, it makes sense at least with respect to the geographical factor, according to which the degree of influence of AmE vs BrE is likely to be the same for countries that are close to each other.

Table 5. Average Americanness rate per variety (the rates below 50% are shown in bold)<sup>14</sup>

ENL		ESL		EFL	
IE	<b>36.95</b>	ZA	<b>47.39</b>	EU	58.26
NZ	<b>37.51</b>	LK	50.56	AS	60.81
AU	<b>39.11</b>	TZ	50.80	AF	61.91
CA	73.23	HK	57.46	NA	65.84
		GH	58.40	SA	70.90
		PK	58.47		
		NG	58.81		
		IN	58.91		
		KE	60.08		
		MY	60.99		
		BD	61.31		
		SG	62.51		
		JM	66.05		
		PH	72.16		

Table 5 provides the average Americanness rate per variety. Within ENL, we can notice a sharp divide between Irish, New Zealand and Australian English on the one hand and Canadian English on the other hand. While the former are mainly influenced by BrE, with an Americanness rate ranging from 37% to 39%, the latter is predominantly influenced by AmE (Americanness rate = 73%). The finding that ENL is more oriented towards the BrE model (see Section 4.1) is thus valid for certain ENL varieties, but clearly not for Canadian English. Within ESL, almost all of the national varieties display the same general tendency as highlighted in Section 4.1, namely a stronger influence of AmE than of BrE. The only real exception is South African English, whose Americanness rate is under 50%. Sri Lankan English and Tanzanian English are only slightly above the 50% threshold, which suggests that they are equally influenced by AmE and BrE. All the other ESL varieties appear to be more influenced by AmE, with Americanness rates varying between 57% (for Hong Kong English) and 72% (for Philippine English). The EFL varieties, finally, are all characterised by an Americanness rate above 50%, ranging from 58% in Europe to 71% in South America.

In order to tentatively explain the variation found within ENL, ESL and EFL, a number of local contextual factors were examined which could potentially affect

14. The labels in the table refer to the following varieties: (for ENL) AU = Australia, CA = Canada, IE = Ireland, NZ = New Zealand; (for ESL) BD = Bangladesh, GH = Ghana, HK = Hong Kong, IN = India, JM = Jamaica, KE = Kenya, LK = Sri Lanka, MY = Malaysia, NG = Nigeria, PH = Philippines, PK = Pakistan, SG = Singapore, TZ = Tanzania, ZA = South Africa; (for EFL) AF = Africa, AS = Asia, EU = Europe, NA = North America, SA = South America.

the relative strength of the influence of AmE or BrE, and for which precise and accurate information could easily be gathered. The factors are as follows: historical context (do the countries under investigation share some colonial history with the US or the UK?), economic relations (do the countries under investigation import and/or export more goods with the US or the UK?) and geographical proximity (are the countries under investigation closer to the US or the UK?). Because these three factors only represent a small selection of all the contextual factors that could influence the orientation towards BrE or AmE, and because they may not even be the most important ones in this respect,<sup>15</sup> this analysis will mainly be exploratory and should only be seen as a first attempt to determine whether norm orientation may also vary according to local contextual factors, in addition to the very general factors considered in Section 4.1 (native vs non-native varieties or ESL vs EFL). Only some of the most interesting findings will be presented here, but detailed tables for each of the three local factors examined can be found in the Appendix.

Starting with the historical factor, all the ENL countries represented in GloWbE have a shared colonial past with the UK, which might explain the overall BrE influence on ENL (Section 4.1). Canada is an exception among the ENL countries, in that it displays a predominantly American influence (73.23%) despite being historically linked to the UK. Of all the ESL countries included in GloWbE, only one is a former American colony, namely the Philippines. History is presumably an important factor in this case, since Philippine English is the ESL variety that has the highest Americanness rate (72.16%). It should however be noted that the former British colonies of the Outer Circle (with the exceptions of South Africa, Sri Lanka and Tanzania) are also more distinctively oriented towards AmE, though with lower rates than is the case in Philippine English. As for the EFL countries, not all of them share a colonial history with either the US or the UK, but among those that were former American colonies, most are located in North America, which could partly explain why North America comes second in terms of Americanness rate (65.84%). The highest Americanness rate (70.90%) is found in South America, although its shared colonial history with the US is limited to Colombia. African, Asian and European countries are, for the most part, not historically connected to the US (Japan is an exception), and several Asian and African countries are in fact former British colonies. This could account for the lower Americanness rate in these countries (around 60%).

---

15. Factors such as media imports or educational systems may be more influential than history, geography or global economy to establish norm orientation, as pointed out by one of the reviewers, but they are more difficult to quantify precisely, especially for individual countries. A more qualitative approach to the issue, as in Hänsel & Deuber (2013), would be valid as well, but not feasible here given the large number of countries investigated.

In terms of economy, all of the ENL and ESL countries and most of the EFL countries considered in this study have more economic contacts with the US than with the UK, which could be related to the overall preference for an American norm (see Section 4.1). In the case of the ESL countries, these economic relations with the US go hand in hand with a preference for AmE (with the exceptions, again, of South Africa, Sri Lanka and Tanzania).<sup>16</sup> As regards EFL, it is interesting to note that the only countries (worldwide) that have closer economic ties with the UK than with the US are a handful of countries in Europe and one in Asia (Cyprus, which some would consider as belonging to Europe rather than Asia). It seems as if this special status on the world map might have some linguistic consequences, since Europe is the EFL continent that shows the lowest Americanness rate (58.26%), followed by Asia (60.81%). For the ENL countries, finally, the economic relations with the US are hardly reflected in the norm orientation since, with the exception of Canada and its Americanness rate of 73.23%, the other countries (Ireland, New Zealand and Australia) all have an Americanness rate below 40%.

The last local contextual factor whose potential impact on norm orientation was examined is geographical proximity. Among the ENL countries, this factor could explain why Canada is predominantly influenced by AmE, despite being historically linked to the UK. It could also account for the predominantly British influence in Ireland (in this case, together with – rather than despite – its shared colonial history with the UK). Australia and New Zealand are geographically closer to the US than to the UK, which does not correspond to their main attraction towards BrE. However, the difference in distance between Australia-US and Australia-UK is extremely small (less than 22 kilometres), and the status of Australia and New Zealand as islands could be argued to neutralise, to some extent, the effect of geographical proximity. Turning to the ESL countries, Jamaica has this particular feature that it is the only country that is geographically closer to the US than to the UK. This feature seems to be reflected in the Americanness rate of Jamaican English (66.05%), which is the second highest rate for ESL. Several ESL countries, however, are closer to the UK but more attracted to AmE (e.g. Bangladesh with an Americanness rate of 61.31%, Singapore with an Americanness rate of 62.51%, or even the Philippines, which has the highest Americanness rate of all ESL countries, with 72.16%). EFL countries, finally, show quite a good correspondence between norm orientation

---

16. Interestingly, of these three exceptions, two are located in Africa, which could be related to Hänsel & Deuber's (2013: 347) finding for Kenya that, as an African country "almost completely" left out of the North America/European Union/East Asia triad affected by globalisation, it is less influenced by AmE than the other varieties considered in their study (Singapore and Trinidad & Tobago). Note, however, that some African countries in the present study display a clearer influence of AmE, including Kenya.

and geographical proximity. The continents with the lowest Americanness rates (Europe, Asia and Africa) include countries that are closer to the UK (with the exception of Guadeloupe), while the continents with the highest Americanness rates (North America and South America) include countries that are all closer to the US.

With respect to the third hypothesis put forward in Section 2 (“The degree of influence of AmE and BrE is expected to vary according to the local context”), two main observations can be made on the basis of the above results. The first one is that local variation is indeed visible. Countries within each of the Kachruvian Circles can present differing rates of Americanness or Britishness, to the extent that certain countries may be predominantly influenced by AmE and others by BrE (compare, for example, Canada and Ireland in the Inner Circle). The second observation is that the local context seems to play a certain role in the degree of influence of AmE vs BrE, with factors like colonial history, economic relations and geographical proximity potentially explaining some of the preferences for an AmE or BrE model. Yet, the link between these factors and the preferred linguistic model is far from being systematic. The economic power of the US worldwide probably contributes to the considerable influence of AmE on most of the varieties considered here (only four countries display an Americanness rate lower than the 50% threshold), which might also explain why the few countries that have more economic relations with the UK than with the US belong to the area (EU) with the lowest Americanness rate in its category (EFL). However, countries with the same profile in terms of colonial history, economic relations and geographical proximity may also show different linguistic influences (compare, for example, South Africa and Singapore, which are both former British colonies, are geographically closer to the UK and have stronger economic relations with the US, but which differ in their degree of attraction towards AmE). This suggests that other factors have a role to play in the choice of a linguistic model and/or that the factors should be examined at a finer level of granularity. Among the additional factors to be considered, we have already mentioned educational or cultural factors. We could also add the evolutionary status of the ESL varieties (Singapore English, for instance, is considered to be slightly more advanced in Schneider’s (2003, 2007) phase of endonormative stabilisation than South African English) and the proficiency level of the EFL learners (a variable that could be investigated on the basis of EFCAMDAT). Examining contextual factors at a finer level of granularity could involve taking into account the duration of colonial rule, the types of goods imported and exported (e.g. media), or the exact distance separating two countries.

## 5. Zooming in on EFL varieties

In this final section before we turn to some conclusions, I would like to zoom in on the type of varieties not included in Mair's (2013) model, namely the EFL varieties. While it was hypothesised in Section 2 that EFL varieties would be less influenced by AmE than ESL varieties, on the grounds that they would be less subject to the forces of globalisation associated with AmE and more subject to the forces of education associated with BrE, the results of the corpus analysis revealed that the average Americanness rate was in fact higher in EFL (63%) than in ESL (58%). This probably reflects two opposing trends. The first one is that the preferred model in the EFL classroom may no longer be BrE, and that AmE may have become a strong competitor, especially in certain parts of the world. Talking about pronunciation, Collins & Mees (2013: 7) argue that "General American is also used as a model by millions of students learning English as a second language – notably in Latin America and Japan, but nowadays increasingly elsewhere".<sup>17</sup> The second trend is that the traditional view of EFL learners in the Expanding Circle as only getting exposed to the English language in the classroom, through instruction, no longer corresponds to the reality of most EFL learners. Because of the global role of English in today's world, it has become almost impossible for EFL learners not to get any exposure to English outside the classroom, be it through popular music, TV series or web content. The EFL varieties may therefore be subject to the forces of globalisation and the resulting influence of AmE, like the ESL varieties. Unlike the ESL varieties, however, in EFL these forces may not be counterbalanced by nativisation forces, through which ESL speakers may want to express their local identity when using English (see Schneider 2003, 2007 on the notion of nativisation). This could perhaps account for the higher Americanness rate in EFL than in ESL. That these forces of globalisation have only really begun to affect EFL populations recently, however, appears from a comparison between the EFCAMDAT data and data from ICLE, which started to be collected earlier than EFCAMDAT. On average, the Americanness rate in ICLE for the 20 selected pairs of items amounts to 49%, as against 63% in EFCAMDAT. This is lower than any of the EFCAMDAT areas considered separately (see Table 5 in Section 4.2) and lower than the average rate for GloWbE-ESL (58%). In ICLE, eleven pairs of items show a preference for the BrE word/construction, as opposed to only seven in EFCAMDAT. For some pairs of items, the difference between EFCAMDAT and ICLE is such that the AmE option is widely preferred in the former but dispreferred in the latter (e.g. *MOVIE*

---

17. As appears from the areas mentioned as examples in this quotation, Collins and Mees (2013) do not use the term "second language" in the sense of Outer Circle variety, but in the sense of an L2, acquired later than the mother tongue.



is preferred with an Americanness rate of 92% in EFCAMDAT, but dispreferred with an Americanness rate of 32% in ICLE). On the whole, the less recent data of ICLE thus display a slight preference for the BrE model and the more recent data of EFCAMDAT a preference for the AmE model, which suggests that a process of Americanisation may have taken place in EFL over the last few years, as was shown to be the case earlier in native English (cf. Leech et al. 2009: 252ff.) and some ESL varieties (e.g. Caribbean English, see Hackert & Deuber 2015).

In addition to the stronger AmE influence in EFL than in ESL, the corpus analysis also suggested that EFL may not be influenced by AmE in the same way as ESL (and ENL). While the latter can be said to display a relatively stable kind of influence across the different pairs of items considered, EFL is characterised by a considerable disparity between certain pairs of items and others. This might be an indication that something different is going on in EFL and in ESL. More precisely, the results seem to point to some “polarizing effects” (Laitinen 2016: 187) in EFL and, in effect, a preference for certain individual AmE words/constructions rather than a true phenomenon of Americanisation affecting EFL across the board. The preference for these AmE words/constructions may be related to the kind of exposure that EFL learners receive. Although access to English has become more widespread in EFL countries, as noted above, it tends to be limited to certain domains, like those of entertainment or technology, as opposed to ESL varieties that use English in a wider range of functional domains. It is thus probably not a coincidence that among those items with the strongest Americanness rate in EFL, we find *CHAT with* (97%), which nowadays is often used in the context of internet communication, and *MOVIE* (92%), which is related to the entertainment industry. Another possible explanation for the item-based preferences found in EFL, which is not incompatible with the previous explanation, is that the preference for these words is in fact not (only) a preference for a word/construction that is typical of AmE, but a preference for a word/construction that is somehow easier for learners to acquire and remember. Arguments to support this view can be offered for several of the items with a high Americanness rate in EFL. Thus, *a half* (66%) corresponds to the usual word order of a determiner followed by an adjective (and a noun), as in *a big (house)*, unlike the use of *half a(n)*. *Math* (79%) is morphologically simpler than its BrE counterpart *maths*. In *free time* (84%), the adjective *free* is more basic and more frequent than the adjective *spare* (the adjective *free* has a relative frequency of 197.66 pmw in the BNC, as against 19.06 pmw for the adjective *spare*). *APARTMENT* (93%) will be easier to remember for learners who have a Romance mother tongue because of its similarity with the L1 equivalent (cf. French *appartement*, Italian *appartamento*, Spanish *apartamento*), whereas *FLAT* might be confusing because of its homonymy with the adjective *flat*. As for *GIVE it a try* (98%), it is undoubtedly more transparent than its British counterpart *GIVE it a go*. Of



course, EFL learners need to first get exposed to these AmE words or constructions before they can actually start using them, but because they end up preferring these AmE items does not mean that they have been exposed to them more often than their BrE counterparts and that they have been more influenced by AmE than by BrE. It might simply be that among two options, and for reasons that could have more to do with degree of complexity than with forces of globalisation, the EFL learners happened to prefer the option that turned out to be characteristic of AmE. Interestingly, a test carried out among 130 Belgian (French- and Dutch-speaking) Bachelor students specialising in English revealed that they were not always aware of the origin (AmE or BrE) of a selection of items presented to them and typical of one variety or the other.<sup>18</sup> Scores varied from 82% (for the word *movie*) to 35% (for *give it a try*). For a majority of the items, the average score was between 50% and 60%. This indicates that EFL learners are unlikely to consciously choose an AmE item because they know it is American. If American influence does occur for certain items (possibly in combination with other factors like preference for a more transparent or morphologically simpler form), this will most probably be without the learner being aware of any such influence. In other words, it does not seem to be the case that learners have two distinct registers, one British oriented and the other American oriented, and that they deliberately pick from one register or the other. As is arguably the case with written vs spoken registers (see Gilquin 2008: 128), it is more likely that they have a single register which consists in an opportunistic combination of AmE and BrE features and that, in this collection of items, it is sometimes the AmE item that is more salient and sometimes the BrE item.<sup>19</sup> This can be illustrated by examples such as (6) and (7), which combine features of AmE and BrE within the same text or even within the same sentence, or by constructions like *HAVE a vacation* (116 occurrences in EFCAMDAT) and *TAKE a holiday* (31 occurrences), which mix the verb of one variety with the noun of the other variety, as shown in (8). This phenomenon of combining AmE and BrE features could explain the considerable variation between pairs of items that was observed in the corpus-based analysis (see Section 4.1). It could also be related to one of the factors that was mentioned in Section 4.2, namely proficiency level: if the learners have not yet internalised consistent distinctions between AmE and BrE, their usage is likely to be less homogeneous than that of users whose knowledge of the language includes a full awareness of such distinctions.<sup>20</sup>

---

18. Vine (1999) makes a similar remark about New Zealanders. See also Hundt (this volume).

19. Modiano (1996) refers to this combination of AmE and BrE language traits as “Mid-Atlantic English”.

20. I thank an anonymous reviewer for bringing this point to my attention.

- (6) *The thief ran towards [BrE] me. He looked dangerous and was holding a gun in his hand. I jumped to side and he was run away. My **apartment** [AmE] was a mess.* (EFCAMDAT-Germany)
- (7) *In **spare time** [BrE] I like reading, listening music, watching **movies** [AmE], **traveling** [AmE] and shopping.* (EFCAMDAT-Mainland China)
- (8) *Finally you should **have** [BrE] a **vacation** [AmE] in a countryside and walking almost three our a day, eat well and sleeping better.* (EFCAMDAT-Italy)

## 6. Conclusion

The aim of this chapter was threefold. First, it aimed to test the prediction of Mair's (2013) World System of Englishes that AmE (rather than BrE) should be the most influential model for all the other varieties of English. The second objective was to see whether Mair's model could be expanded to include EFL varieties, and how these would fare with respect to AmE vs BrE influence. Finally, the chapter sought to explore the influence of AmE vs BrE more locally and to relate possible differences between countries to their historical, economic and geographical relations with the US and the UK. The analysis of corpus data from GloWbE and EFCAMDAT revealed that globally AmE had more influence than BrE on World Englishes, although, when a distinction was drawn between ENL, ESL and EFL, this turned out to be true only for the non-native varieties. Among these, the influence of AmE was stronger on EFL, which was however characterised by a greater disparity between the different linguistic items. Within each of the three types of varieties, local variation in terms of norm orientation emerged, which could partly be accounted for by the contextual factors that were examined (shared colonial history with the US or the UK, value of imports and exports, and geographical proximity), but probably also by other factors (including educational or cultural factors) that could not be quantified. While this study has thus partly confirmed the position of AmE as a hyper-central variety, as posited in Mair's model, it has also shown that a great deal of variation exists and should be taken into account in the model, for example through the consideration of local contextual factors, in addition to the general factors of demographic weight and institutional support that are examined in the model. As for the EFL varieties, they appear to be worthy of inclusion in the World System of Englishes, although their position within the hierarchy would need to be defined and the real impact of Americanisation (instead of, or in addition to, other factors like preference for simple or transparent forms) would have to be clarified. Ideally, a multifactorial analysis should be performed with a view to identifying the possible interactions between the pairs of items, the English varieties, and any other factors that may have a role to play in the choice of

a certain norm orientation. Such an approach would allow for a more robust study and would provide more precise and reliable results, while also making it possible to predict preferences for an AmE or BrE item.

To finish with, it should be underlined that the results obtained in this study should only be taken as a first indication of the possible influence of AmE and BrE on some World Englishes. For one thing, only twenty pairs of items were used as a basis for the analysis. Not all of these items were typical of AmE/BrE to the same extent (although they all appeared to be significantly distinctive of one variety or the other) and they also differed in their Americanness rate within one and the same English variety. This means that a different selection of items may have provided different results. For another thing, it was demonstrated that different countries displayed different Americanness rates, so that the results for ENL, ESL and EFL may be said to depend on the combination of countries represented in the corpora used. Again, a different choice of corpora may have changed the results. Finally, a level of analysis that could also have an impact on the results but that was not taken into account here is that of the individual speakers. Each speaker, depending on his/her personal history, may be more or less influenced by a certain variety, be it AmE, BrE, or any other English variety for that matter. Depending on the particular combination of speakers represented in a corpus, the global results may display a more or less marked American influence. While general linguistic models abstracting away from individual variation are certainly useful for the bird's-eye view they offer, we should not forget that ultimately, it is the language of *people* that these models seek to describe, and that the specific context in which these individual language users evolve is likely to tip the scale – or scales.<sup>21</sup>

## Acknowledgements

I would like to thank the editor of this volume, Sandra Deshors, as well as two anonymous reviewers for their comments on an earlier version of this chapter.

---

21. Although, according to GloWbE, *TIP the scales* is the preferred form in both AmE and BrE, the singular form *TIP the scale* is more characteristic of AmE, where it represents 31% of the uses of *TIP the scale(s)*, than of BrE, where it represents only 9%.

## References

- Algeo, J. 2006. *British or American English? A Handbook of Word and Grammar Patterns*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511607240>
- Appadurai, A. 1996. *Modernity at Large: Cultural Dimensions of Globalisation*. Minneapolis MN: University of Minnesota Press.
- Braine, G. 2005. Introduction. In *Teaching English to the World: History, Curriculum, and Practice*, G. Braine (ed.), xi–xx. London: Routledge.
- Clark, R. 1998. Best of all worlds? *English Today* 14(4): 18.
- Collins, B. & Mees, I. M. 2013. *Practical Phonetics and Phonology*, 3rd edn. London: Routledge.
- Davies, M. & Fuchs, R. 2015a. 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. <https://doi.org/10.1075/eww.36.1.01dav>
- Davies, M. & Fuchs, R. 2015b. A reply. *English World-Wide* 36(1): 45–47. <https://doi.org/10.1075/eww.36.1.03dav>
- Davydova, J. 2012. English in the Outer and Expanding Circles. *World Englishes* 29(4): 546–556.
- Deshors, S. C. 2014. A case for a unified treatment of EFL and ESL: A multifactorial approach. *English World-Wide* 35(3): 277–305. <https://doi.org/10.1075/eww.35.3.02des>
- Deshors, S. C. & Götz, S. Forthcoming. Common ground across globalized English varieties: A multivariate exploration of mental predicates in World Englishes. *Corpus Linguistics and Linguistic Theory*.
- Deshors, S. C. & Gries, S. T. 2016. Profiling verb complementation constructions across New Englishes: A two-step random forests analysis of *ing* vs. *to* complements. *International Journal of Corpus Linguistics* 21(2): 192–218. <https://doi.org/10.1075/ijcl.21.2.03des>
- Edwards, A. & Laporte, S. 2015. Outer and Expanding Circle Englishes: The competing roles of norm orientation and proficiency levels. *English World-Wide* 36(2): 135–169. <https://doi.org/10.1075/eww.36.2.01edw>
- Geertzen, J., Alexopoulou, T. & Korhonen, A. 2014. Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT). In *Selected Proceedings of the 2012 Second Language Research Forum: Building Bridges between Disciplines*, R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. Marcos Miguel, A. M. Tseng, A. Tuninetti, & D. Walter (eds), 240–254. Somerville MA: Cascadilla Proceedings Project.
- Gilquin, G. 2008. Hesitation markers among EFL learners: Pragmatic deficiency or difference? In *Pragmatics and Corpus Linguistics: A Mutualistic Entente*, J. Romero-Trillo (ed.), 119–149. Berlin: Mouton de Gruyter.
- Gilquin, G. 2015. At the interface of contact linguistics and second language acquisition research: New Englishes and Learner Englishes compared. *English World-Wide* 36(1): 91–124. <https://doi.org/10.1075/eww.36.1.05gil>
- Grzega, J. 2005. Reflections on concepts of English for Europe: British English, American English, Euro-English, Global English. *Journal for EuroLinguistiX* 2: 44–64.
- Hackert, S. 2015. Pseudotitles in Bahamian English: A case of Americanization? *Journal of English Linguistics* 43(2): 143–167. <https://doi.org/10.1177/0075424215577966>
- Hackert, S. & Deuber, D. 2015. American influence on written Caribbean English. A diachronic analysis of newspaper reportage in the Bahamas and in Trinidad and Tobago. In *Grammatical Change in English World-Wide* [Studies in Corpus Linguistics 67], P. Collins (ed.), 389–410. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.67.16hac>

- Hänsel, E. C. & Deuber, D. 2013. Globalization, postcolonial Englishes, and the English language press in Kenya, Singapore, and Trinidad and Tobago. *World Englishes* 32(3): 338–357. <https://doi.org/10.1111/weng.12035>
- Jayapalan, K. & Pillai, S. 2011. The state of teaching and learning English pronunciation in Malaysia: A preliminary study. *Malaysian Journal of ELT Research* 7(2). <<http://journals.melta.org.my/index.php/majer/article/view/155>>
- Kachru, B. B. 1983. Regional norms for English. *Studies in Language Learning* 4(2): 54–76.
- Kachru, B. B. 1985. Standards, codification and sociolinguistic realism: The English language in the Outer Circle. In *English in the World: Teaching and Learning the Language and Literatures*, R. Quirk & H. G. Widdowson (eds), 11–30. Cambridge: CUP.
- Kahane, H. 1992. American English: From a colonial substandard to a prestige language. In *The Other Tongue. English across Cultures*, 2nd edn, B. B. Kachru (ed.), 211–219. Urbana IL: University of Illinois Press.
- Laitinen, M. 2016. Ongoing changes in English modals: On the developments in ELF. In *New Approaches to English Linguistics: Building Bridges* [Studies in Language Companion Series 177], O. Timofeeva, A.-C. Gardner, A. Honkapohja & S. Chevalier (eds), 175–196. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.177.07lai>
- Larsson, T. 2012. On spelling behavior(u)r: A corpus-based study of advanced EFL learners' preferred variety of English. *Nordic Journal of English Studies* 11(3): 127–154.
- Leech, G., Hundt, M., Mair, C. & Smith, N. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511642210>
- Lim, L. 2012. Standards of English in South-East Asia. In *Standards of English: Codified Varieties Around the World*, R. Hickey (ed.), 274–293. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139023832.015>
- Mair, C. 2013. The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34(3): 253–278. <https://doi.org/10.1075/eww.34.3.01mai>
- Mair, C., Mukherjee, J., Nelson, G. & Peters, P. 2015. Responses to Davies and Fuchs. *English World-Wide* 36(1): 29–44. <https://doi.org/10.1075/eww.36.1.02mai>
- Mauranen, A. 2011. Learners and users – Who do we want corpus data from? In *A Taste for Corpora: In Honour of Sylviane Granger* [Studies in Corpus Linguistics 45], F. Meunier, S. De Cock, G. Gilquin & M. Paquot, 155–172. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.45.11mau>
- Modiano, M. 1996. The Americanization of Euro-English. *World Englishes* 15(2): 207–215. <https://doi.org/10.1111/j.1467-971X.1996.tb00106.x>
- Modiano, M. 1999. Standard English(es) and educational practices for the world's lingua franca. *English Today* 15(4): 3–13. <https://doi.org/10.1017/S0266078400011196>
- Modiano, M. 2007. Euro-English from a 'deficit linguistics' perspective? *World Englishes* 26(4): 525–533. <https://doi.org/10.1111/j.1467-971X.2007.00526.x>
- Mukherjee, J. & Hundt, M. (eds). 2011. *Exploring Second-language Varieties of English and Learner Englishes: Bridging a Paradigm Gap* [Studies in Corpus Linguistics 44]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.44>
- Ranta, E. 2010. English in the real world vs. English at school: Finnish English teachers' and students' views. *International Journal of Applied Linguistics* 20(2): 156–177. <https://doi.org/10.1111/j.1473-4192.2009.00235.x>

- Reuter, V. 2016. The Influence of British and American English on some Varieties of English on the Basis of the GloWbE Corpus. MA thesis, Université catholique de Louvain.
- Schneider, E. W. 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79(2): 233–281.
- Schneider, E. W. 2007. *Postcolonial English. Varieties around the World*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511618901>
- Schneider, E. W. 2014. New reflections on the evolutionary dynamics of world Englishes. *World Englishes* 33(1): 9–32. <https://doi.org/10.1111/weng.12069>
- Simo Bobda, A. 1998. British or American English: Does it matter? *English Today* 14(4): 13–18. <https://doi.org/10.1017/S0266078400010464>
- Sridhar, K. K. & Sridhar, S. N. 1986. Bridging the paradigm gap: Second language acquisition theory and indigenized varieties of English. *World Englishes* 5(1): 3–14. <https://doi.org/10.1111/j.1467-971X.1986.tb00636.x>
- Szpyra-Kozłowska, J. 2015. *Pronunciation in EFL Instruction. A Research-Based Approach*. Bristol: Multilingual Matters.
- Trudgill, P. & Hannah, J. 2017. *International English: A Guide to Varieties of English Around the World*, 6th edn. London: Routledge. <https://doi.org/10.4324/9781315192932>
- Vine, B. 1999. Americanisms in the New Zealand English lexicon. *World Englishes* 18(1): 13–22. <https://doi.org/10.1111/1467-971X.00118>
- White, R. G. 1870. *Words and Their Uses, Past and Present: A Study of the English Language*. New York, NY: Sheldon and Company.
- Xu, Z. 2010. Chinese English: A future power? In *The Routledge Handbook of World Englishes*, A. Kirkpatrick (ed.), 282–298. London: Routledge.

## Appendix: Local contextual factors

**Table 6.** Shared colonial history between the US/UK and the countries under study

US	UK
<u>ENL</u> : /	<u>ENL</u> : Australia, Canada, Ireland, New Zealand
<u>ESL</u> : Philippines	<u>ESL</u> : Bangladesh, Ghana, Hong Kong, India, Jamaica, Kenya, Malaysia, Nigeria, Pakistan, Singapore, South Africa, Sri Lanka, Tanzania
<u>EFL</u> : Colombia (SA), Cuba (NA), Dominican Republic (NA), Honduras (NA), Japan (AS), Mexico (NA), Nicaragua (NA), Panama (NA)	<u>EFL</u> : Afghanistan (AS), Bahrain (AS), Cyprus (AS), Egypt (AF), Iraq (AS), Jordan (AS), Kuwait (AS), Myanmar (AS), United Arab Emirates (AS), Yemen (AS)

*This table is based on information found at [https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_colonial\\_possessions](https://en.wikipedia.org/wiki/List_of_United_States_colonial_possessions) and [https://en.wikipedia.org/wiki/List\\_of\\_countries\\_that\\_have\\_gained\\_independence\\_from\\_the\\_United\\_Kingdom](https://en.wikipedia.org/wiki/List_of_countries_that_have_gained_independence_from_the_United_Kingdom). The countries that have a shared colonial history with the US are listed in the first column and those that have a shared colonial history with the UK are listed in the second column.*

**Table 7.** Economic relations between the US/UK and the countries under study

US	UK
<u>ENL</u> : all countries considered	<u>ENL</u> : /
<u>ESL</u> : all countries considered	<u>ESL</u> : /
<u>EFL</u> : most countries considered	<u>EFL</u> : Cyprus (AS), Czech Republic (EU), Latvia (EU), Moldova (EU), Norway (EU), Poland (EU), Slovakia (EU)

*This table is based on the 2015 figures of the World Integrated Trade Solution website (<http://wits.worldbank.org>). It takes into account the combined value of imports and exports between the US/UK and each of the countries. If this value is higher with the US than with the UK, the countries are listed in the first column; otherwise, they are listed in the second column. No information was available for the following countries: Guadeloupe, Macau, Palestine and Taiwan.*

**Table 8.** Geographical proximity between the US/UK and the countries under study

US	UK
<u>ENL</u> : Australia, Canada, New Zealand	<u>ENL</u> : Ireland
<u>ESL</u> : Jamaica	<u>ESL</u> : Bangladesh, Ghana, Hong Kong, India, Kenya, Malaysia, Nigeria, Pakistan, Philippines, Singapore, South Africa, Sri Lanka, Tanzania
<u>EFL</u> : Guadeloupe (EU), all North American countries considered, all South American countries considered	<u>EFL</u> : all African countries considered, all Asian countries considered, most European countries considered

*This table is based on information found at <https://www.freemaptools.com/how-far-is-it-between.htm>. The distance between each country and the US/UK is calculated “as the crow flies”, with the centre of the country taken as a point of reference. The countries that are found to be closer to the US are listed in the first column, whereas those that are closer to the UK are listed in the second column.*



# It is time that this (*should*) be studied across a broader range of Englishes

## A global trip around mandative subjunctives

Marianne Hundt

Zurich University

In subordinate clauses after mandatives such as *recommend* or *require*, English speakers can choose between a subjunctive or a modal construction. Studies of this alternation in second-language varieties of English (ESL) are scarce. Previous research looked at the distribution of subjunctives across varieties, speech and writing, their co-occurrence with passive voice and individual triggers. However, these factors have not been studied as predictor variables for the subjunctive: modal alternation. This study uses corpus data from the *International Corpus of English* and the *Global Web-based English corpus* to model the relative strength of external predictor variables for the mandative alternation across a broad range of Englishes. The findings do not lend themselves to straightforward interpretation within an individual model of World Englishes.

**Keywords:** mandative sentences, subjunctive: modal alternation, probabilistic grammar approach

### 1. Introduction

The subjunctive in English has often been described as a ‘moribund’ (Fowler 1926) or near-extinct grammatical category (e.g. Harsh 1968). This view is occasionally maintained up until the end of the twentieth century; Denison (1998: 263) claims that “[i]n Br[itish] E[nglish] the present subjunctive [...] has retreated to high-flown literary or legal language [...]”. Yet, in subordinate clauses following mandative verbs, nouns and adjectives such as *demand* or *important* the subjunctive form has been increasing again in the twentieth century (e.g. Övergaard 1995, Leech et al. 2009).<sup>1</sup> The mandative phrase *it is time* from the title can be used to

---

1. A notable exception to this view is a recent study by Kastronic and Poplack (2014), who combine data from corpora of Early and Late Modern (British) English with a corpus of Present-Day



illustrate the variable context, i.e. the choice between a subjunctive form (typically distinct after third person singular subjects, as in Example (1)), a periphrastic construction with a modal verb (typically *should*, as in (2)) or, in some varieties, an indicative, as in Example (3):<sup>2</sup>

- (1) The rights and responsibilities of marriage are civil, they are legal, and now *it is time that they be made equal*. (GloWbE, US, General)
- (2) This is the English practice and, if it is not yet definitely settled in this country, *it is time that it should be*. (GloWbE, IE, General)
- (3) I recognise *it is time that* the unabated issue of dangerous dogs *is grabbed* by the scruff of the neck and brought firmly under control. (GloWbE, GB, General)

Previous research has shown that the revival of the mandative subjunctive can be observed in varieties of English as a first (ENL) or institutionalized second (ESL) language, and that this recent/ongoing change is spearheaded by American English (AmE) (e.g. Övergaard 1995; Hundt 1998). Various corpus-based studies have looked into the diachronic spread of subjunctive and regional variation, but the range of ESL varieties in previous research has typically been restricted to only one (e.g. Sayder 1989; Schneider 2005, 2011) or a few selected varieties (e.g. Peters 2009). Moreover, no previous research has looked into the relative importance that contextual and linguistic factors play in the choice between a subjunctive and the periphrastic construction with modal *should*. It is the aim of this chapter to address this research gap using data from the *International Corpus of English* (ICE) and the corpus of *Global Web-based English* (GloWbE).

In the following, I will briefly summarize the most important findings from previous corpus-based research into mandative subjunctives (Section 2) and give a detailed definition of the variable context, data retrieval and background on the factors included in the analysis (Section 3). In addition to summary statistics, I will combine random forest analysis and conditional inference trees to probe into the relative importance that contextual and linguistic factors play in the choice between subjunctive and modal periphrasis (Section 4). Which model of English as a World Englishes might best explain the patterns of variation found in the corpus data will be the subject of the concluding discussion (Section 5). As Mair

---

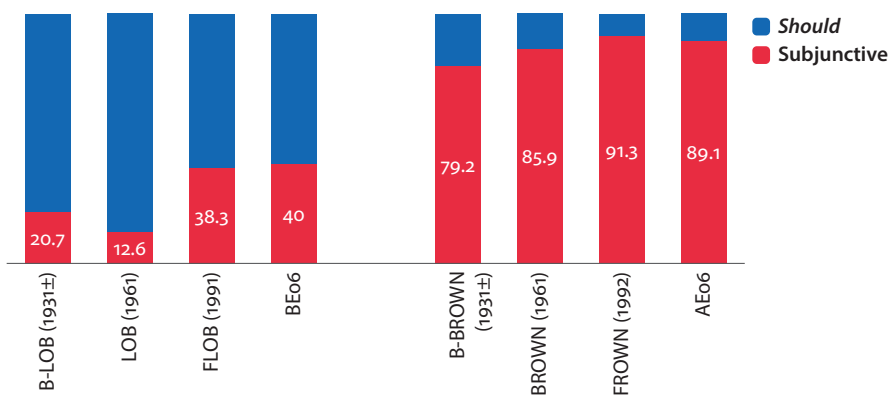
Canadian English and, using a very different methodology from previous studies, argue that the mandative subjunctive has not increased over the past few hundred years. This discrepancy is largely due to a very different way of defining the variable context (see Section 3).

2. The choice can also be avoided in various ways, e.g. by resorting to a *for-to*-infinitival construction (*They asked for Mr Robinson to resign*) or a nominal instead of a clausal complement (*They asked for Mr Robinson's resignation*). These options are not included as variants in this paper.

(2017: 105) points out, while models of World Englishes are not usually developed with a view to corpus-based research, they “can be seen as presenting a blueprint for a corpus-linguistic research agenda in the field.”

## 2. Previous research: Recent and ongoing change in World Englishes

One of the earliest studies, based on data from the Brown and LOB corpora and a set of pre-defined mandative triggers, found that the mandative subjunctive was used more often in American than in British English (Johansson & Norheim 1988). Follow-up studies using the same set of triggers and evidence from the Brown family of corpora (Hundt 1998; Hundt & Gardner 2017; Waller 2017) show that (a) the significant increase in AmE took place in the early years of the twentieth century and (b) the mandative subjunctive has been gaining ground in British English (BrE) in the second half of the twentieth century. However, the change in BrE happens at a much slower rate than in AmE and might already be levelling off at a lower level, judging from the web-based BE06 evidence (see Figure 1).<sup>3</sup>



**Figure 1.** Mandative subjunctives (%) in the Brown-family corpora; variable contexts: B-LOB (1931±) = 92; LOB (1961) = 111; FLOB (1991) = 117; BE06 = 70; B-Brown (1931±) = 96; Brown (1961) = 134; Frown (1992) = 115; AE06 = 92

3. The data for LOB and Brown are from Johansson and Norheim (1988), those for FLOB and Frown from Hundt (1998); data for the extended Brown family (for the 1930s and 2006, respectively) come from Hundt & Gardner, (2017) and Waller (2017). Note that Waller (2017) finds that proportions of variants in mandative sentences are based on slightly different definitions of the variable across studies, for instance with respect to the inclusion/exclusion of bare forms with plural subjects after past tense triggers (see Example (8) below). The overall development and general regional variation are not affected by these differences.

Corpus data also provides evidence that settler varieties such as Australian (AusE) and New Zealand English (NZE) (e.g. Hundt 1998; Peters 1998; Collins 2015) occupy an intermediate stage between BrE and AmE, with a slight preference of subjunctives over periphrastic constructions. Ongoing change in ENL varieties has typically been explained in terms of American influence/Americanization (see e.g. Övergaard 1995: 89 or Mair 2006: 193).

For ESL varieties, previous research based on Brown-type corpora found that they either used similarly low proportions of mandative subjunctives as BrE (e.g. Sayder 1989 or Schneider 2000 on Indian English) or that they aligned with AmE (e.g. Schneider 2005, 2011 and Collins et al. 2014 on Philippine English). Typically, studies on ESL varieties compare only one of these with ENL reference varieties. A notable exception is Peters (2009), who compares evidence on Philippine English (PhilE), Indian English (IndE) and Singapore English (SingE) with three ENL varieties (BrE, NZE and AusE). Her study, which uses six lexical triggers<sup>4</sup> and both written and spoken evidence from the respective components of ICE confirms previous results for PhilE and IndE, i.e. the affinity with the respective matrulect.<sup>5</sup> However, SingE turns out to prefer subjunctives over modal constructions, thus casting some doubt on the historical lineage as the best explanation for the usage patterns in ESL varieties. Moreover, for IndE, her study reveals regular use of deontic and quasi-modals in subordinate clauses after mandative triggers, i.e. variants not typically included in the study of the variable context; she also finds evidence of nativized patterns that avoid the choice between a subjunctive and the modal construction in ways different from ENL varieties (Peters, 2009: 130).

With respect to variation across speech and writing, earlier studies have shown that the subjunctive is not limited to formal, written language. On the contrary, corpus data have consistently shown that the subjunctive is regularly attested in speech (e.g. Hoffmann 1997; Hundt 1998; Schneider 2005), though how this factor may interact with regional variety has not yet been discussed on the basis of conclusive evidence (see Peters 2009: 130).

This short review of previous research shows that additional investigation on the mandative subjunctive in World Englishes is warranted for various reasons. First, comparison across different regional varieties is confounded by the fact that previous research is not necessarily based on the same set of mandative triggers (see also Section 3). In addition, studies typically focussed on frequency differences and/or historical developments but rarely looked at factors that might influence

---

4. She searched the ICE corpora for variant forms of the verbs *demand*, *move*, *recommend*, *request*, *require* and *suggest* as well as their related nouns (Peters, 2009: 133).

5. Note that Sedlatschek (2009: 286–88), on the basis of a small set of verbs invested in internet-based data, finds IndE to be more conservative than its matrulect.

the choice between a subjunctive and the periphrastic variant. In other words, there is very little research across ENL and ESL varieties on micro-variation in the use of mandative subjunctives in relation to register, mandative trigger, negation, etc. and none on the relative importance that such factors may have. Is register, for instance, a more important factor than regional variation, as in various other studies on grammatical variation across ENL and ESL varieties? Does the verb *be*, the most clearly subjunctive form after mandative triggers, provide a stronghold for the subjunctive across all World Englishes, as previous research suggests (e.g. Turner 1980: 275)? How frequent are negated subjunctives and are they typical of formal, written language? Is *that*-omission in mandative sentences associated with a particular region and/or register? This chapter uses evidence from a broad range of ENL and ESL varieties to answer these questions against the backdrop of existing models of World Englishes.

### 3. Data and methodology

#### 3.1 Corpus data of World Englishes

This paper uses data from a total of 10 ICE components, representing five ENL varieties (ICE Canada, Great Britain, Ireland, New Zealand and Australia), four ESL varieties (ICE Hong Kong, India, Philippines and Singapore) and one country where standard English is used as a second dialect alongside an English-based creole (ICE Jamaica). In terms of regional distribution, this means that the focus of the investigation will be on North American Englishes, Britain, Ireland, Southern Hemisphere settler varieties, on the one hand, and Asian Englishes on the other hand. One condition for selecting the ICE components to be included was that a complete sample (i.e. both the written and spoken part) had to be available, as one of the research questions for this paper was to include the factor ‘medium/register’ in the analysis, even though this meant that US English could not be included. At approximately 1 million words each, however, the ICE corpora are too small to yield sufficient evidence on variability with individual triggers.<sup>6</sup> More precisely, the overall number of variable contexts for the triggers turned out to vary greatly (see Section 4.1). In a follow-up study, additional data were therefore retrieved from GloWbE to probe into the effect that the trigger might have on the choice of

---

6. Exact (and comparable) word counts for some ICE components are currently difficult to obtain. The online version of ICE components to be made available at [www.ice-corpora.uzh.ch](http://www.ice-corpora.uzh.ch) will address this issue. For the current study, the dependent variable is defined as a choice context. This allows comparisons without normalization.

verb form in mandative sentences. GloWbE provides subcorpora of varying size for twenty varieties of English, sampling (informal) blog data, on the one hand, and ‘general’, on the other.<sup>7</sup> In order to keep the amount of data for this follow-up study manageable, the focus in this part of the paper is on AmE and BrE as reference varieties, as well as Indian English as an ESL variety.

### 3.2 Definition of the variable context

As pointed out above, previous studies have used different approaches in defining the variable context. This applies both to the selection of the triggering context and the variants that were included in the datasets. However, in order to allow for replicability, it is important to not only define the variable but also to provide sufficient detail on how data for the present paper were retrieved from the corpora. Johansson and Norheim (1988), for instance, used a predefined set of 17 *suasive* verbs, 11 nouns and 5 adjectives to retrieve mandative contexts from LOB and Brown, presumably on the basis of lists found in standard grammars. In the interest of replicability, this set of triggers was then used in diachronic studies of Brown-family corpora (e.g. Hundt 1998; Hundt & Gardner 2017; Waller 2017). Algeo (1992) lists 25 verbs, 29 nouns and 18 adjectives that may all trigger a mandative subjunctive, Crawford (2009) a total of 108 triggers and Kastronic & Poplack (2014) 240.<sup>8</sup> Hoffmann (1997) uses a corpus-based approach to arrive at the set of triggers he investigates: he first searched for instances of subjunctive *be*, and from the list of 276 items then used the most frequently attested triggers to retrieve his mandative sentences. However, even such an approach would miss instances like the following, that were retrieved searching for the sequence *that + personal pronoun + not*:

---

7. Note that the ‘general’ part of the corpus might occasionally also include blog data. For background on GloWbE and the advantages and disadvantages of this resource for World Englishes research, see Davies & Fuchs (2016) and the responses to their article by Christian Mair, Joybrato Mukherjee, Gerald Nelson & Pam Peters in *English World Wide* 36(1).

8. Crawford’s list of triggers is based on an unpublished paper by Albakry & Crawford 2004, but the list provided in the appendix of Crawford (2009: 274) shows that it is not derived in a bottom-up fashion from corpus data but based on the lists provided in Quirk et al. (1985) and Övergaard (1995). Moreover, Crawford (2009) only uses the 16 nouns, 11 verbs and 6 adjectives that are attested at least once with the subjunctive in his news corpus (2009: 261). Kastronic and Poplack (2014) used previous studies as well as corpus data (following Hoffmann’s 1997 approach), but they fail to list the specific triggers that their study is based on. In a footnote they mention that *say*, *think*, *feel* initially were among the triggers but go on to say that they were not included in the rate calculations as they were overwhelmingly used with the indicative (Kastronic & Poplack, 2014: 77).

- (4) On reviewing those remarks I *regret that I not provide* sufficient airing of Lessing's notions on that subject, and remained content with a cryptic reference. (ICE-SING, W2D-019)
- (5) ... we pray and *pray* intensely *that we not be found wanting*. (ICE-PHI, S2B-023)

Ultimately, a complete set of mandative sentences is impossible to obtain from a corpus as mandative subjunctives can also be used without overt triggers (see Övergaard 1995: 82). Moreover, in order to cover a relatively broad range of Englishes, the number of lexical triggers used in this paper had to be limited to keep the amount of data that had to be manually post-edited manageable. I opted to restrict data retrieval to eleven verbs and three adjectives.<sup>9</sup> These were selected on the basis of having been shown in previous studies to regularly trigger subjunctives. For data retrieval from GloWbE, one hundred variable contexts were retrieved from the three sub-sections for the following verbs (all forms): *demand*, *order*, *propose*, *recommend*, *request* and *require*, giving a total of 1800 hits. Note that the deontic force of these triggers is quite different (i.e. stronger for *demand*, *order* and *require* than for *propose* and *recommend*). The former might therefore more frequently trigger a subjunctive than the latter.

Not all subordinate clauses following these triggers are instances of choice contexts, however. Waller (2017: 81) distinguishes four identifying contexts for the mandative subjunctive and introduces his own labels for them: a third person singular noun phrase followed by an unmarked verb (iNO-S), as in Example (6), unmarked use of *be* (iBE) as in (7), an unmarked verb following a past-tense trigger (iST), as illustrated in Example (8), and pre-verbal *not*-negation (iNEG), exemplified in Example (9):<sup>10</sup>

- (6) May I *ask* that *the uh prosecutor furnish* us at least a copy of what they have furnished. (ICE-PHI, S1B-062)
- (7) Dewey *requires* that pupils *be* given wide opportunities for purposive inquiry. (ICE-IRE, S2B-035)
- (8) In short he *ordered* that *his troops use* terrorism. (ICE-CAN, W1A-005)
- (9) It's *important* that *you not wait* until severe soiling has occurred before cleaning your upholstery. (GloWbE, CA G)

9. The verbs (all variant forms) included were *ask*, *demand*, *dictate*, *insist*, *order*, *propose*, *recommend*, *request*, *require*, *suggest*, *urge*; the adjectives were *essential*, *imperative*, *important*.

10. Kastronic and Poplack (2014: 73) simply define the subjunctive as a morphological category, i.e. as all instances of *be* and bare forms following a third-person singular subject NP. This means that their study did not include instances with disambiguating tense sequences as in Examples (8) and (7).

These identifying contexts can also co-occur, as in the following example with a past tense in the main clause, pre-verbal negation and use of unmarked *be*:

- (10) ... he suggests opening up the least vulnerable part and *recommended* that all visitors *not be* concentrated on the same spot. (ICE-IRE, W1A-018)

However, in the following example with a past-tense trigger, the speaker self-corrects from a subjunctive to a past tense verb; I therefore decided not to include this particular instance:

- (11) And he *suggested that* as soon as I got here I came and *see saw* you. (ICE-GB, S1A-051)

Extensive manual post-editing of the concordances is necessary to include only true subjunctives and those instances with *should* that would also result in an unambiguous mandative had the modal been omitted. Thus, Example (12) is ambiguous because the verb following the first-person subject is not *be* and could therefore be either indicative or subjunctive. Example (13) is ambiguous because the subject is a collective noun that is ambiguous in terms of number (collectives in English can be used with both plural and singular verbs and pronouns). Example (14) was excluded because there is no disambiguating context that would result in a mandative subjunctive were the modal to be left out.

- (12) ... what do you recommend *I read* to go into that more. (ICE-CAN, S1B-012)
- (13) The PQ is once again demanding that *Quebec opt out* of National Agricultural Program. (ICE-CAN, S1B-028)
- (14) *We propose that we should have* a meeting together to discuss the proposal further. (ICE-HK, W1B-023)

With collective nouns, pronominal clues in the immediate context can disambiguate the verb, making *sit up*, *use* and *allow* subjunctive; such instances were included in the final analysis:

- (15) She *demands that her audience sit up, use its* intellect and *allow* her plays to challenge. (ICE-NZ, W1A-002)

In addition to ambiguous instances, occurrences in which the trigger did not have mandative meaning were also removed during manual post-editing, including the following examples:

- (16) However the unions are *insisting* that the order *means* a return to the status quo of seven-member teams (ICE-JAM, S2B-005)



- (17) We have studies by psychologists about how jurors make up their mind *suggesting* [=‘implying’] *that* jurors typically decide by constructing or testing the alternative ... (ICE-GB, S2A-044)

I also decided to remove instances where the trigger was separated from the following context by a punctuation mark; typically, this concerns lists, as in the following example:

- (18) In his experiments with the culture of *tikog*, Herminio Pava, professor VI of Central Mindanao University in Musuan, Bukidnon, *recommends* the following:  
\*Planting distance should be at least 8 to 10 away from each other. (ICE-PHI, W2D-016-017)

ESL varieties provide additional cases that need to be carefully considered during manual post-editing. At times, the triggers are used with different semantics than we would expect from ENL varieties, as in the following example from ICE-IND, where *request* is used in the sense of ‘ask’.

- (19) So second September immedietly [sic!] I requested *them you send the <, > some of sample* which you’ve collected. (ICE-IND, S1B-029)

What complicates the analysis in this particular case is that the trigger is followed by what could either be interpreted as a subordinate clause (in which case *request* would be used as a transitive verb) or – since it is from the spoken part of the corpus – a quotation of what the speaker originally uttered as a request. For this reason, this particular occurrence was considered to be ambiguous and therefore not included in the analysis.

As pointed out in the introduction, English allows for the choice between a subjunctive, a modal construction or an indicative after mandative triggers. Other modals may also occur in mandative sentences, as the following examples show, including semi-modal *ought to*:

- (20) Mr Heseltine has been no less resolute than Mrs Thatcher in insisting that the dictator *must* withdraw from Kuwait. (ICE-GB, W2E-004)
- (21) We are simply asking that there *ought to* be a mechanism whereby our salary could be determined and adjusted in future ... (ICE-HK, S1B-042)

For the modelling of variation across the varieties (Section 4.2) the envelope of variation was narrowed to include only the subjunctive and the modal construction with *should*. The range of possible patterns that can follow a mandative trigger may, occasionally, even include a past tense indicative, as in (22) or (23).<sup>11</sup>

11. The past indicative is followed by a *be*-subjunctive in Example (23); the latter was included in the analysis.



- (22) and on the basis that there is apparently no new money available / *i would be loath to recommend* at this point / particularly as i'm departing / *that money was taken* from other schemes (ICE-NZ, S1B-07)
- (23) After the doctor's report, the company asked that he *resigned* or *be fired*. (ICE-JAM, W2C-018)

The use of a past indicative in (22) may have to do with the fact that the trigger occurs in a hypothetical sentence and the subordinate clause occurs at quite a distance from the trigger.

### 3.3 Predictor variables

In a second step, the mandative sentences retrieved from the ICE corpora were coded for two external predictor variables (variety and medium/register) and six contextual variables. The factor 'variety' is self-explanatory, but the external predictor 'medium/register' needs further commentary. The ICE sampling frame distinguishes a total of eight written and four spoken categories (see e.g. Greenbaum 1996). For the medium/register analysis in this paper, however, the four macro-categories 'spoken dialogue', 'spoken monologue', 'written unpublished' and 'written published' were used to avoid the risk of having too sparsely populated cells in the subsequent statistical modelling. The contextual linguistic variables included in the data coding were 'trigger' (with the values of the individual lexical items), 'trigger type' (verb vs. adjective), 'controlling subject' (with the values 'third person singular' vs. 'non-third person'), 'verb' (for the lexical verbs in the subjunctive/modal construction, with the values 'be' vs. 'other'), 'negation' (with the values 'negative' vs. 'affirmative') and 'subordination' (with the values 'that' vs. 'zero').

### 3.4 Data retrieval

#### 3.4.1 Core study

In order to include instances where the subordinating conjunction had been omitted (see Examples (24)–(26)), data retrieval relied on the mandative triggers, only.

- (24) Initially the captors demanded  $\emptyset$  the cash *be dropped* by plane (ICE-AUS, S2B-005)
- (25) I insisted  $\emptyset$  he come and see our establishment before he did that. (ICE-IND, W2F-009)
- (26) ... the council's planning consultant recommended  $\emptyset$  the consent *exclude* general engineering fibreglassing spray painting and steam or sand blasting operations. (ICE-NZ, W2C-008)

One disadvantage of this approach is that it retrieves a relatively large number of false positives. For the lemma *ASK*, for instance, the search yields well over 2,000 hits, of which only 37 showed variation between *should* and a subjunctive. Numerous instances were either not mandative uses of *ASK* (see (27)) or avoided the choice between the subjunctive and a modal construction by using a non-finite complement (as in (28) and (29)).<sup>12</sup>

- (27) I asked Malang *what he thinks* the peoples of the planet should do to strive for sustainability. (ICE-CAN, S2B-038)
- (28) We would certainly ask *for* those *to reduce* the time. (ICE-GB, S1A-024)
- (29) Then after the interview they asked me *to join* in the month of July. (ICE-IND, S1A-035)

Complementation of the trigger *URGE* provides further interesting evidence of the possibility of avoiding the choice between a subjunctive and modal periphrasis. In addition to combining an agent NP with a non-finite construction as in (30), it is also possible to use a deverbal noun and move the agent to a *by*-phrase, as in (31).

- (30) I would urge *members to give consideration* to nominating as a Director. (ICE-AUS, W1B-024)
- (31) He said problems with boarding houses were highlighted by the Burdekin report on the rights of mentally ill people, which urged *stringent licensing and regulation by State governments*. (ICE-AUS, W2C-017)

Statistical analysis of the ICE data showed that the trigger is a significant factor (see Section 4.2), but one that needed to be further controlled as the raw number of variable contexts per trigger in ICE varied greatly.

### 3.4.2 Follow-up study

For the follow-up study, the goal was to retrieve 100 variable contexts per trigger and variety. This meant that the number of triggers and varieties had to be reduced to keep the amount of data manageable. The focus will be on AmE and BrE as metropolitan reference varieties and IndE as an institutionalised second-language variety. The six triggers (*demand*, *order*, *propose*, *recommend*, *request* and *require*) were selected because they were relatively frequently attested in the variable context in previous research.<sup>13</sup> In order to increase precision and reduce the number of

12. Note that in some ESL varieties and JamE, *request* is also regularly used in the sense of ‘ask’ with the same complementation pattern, i.e. a direct object followed by a *to*-infinitive rather than a *that*-complement clause: “I request you to convey your sanction at the earliest” (ICE-IND, W1B-021).

13. Data were originally also collected for *urge*, but since these did not amount to the required number of 100 variable contexts per variety, these were not included.

sentences that had to be removed at the manual post-editing stage, data extraction for the follow-up study was limited to instances where the trigger was followed directly by a subordinating conjunction. The GloWbE data were manually post-edited and coded for five predictor variables: ‘variety’, ‘trigger’, ‘controlling subject’, ‘verb’ and ‘negation’. A predictor variable ‘register’ (general vs. blog) was not included as about 20% of the material included in the ‘general’ category also comes from blogs (see Davies & Fuchs, 2015: 4).

#### 4. Results: Contextual and/or linguistic factors?

I will first present an overview of the distribution of subjunctives and periphrastic constructions with *should* in the ICE data, then move on to the statistical analysis of this data set (Sections 4.1 and 4.2). The additional evidence from GloWbE on the potential impact that the lexical triggers may have on the choice of a subjunctive will be presented in Section 4.3.

##### 4.1 Distribution of mandative subjunctives in ICE by predictor variable

The search for mandative sentences with an unambiguous mandative verb or periphrastic *should* construction in the 10 ICE corpora yielded a total of 403 instances, of which the majority (70.7%) are subjunctives. Not all regional varieties contribute to the dominance of the subjunctive, though, as the distribution of the variants across the WEs examined (see Table 1) shows: While CanE shares the preference with its American neighbour for the subjunctive, periphrastic *should* is preferred in ICE-GB. In the ICE data, NZE and AusE also clearly prefer the subjunctive. As in previous studies (Schneider 2005, 2011; Collins et al. 2014), PhilE aligns with the ‘American’ usage pattern in its strong preference for the subjunctive. Varieties showing a more even distribution of the two options are IrE, IndE and HKE.

**Table 1.** Distribution of subjunctives and *should*-constructions across varieties (raw frequencies) in ICE

	CanE	BrE	IrE	NZE	AusE	HKE	IndE	PhilE	SingE	JamE
<i>should</i>	6	16	18	15	5	16	16	11	10	5
subjunctive	35	10	18	47	35	14	16	46	26	38

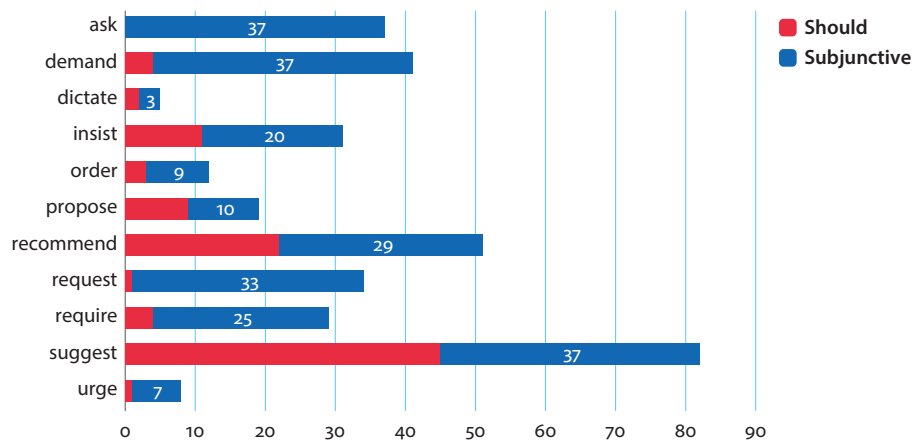
With respect to the distribution of subjunctives against modal variants across speech and writing (see Table 2) the ICE data do not reveal a marked difference. Interestingly, the informal spoken and written registers yield an even higher proportion of

subjunctives than the formal ones, but this is not a result that finds a straightforward explanation, e.g. in the innovative varieties providing proportionally more evidence from informal spoken and written contexts. The fact that the largest total number of choice contexts comes from written published English probably explains why people associate the subjunctive with this register.

**Table 2.** Distribution of subjunctives and *should*-constructions across ICE registers (macro-categories)

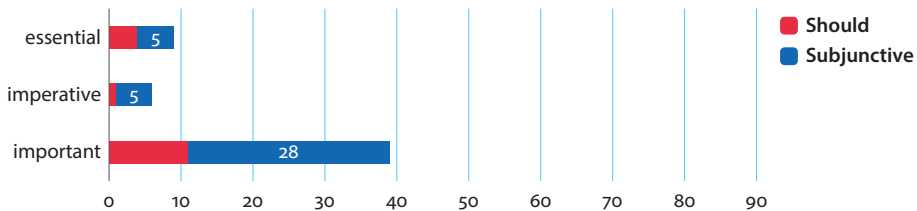
	dialogue	monologue	unpublished	published
<i>should</i>	16 (22.5%)	41 (37.6%)	13 (20.6%)	48 (30%)
subjunctive	55 (77.5%)	68 (62.4%)	50 (79.4%)	112 (70%)

Preliminary evidence from ICE indicates that the lexical trigger plays an important role in the choice between the two variants (see Figures 2a and 2b), with *urge*, *request*, *dictate* or *imperative* favouring the subjunctive and *recommend* or *suggest* showing a more even distribution of the variants. However, as the amount of data for individual triggers varies greatly, additional evidence on this factor for a limited number of varieties is discussed in Section 4.3 below.



**Figure 2a.** Distribution of *should* vs. subjunctive (raw frequencies) across verbal triggers in ICE

When we look at the role that the controlling subject may play (see Table 3), we see that the third person is an obvious stronghold of the subjunctive (61% of all subjunctives occur with a third person subject). However, in terms of choice contexts, non-third person subjects proportionally occur more often with the subjunctive than with a modal.



**Figure 2b.** Distribution of *should* vs. subjunctive (raw frequencies) across adjectival triggers in ICE

**Table 3.** Distribution of subjunctives and *should*-constructions by controlling subject in ICE

	third person singular	other
<i>should</i>	80 (31.5%)	38 (25.5%)
subjunctive	174 (68.5%)	111 (74.5%)

With respect to the verb used in the choice context, the summary statistics in Table 4 show that, in absolute terms, *be* is a stronghold for the subjunctive, as in previous research; however, proportionally, the choice for the subjunctive is not more likely with *be* than with other verbs.

**Table 4.** Distribution of subjunctives and *should*-constructions by variable verb (*be* vs. other)

	<i>be</i>	other
<i>should</i>	90 (32%)	28 (23%)
subjunctive	191 (68%)	94 (77%)

As far as the possibility of *that*-omission is concerned, Table 5 shows that overall it is relatively infrequent at 11.2% of all mandative contexts in the ICE data. More importantly, there is no substantial difference between subjunctives and periphrastic constructions with *should* as far as this factor is concerned. Interestingly, some regional varieties allow for more *that*-omission (New Zealand, Australian and Philippine English) whereas others strongly favour overt subordination in mandative constructions (namely British, Irish and Jamaican English).

**Table 5.** Distribution of subordination (zero vs. *that*) across varieties (figures in brackets for varieties give the subjunctives, only)

	<i>should</i>	subj.	CanE	BrE	IrE	NZE	AusE	HKE	IndE	PhilE	SingE	JamE
<i>that</i>	107 (90.7)	251 (88.1)	38 (32)	25 (10)	36 (18)	49 (37)	31 (26)	25 (11)	30 (14)	49 (41)	33 (24)	42 (38)
zero	11 (9.3)	34 (11.9)	3 (3)	1 (0)	0 (0)	13 (10)	9 (9)	5 (3)	2 (2)	8 (5)	3 (2)	1 (0)

Finally, the ICE data show that the subjunctive strongly disprefers negative contexts: out of 285 subjunctives, only 3 (1.1%) were negated, compared with 12 (10.2%) negated instances of *should* (out of a total of 118) in mandative constructions. In addition to the one attestation of a negated subjunctive ICE-IRE quoted above (see Example (10)) there is one each in ICE-CAN and ICE-JAM:

- (32) Since this story is popular with men, it is *important that* the woman *not become* too independent. (ICE-CAN, W1A-017)
- (33) Oh, but the crab catchers have *suggested that you not try* that method, have they? (ICE-JAM, W2F-006)

Corpus data from ICE thus confirm previous corpus-based finding (e.g. Hundt 1998, Leech et al. 2009), which also showed negated subjunctives to be rare. Moreover, all three instances are from the written part of the corpus. But both Example (33) and the following instance from a chatty article in the Indian edition of *Cosmopolitan* magazine show that negated subjunctives are not necessarily limited to formal contexts in writing:

- (34) If you are one of those women who complain about not finding a guy who knows how to cook and clean, chances are you aren't hanging out with the right kind of man. In all honesty, tradition *demanded that* Indian men *not step* into the kitchen but if you are going to break every other kind of tradition then the least you can do is not pick this one as the only tradition to maintain. (GloWbE, www.cosmopolitan.in)

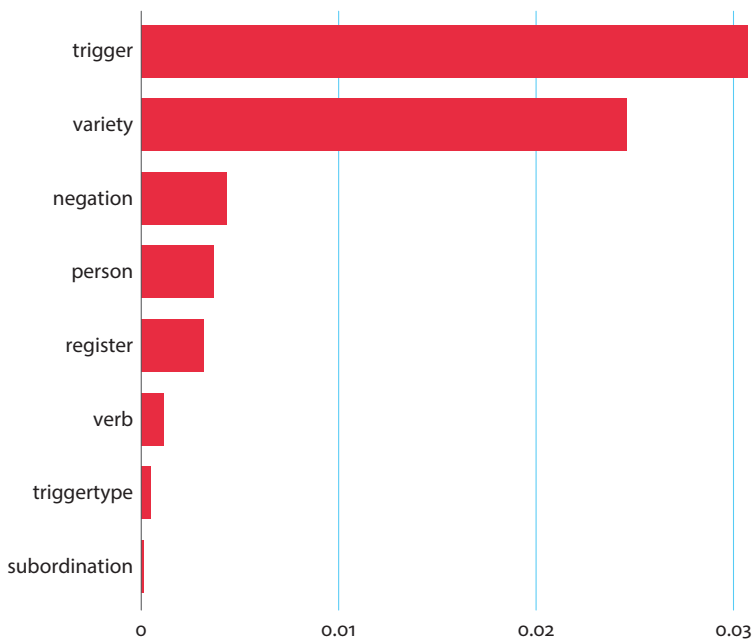
The question remains, however, what the relative importance of the factors is and how usage of subjunctives plays out across the different regional varieties.

#### 4.2 Statistical analysis: Variable importance

A traditional approach to modelling variable importance in a multivariate approach is through variable rule or regression analysis. There are two problems inherent in logistic regression models, however, as Szmrecsanyi et al. (2016) point out, namely predictor multicollinearity (the risk that predictors in the model are correlated) and data overfitting. In order to model probability grammar in World Englishes, they use both a random forest analysis and a conditional inference tree (ctree) analysis, an approach also recommended by Tagliamonte and Baayen (2012). The random forest analysis is a variant of permutation testing, which has the advantage that it does not assume a certain data distribution but instead builds the distribution by recursively resampling the data. It provides information (on the basis of a large number of trees) on the relative importance of predictor variables but does not indicate how these might interact with each other (e.g. Strobl, Malley & Tutz 2009

or Strobl, Hothorn & Zeileis 2009). The conditional inference tree (ctree) analysis returns a single tree which, on the basis of recursive partitioning of the underlying data, makes predictions in the form of binary splits of the data in a hierarchical fashion, thus showing predictor interaction. For the purposes of this study, interaction between ‘variety’ and other predictor variables is of particular interest. The two different kinds of analysis thus complement each other. They were computed using R’s party and partykit packages.

The relative ranking of the factors in the random forest analysis (with  $n\text{tree} = 500$  and  $m\text{try} = 2$ ) in Figure 3 returns ‘trigger’ as the most important predictor variable in the choice between subjunctive and *should* periphrasis, with ‘variety’ coming second. All the other factors are far less important.



**Figure 3.** Variable factor importance (Random Forest Analysis) for the ICE data

The variable importance plot in Figure 3 is based on the run with all predictor variables. Testing for model fit with Somers2 Dxy returns a prediction accuracy of 0.75 and a C-index value of 0.875, which is above the 0.8 level recommended e.g. in Tagliamonte & Baayen (2012: 156) thus already indicating a good model fit. The best model fit (with a C-index of 0.896 and a prediction accuracy of 0.79) was returned in a run with only ‘trigger’, ‘variety’ and ‘register’ and ‘person’ as predictor variables.

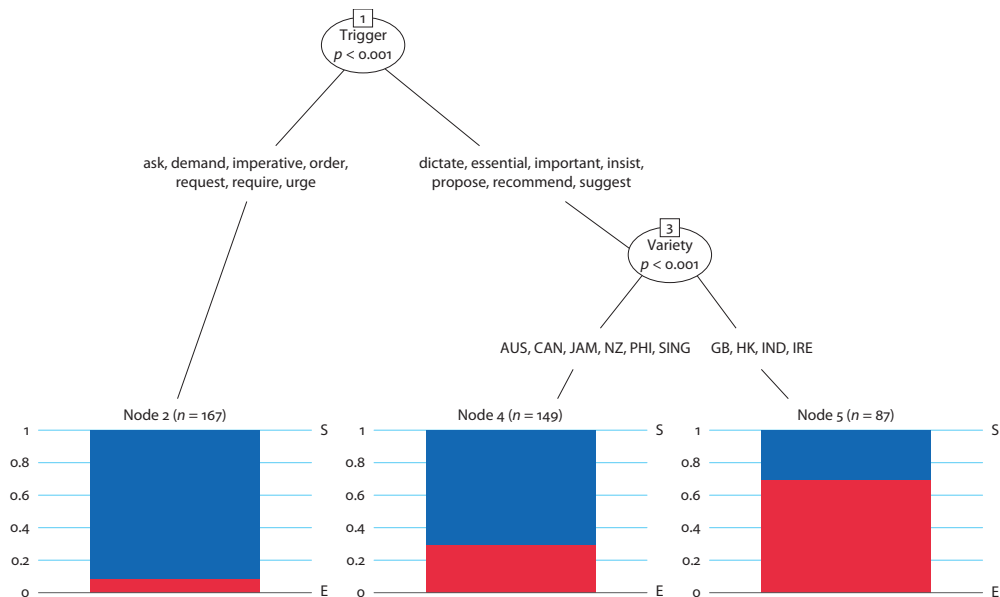


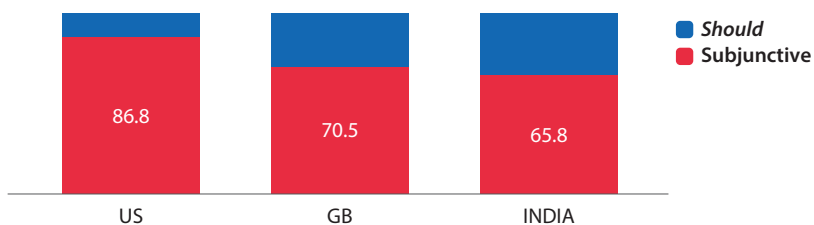
Figure 4. Plotting predictor interaction for mandates in ICE (ctree)



The single tree (with  $\text{maxdepth} = 4$ ,  $\text{mincriterion} = 0.95$ ) also selects ‘trigger’ at the first split, but only ‘variety’ as the other predictor variable (note, however, that prediction accuracy for the ctree is at 0.425 and thus slightly below the 0.5 level, with a C-index of 0.71). Figure 4 also shows that the triggers *ask*, *demand*, *imperative*, *order*, *request*, *require* and *urge* strongly favour the subjunctive (regardless of variety), whereas triggers *dictate*, *essential*, *important*, *insist*, *propose*, *recommend* and *suggest* are predominantly followed by a subjunctive in AusE, CanE, JamE, NZE, PhilE and SingE, while BrE, HKE, IndE and IrE show a preference for the modal construction with this set of triggers.

### 4.3 Gauging the effect of the trigger

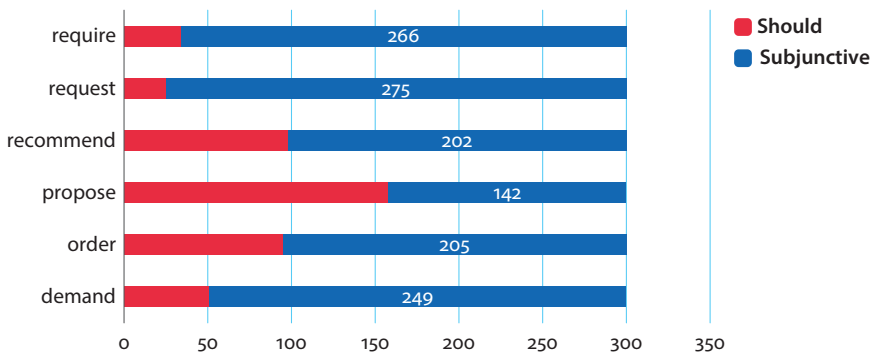
Figure 5 summarizes the distribution of subjunctives and *should*-constructions across the three varieties in the GloWbE data. As expected, AmE shows the highest proportion of subjunctives. Interestingly, the proportion of subjunctives in GloWbE is higher for BrE and IndE than what previous studies, based on standard reference corpora, would suggest. This could be the result of the continued spread of the subjunctive in these varieties, as the GloWbE data are more recent than the material sampled in the Brown family and ICE corpora. Alternatively, it might have to be attributed to the nature of the data included in GloWbE (i.e. the sampling frame, while trying to emulate the ICE criteria, might not have succeeded in providing a stylistically comparable set of data). Overall, the differences between AmE and BrE still prove significant at  $p < 0.01$  in a chi-square test, whereas those between BrE and IndE do not.



**Figure 5.** Mandative subjunctives (%) in GloWbE (600 variable contexts each per variety)

If we turn to the distribution by lexical trigger (see Figure 6), we see that there are verbs that strongly prefer the subjunctive across the GloWbE data analysed here, namely *require* (88.7%) and *request* (91.6%), to a somewhat lesser extent *demand* (83%); *recommend* and *order* trigger subjunctives at lower rates, namely at 67.3% and 68.3%, respectively; *propose*, finally, has the highest number of periphrastic

*should*-constructions at just over 50% (52.6%). The lexical trigger thus turns out to have a significant effect on the choice of a subjunctive over a periphrastic construction in a larger dataset that provides samples of equal size per trigger.



**Figure 6.** Mandative subjunctives (%) in GloWbE (300 variable contexts per verb, 100 from each variety)

As in the ICE data, the verb *be* is not used significantly more often as a subjunctive than other verbs (the proportion of subjunctives for *be* and other verbs is 74%); negation does not prove significant in a chi square test, either (overall, negative subjunctives only amount to just under 4% of all subjunctives). Third person subjects, finally, are followed significantly more often by a subjunctive (76.7%) than other subjects (70.4%) in the GloWbE data.

The random forest analysis (with *ntree* = 500 and *mtry* = 2) shows that ‘trigger’ is the most important predictor variable, with ‘variety’ coming second (see Figure 7). Model validation with Somers2 returns an accuracy level of 0.546 and a C-index of 0.773 (i.e. above the 0.5 level of random assignment but very slightly below the 0.8 level recommended by Tagliamonte and Baayen, 2012), i.e. shows that the analysis with all predictor variables already results in a good model fit. The best model fit is achieved with ‘trigger’, ‘variety’, ‘verb’ and ‘person’ as predictor variables (accuracy level of 0.558 and a marginally higher C-index of 0.779).

The single tree in Figure 8 based on the GloWbE data and including all predictor variables splits the data into two groups: triggers that strongly favour the subjunctive (nodes 2–4) vs. those that do so to a lesser extent, particularly so in BrE and IndE (nodes 13–15). Interestingly, AmE and BrE show a similarly strong preference for subjunctives with the triggers *demand*, *request* and *require* (nodes three and four), whereas with verbs that trigger the periphrastic construction more often (i.e. *order*, *recommend* and *propose*), BrE and IndE align. In other words, IndE is the most conservative variety throughout, even with verbs that strongly lean

towards the use of the subjunctive (see node 2), AmE is the most advanced, whereas BrE takes up an intermediate position. The late selection of the factors ‘person’ and ‘verb’ (nodes 14 and 15) indicates that these factors are less important than ‘trigger’ and ‘variety’, and fits in well with the random forest analysis. Finally, ‘negation’ is not selected in the single tree, either.

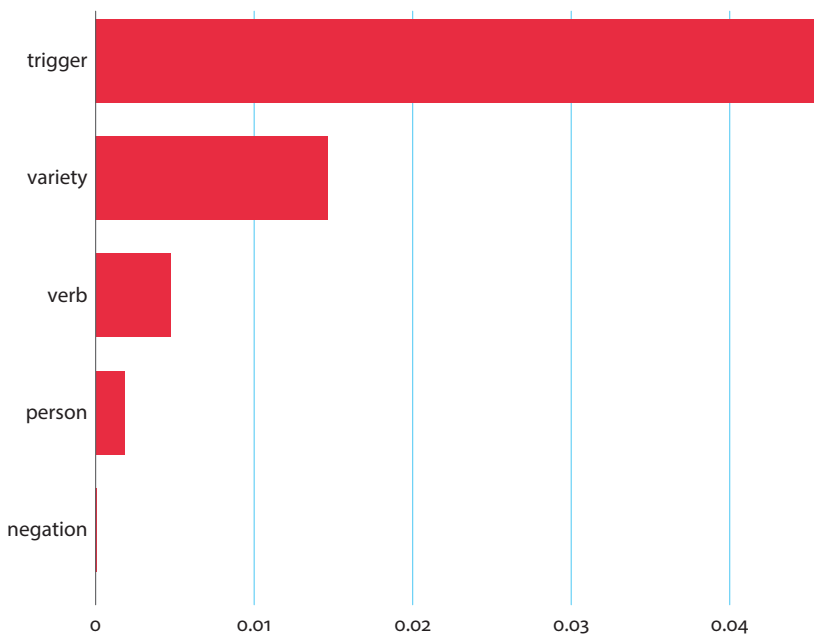


Figure 7. Variable factor importance (Random Forest Analysis) for the GloWbE data

## 5. Discussion: Mandative subjunctives and models of World Englishes

An important finding of the present study is that, on the basis of the ICE data, there is no significant effect of register on the choice between a subjunctive and the periphrastic construction with *should*. This fits in with previous research (Szmrecsanyi et al. 2016: 133), which also found that ‘variety’ consistently ranked higher than ‘register’ as a variable predicting the choice for the dative and genitive alternations as well as particle placement across World Englishes. The fact that variability is strongly determined by the lexical trigger also fits in well with previous research in that lexico-grammatical variation typically shows up as an indicator of indigenization (see e.g. Schneider 2007), in this case ‘probabilistic indigenization’ in the sense of Szmrecsanyi et al. (2016):

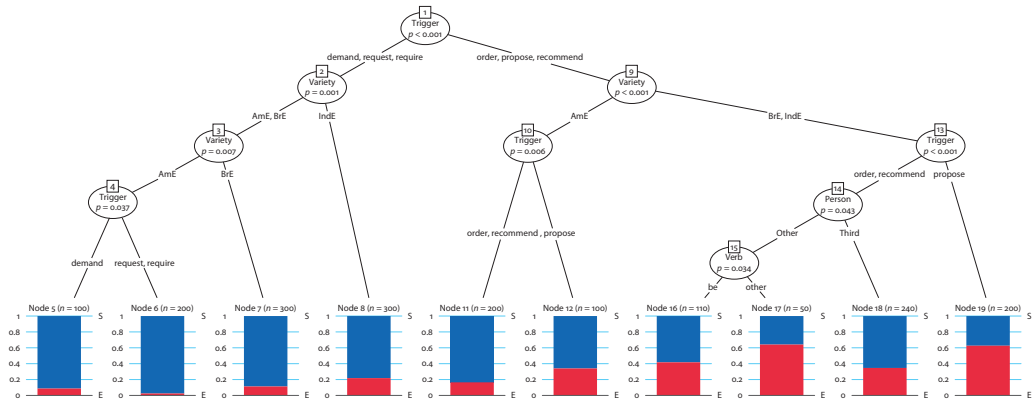


Figure 8. Plotting variable importance (trigger and variety) in GloWbE data

... the process whereby stochastic patterns of internal linguistic variation are re-shaped by shifting usage frequencies in speakers of post-colonial varieties. To the extent that patterns of variation in a new variety A, e.g. the probability of item x in context y, can be shown to differ from those of the mother variety, we can say that the new pattern represents a novel, if gradient, development in the grammar of A. These patterns need not be consistent or stable ..., but they nonetheless reflect the emergence of a unique, region-specific grammar.

The results from the ICE corpora reveal that variation in mandative sentences cuts across ENL, ESL and ESD varieties:<sup>14</sup> SL IndE aligns closer to BrE than to SingE, another SL variety, for instance. While the tripartite distinction does not help explain the variability in this area of morphosyntax, neither does a genetic model (e.g. Strevens 1992), that groups varieties into how they derive from BrE and AmE as matrilects. Variability in the use of mandative subjunctives also cuts across these 'genetic' distinctions: neither NZE nor SingE align with their 'matrilect' (BrE).

Theoretically, language contact might provide an explanation as to why the subjunctive is preferred in SingE: since the mandative subjunctive is identical in form with the base form of the verb, speakers with a variety of Chinese as their first language might prefer unmarked verbs seeing that Chinese does not mark the subjunctive on the verb. The subjunctive, in this interpretation, would also fit in well with a general tendency in contact varieties of English towards simplification. However, this seemingly straight-forward explanation starts to crumble when we look at HKE, which aligns with BrE in its preference for a periphrastic *should* construction in mandative sentences, despite the fact that it shares its main substrate (varieties of Chinese) with SingE.

Traditional models of WEs have been criticized for being rather static. So an obvious question is whether a more dynamic model, such as Schneider's (2007) is better suited to account for the variability found in mandative subjunctive use. Figure 9 lists the postcolonial varieties according to the developmental stage they occupy.<sup>15</sup>

14. The distinction, according to Schneider (2017: 39), was introduced by Quirk et al. (1972).

15. Irish English is not typically discussed as an example of a postcolonial English within Schneider's Dynamic Model, and does not seem to fit it easily, among other things because of its much longer history of colonization (i.e. since the middle ages). Ronan (2016) argues that IrE can be said to have reached Phase 5 (differentiation), following a somewhat different trajectory of change. While SingE may have progressed further towards stage 5 and PhilE towards stage 4 since the publication of Schneider's (2007) monograph, data collection for the ICE corpora took place in the last decades of the twentieth century, making the original classification in Schneider (2007) a suitable point of comparison.

	Phase 2	Phase 3	Phase 4	Phase 5
CanE				-----
NZE				-----
AusE				-----
IrE				-----
SingE				-----
HKE			-----	
PhilE	-----			
IndE			-----	
JamE				-----

(CanE = Canadian English; NZE = New Zealand English; AusE = Australian English; IrE = Irish English; SingE = Singapore English; HKE = Hong Kong English; PhilE = Philippine English; IndE = Indian English; JamE = Jamaican English; classification on the basis of Schneider 2007 and Ronan 2016)

**Figure 9.** Developmental phase of post-colonial varieties of English according to Schneider's (2007) model

One of the basic tenets of Schneider's model is that, as a variety progresses along the developmental cycle, it becomes nativized and thus diverges, structurally, from its matrilect. For lexicogrammatical features (e.g. collostructional variation), greater distance from the input variety has been shown to tally well with developmental stage (see e.g. Mukherjee & Gries 2009), albeit with register as a factor that is likely to add 'noise' to recorded variation in that nativization is more palpable in spoken than in written registers (see Gries & Mukherjee 2010: 542 or van Rooy 2010).<sup>16</sup> The model would account for the relatively conservative nature of HKE, PhilE and IndE (as adhering to their respective matrilects during stage 3)<sup>17</sup> as well as the divergence of SingE away from the BrE model (during stage 4). However, we would also expect to see divergent development for IrE (phase 5) and JamE (phase 4), but both varieties closely align with their respective matrilects.

AmE, the result of what Mesthrie (2006: 388) calls 'the third crossing', has moved beyond phase 5 in Schneider's Dynamic Model and risen to a global standard competing with the original matrilect from the British Isles: "globalization seems to be propelling US English into a position as a potential rival to standard

16. We saw above that the spoken vs. written mode do not play a significant role in this case.

17. Sedlatschek (2009: 287) reports that local textbooks and usage guides explicitly endorse the periphrastic construction and warn against the modal variant as being too formal.

southern British English” (ibid.). An alternative approach might therefore be to simply view the observed variation as a result of Americanisation, an interpretation which would also fit in with Mair’s (2013) model of World Englishes in which AmE serves as a hub variety for ongoing change across Englishes.<sup>18</sup> Americanisation has been proposed as an explanation in previous research on the mandative subjunctive, e.g. to explain the increase in BrE (Kjellmer 2009: 256). It has also been claimed to play a role, more generally, in ongoing change in CanE (e.g. Boberg 2004), AusE (e.g. Collins 2009) and NZE (e.g. Bayard 1999; Green & Bayard 2000). While for SingE, Americanisation has been observed in the areas of phonology (Tan 2015) or spelling and vocabulary (Hänsel & Deuber 2013), the case seems to be less clear for a variety like IndE (see e.g. Cowie 2007 on the development of a regionally ‘neutral’ accent in call centres).

One problem with the concept of Americanisation is that it is difficult to verify, at least if defined as a process that can be traced to speakers’ conscious choices. Vine (1999), for instance, shows that the use of ‘American’ lexical items in New Zealand does not necessarily go hand in hand with speakers’ awareness of them as being Americanisms. Similarly, Hundt (1998: 94) reports on evidence from a small-scale elicitation experiment that showed how informants in New Zealand were unaware of the subjunctives’ association with AmE. This confirmed a previous claim by Algeo (1992: 603–604), who suspected that speakers were unlikely to be aware of the subjunctive’s regional association with AmE and the periphrastic construction’s connection with BrE:

Americans are aware that the British talk differently, but if asked to specify particular forms of difference, few could cite any more than a few hoary old chestnuts like British *lift* for *elevator*. That any except grammarians would have an awareness of the mandative subjunctive is highly improbable.<sup>19</sup>

Thus, while ‘Americanisation’ might be an attractive explanatory concept for the distribution of variants in mandative sentences across the varieties in ICE, it is unlikely that speakers in Australia, New Zealand and Singapore are either aware

---

18. This is the more wide-spread use of the term. Mufwene (2009: 365) uses it to refer to the indigenization of English in North America: “I am using the term **Americanization** here in the sense of ‘becoming American in character’. In the case of English, it means becoming different from British varieties by acquiring characteristics that make it particularly American.”

19. Note that, while there is some discussion of the mandative’s association with AmE on some language-related websites, such as the *English Language & Usage* forum or the advice page *Perfect your English*, these can be considered ‘expert’ discussions of the topic; see <<https://english.stackexchange.com/questions/76578/why-is-american-english-so-wedded-to-the-subjunctive>> and <<http://www.perfectyourengish.com/writing/american-and-british-grammar.htm>>.

of this as being an Americanism nor that they are consciously adopting the mandative subjunctive in order to sound more American. The question is whether 'Americanisation' is therefore the right choice of term, in the first place.<sup>20</sup> Moreover, Americanisation – if at work – might be working alongside other areal factors. Peters (2009: 135), for instance, accounts for the relatively high proportion of mandative subjunctives in both PhlE and SingE as being possibly the result of regional alliances (i.e. SEAMEO, the South-East Asian Ministers of Education Organization). If anything, this possibility would lend support to yet another model of World Englishes as a network of local centres which speakers might be relating to. However, empirically verifying norm orientation to a more local standard than the traditional metropolitan standards of BrE and AmE using corpus data as the only source of evidence is just as difficult as proving 'Americanisation' (see e.g. Hundt 2013).

## Acknowledgement

This paper is dedicated to Christian Mair on the occasion of his 60th birthday.

I would like to thank Nina Benisowitsch and Carlos Hartmann for help with data extraction at various stages of the project.

## Corpus

ICE. *The International Corpus of English*. <http://www.ice-corpora.uzh.ch>.

GloWbE. *Corpus of Global Web-based English*. <<http://corpus.byu.edu/glowbe/>>.

## References

- Albakry, M. & Crawford, W. 2004. Subjunctive triggers in American English newspapers. Paper presented at the *American Association for Applied Linguistics*, Portland, OR.
- Algeo, J. 1992. British and American mandative constructions. In *Language and Civilization: A Concerted Profusion of Essays and Studies in Honour of Otto Hietsch*, C. Blank (ed.), 599–617. Frankfurt: Peter Lang.
- Bayard, D. 1999. Getting in a flap or turning off the tap in Dunedin? Stylistic variation in New Zealand English intervocalic (-T-). *English World-Wide* 20(1): 125–155.  
<https://doi.org/10.1075/eww.20.1.05bay>

---

20. For a careful discussion of the methodology and theoretical underpinnings needed to 'prove' Americanisation/globalisation, see Meyerhoff & Niedzielski (2003).



- Boberg, C. 2004. English in Canada: Phonology. In *A Handbook of Varieties of English*, Vol. 1: Phonology, B. Kortmann, E. Schneider, K. Burridge, R. Mesthrie & C. Upton (eds), 351–365. Berlin: De Gruyter.
- Collins, P. 2009. Modals and quasi-modals. In *Comparative Studies in Australian and New Zealand English: Grammar and Beyond* [Varieties of English Around the World G39], P. Peters, P. Collins, & A. Smith (eds), 73–87. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g39.05col>
- Collins, P. 2015. Diachronic variation in the grammar of Australian English. Corpus-based explorations. In *Grammatical Change in English World-Wide* [Studies in Corpus Linguistics 67], P. Peters, P. Collins, & A. Smith (eds), 15–42. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.67.02col>
- Collins, P., Macalinga Borlongan, A., Lim, J. H. & Yao, X. 2014. The subjunctive mood in English. A diachronic analysis. In *Contact, Variation and Change in the History of English* [Studies in Language Companion Series 159], S. E. Pfenninger, O. Timofeeva, A. C. Gardner, A. Honkaphoja, M. Hundt, & D. Schreier (eds), 259–280. Amsterdam: John Benjamins.
- Cowie, C. 2007. The accents of outsourcing: The meanings of ‘neutral’ in the Indian call centre industry. *World Englishes* 26(3): 316–330. <https://doi.org/10.1111/j.1467-971X.2007.00511.x>
- Crawford, W. J. 2009. The mandative subjunctive. In *One Language, Two Grammars? Differences between British and American English*, G. Rohdenburg & J. Schlüter (eds), 257–276. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511551970.015>
- Davies, M. & Fuchs, R. 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. <https://doi.org/10.1075/eww.36.1.01dav>
- Denison, D. 1998. Syntax. In *The Cambridge History of the English Language*, Vol. IV, 1776–1997, S. Romaine (ed.), 92–329. Cambridge: CUP.
- Fowler, H. 1926. *A Dictionary of Modern English Usage*. Oxford: Clarendon Press.
- Green, J. A. & Bayard, D. 2000. Zed to zee: Notes from Dunedin on the Americanisation of the NZE lexicon through time. *Te Reo: Journal of the Linguistics Society of New Zealand* 43: 63–70.
- Greenbaum, S. (ed.). 1996. *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon.
- Gries, S. T. & Mukherjee, J. 2010. Lexical gravity across varieties of English. An ICE-based study of n-grams in Asian Englishes. *International Journal of Corpus Linguistics* 15(4): 520–548. <https://doi.org/10.1075/ijcl.15.4.04gri>
- Hänsel, E. C. & Deuber, D. 2013. Globalization, postcolonial Englishes, and the English language press in Kenya, Singapore, and Trinidad and Tobago. *World Englishes* 32(3): 338–357. <https://doi.org/10.1111/weng.12035>
- Harsh, W. 1968. *The Subjunctive in English*. Tuscaloosa AL: University of Alabama Press.
- Hoffmann, S. 1997. Mandative Sentences. A Study of Variation on the Basis of the British National Corpus. Lizentiats-Thesis, University of Zurich.
- Hundt, M. 1998. It is important that this study (*should*) be based on the analysis of parallel corpora: on the use of the mandative subjunctive in four varieties of English. In *The Major Varieties of English*, L. Hans, S. Klintborg, M. Levin and M. Estling (eds), 159–175. Växjö: Växjö University.
- Hundt, M. 2013. The diversification of English: Old, new and emerging epicentres. In *English as a Contact Language*, D. Schreier & M. Hundt (eds), 182–203. Cambridge: CUP.
- Hundt, M. & Gardner, A. C. 2017. Corpus-based approaches: Watching English change. In *English Historical Linguistics: Approaches and Perspectives*, L. Brinton (ed.), 96–130. Cambridge: CUP.

- Johansson, S. & Norheim, E. H. 1988. The subjunctive in British and American English. *ICAME Journal* 12: 27–36.
- Kastronic, L. & Poplack, S. 2014. The (north) American English mandative subjunctive in the 21st century: Revival or remnant? *University of Pennsylvania Working Papers in Linguistics* 20(2): 71–80.
- Kjellmer, G. 2009. The revived subjunctive. In *One Language, Two Grammars? Differences between British and American English*, G. Rohdenburg & J. Schlüter (eds), 246–256. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511551970.014>
- Leech, G., Hundt, M., Mair, C. & Smith, N. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511642210>
- Mair, C. 2006. *Twentieth-Century English: History, Variation and Standardization*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511486951>
- Mair, C. 2013. The world system of Englishes. Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34(3): 253–278. <https://doi.org/10.1075/eww.34.3.01mai>
- Mair, C. 2017. World Englishes and corpora. In *The Oxford Handbook of World Englishes*, M. Filppula, J. Klemola, & D. Sharma (eds), 103–122. Oxford: OUP.
- Mesthrie, R. 2006. World Englishes and the multilingual history of English. *World Englishes* 25: 381–390. <https://doi.org/10.1111/j.1467-971X.2006.00477.x>
- Meyerhoff, M. & Niedzielski, N. 2003. The globalisation of vernacular variation. *Journal of Sociolinguistics* 7(4): 534–555.
- Mufwene, S. S. 2009. The indigenization of English in North America. In *World Englishes – Problems, Properties and Prospects*, T. Hoffmann & L. Siebers (eds), 353–368. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g40.21muf>
- Mukherjee, J. & Gries, S. T. 2009. Collostructional nativisation in New Englishes: Verb-construction associations in the International Corpus of English. *English World-Wide* 30(1): 27–51. <https://doi.org/10.1075/eww.30.1.03muk>
- Övergaard, G. 1995. *The Mandative Subjunctive in American and British English in the 20th Century*. Uppsala: Almqvist and Wiksell.
- Peters, P. 1998. The survival of the subjunctive: Evidence of its use in Australia and elsewhere. *English World-Wide* 19: 87–103. <https://doi.org/10.1075/eww.19.1.06pet>
- Peters, P. 2009. The mandative subjunctive in spoken English. In *Comparative Studies in Australian and New Zealand English: Grammar and Beyond* [Varieties of English Around the World G39], P. Peters, P. Collins, & A. Smith (eds), 125–137. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g39.08pet>
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1972. *A Grammar of Contemporary English*. London: Longman.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985 *A Comprehensive Grammar of the English Language*. London: Longman.
- Ronan, P. 2016. Language relations in early Ireland. In *Sociolinguistics in Ireland*, R. Hickey (ed.), 133–153. Cambridge: CUP.
- Sayder, S. 1989. The subjunctive in Indian, British and American English: A corpus-based study. *Linguistische Arbeiten* 69: 58–66.
- Schneider, E. W. 2000. Corpus linguistics in the Asian context: Exemplary analyses of the Kolhapur Corpus of Indian English. In *Parangal kang Brother Andrew: Festschrift for Andrew Gonzalez*, M. L. Bautista, T. A. Llamzon, & B. P. Sibayan (eds), 115–137. Manila: Linguistic Society of the Philippines.

- Schneider, E. W. 2005. The subjunctive in Philippine English. In *Linguistics and Language Education in the Philippines and Beyond: A Festschrift in Honor for Ma. Lourdes S. Bautista*, D. T. Dayag & S. Quackenbush (eds), 27–40. Manila: Linguistic Society of the Philippines.
- Schneider, E. W. 2007. *Postcolonial English. Varieties Around the World*. Cambridge: CUP.  
<https://doi.org/10.1017/CBO9780511618901>
- Schneider, E. W. 2011. The subjunctive in Philippine English – An updated assessment. In *Studies in Philippine English: Exploring the Philippine Component of the International Corpus of English*, M. L. S. Bautista (ed.), 159–173. Mandaluyong City: Anvil Publishing.
- Schneider, E. W. 2017. Models of English in the world. In *The Oxford Handbook of World Englishes*, M. Filppula, J. Klemola, & D. Sharma (eds), 35–57. Oxford: OUP.
- Sedlatschek, A. 2009. *Contemporary Indian English. Variation and Change* [Varieties of English Around the World G38]. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g38>
- Stevens, P. 1992. English as an international language: directions in the 1990s. In *The Other Tongue. English Across Cultures*, B. B. Kachru (ed.), 27–47. Urbana IL: University of Illinois Press.
- Strobl, C., Hothorn, T. & Zeileis, A. 2009. Party on! A new, conditional variable-important measure for random forests available in the party package. *The R Journal* 1(2): 14–17.
- Strobl, C., Malley, J. & Tutz, G. 2009. Introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging and random forests. *Psychological Methods* 14(4): 323–348. <https://doi.org/10.1037/a0016973>
- Szmrecsanyi, B., Grafmiller, J., Heller, B. & Röthlisberger, M. 2016. Around the world in three alternations: Modelling syntactic variation in varieties of English. *English World-Wide* 37(2): 109–137. <https://doi.org/10.1075/eww.37.2.015zm>
- Tagliamonte, S. & Baayen, H. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24: 135–178.  
<https://doi.org/10.1017/S0954394512000129>
- Turner, J. F. 1980. The marked subjunctive in contemporary English. *Studia Neophilologica* 52: 271–277. <https://doi.org/10.1080/00393278008587777>
- Tan, Y-Y. 2015. The Americanization of the phonology of Asian Englishes: Evidence from Singapore. In *Communicating with Asia: The Future of English as a Global Language*, G. Leitner, A. Hashim, & H-G. Wolf (eds), 120–134. Cambridge: CUP.
- van Rooy, B. 2010. Societal and linguistic perspectives on variability in World Englishes. *World Englishes* 29(1): 3–20. <https://doi.org/10.1111/j.1467-971X.2009.01621.x>
- Vine, B. 1999. Americanisms in the New Zealand English lexicon. *World Englishes* 18(1): 13–22.  
<https://doi.org/10.1111/1467-971X.00118>
- Waller, T. 2017. Studies of the subjunctive in present-day English: A critical analysis of recent research, leading to a new diachronic investigation of the mandative subjunctive. PhD Dissertation, University College London.

# A corpus-linguistic account of the history of the genitive alternation in Singapore English

Stefan Th. Gries<sup>1,2</sup>, Tobias Bernaisch<sup>2</sup> and Benedikt Heller<sup>2</sup>

<sup>1</sup>University of California at Santa Barbara / <sup>2</sup>Justus Liebig University Giessen

In this paper, we are exploring the history of the genitive alternation (*of-* vs. *s-*genitive) in Singapore English based on corpus data covering both British English (as the historical input variety) and Singapore English (as the target variety whose diachronic development we are interested in). Specifically, while earlier research has produced partly diachronic accounts of genitive variability, the diachronic development of the genitive has so far not been studied in ESL contexts, a gap which this study attempts to fill. Nearly 7000 instances of *of-* and *s-*genitives were annotated for a large number of predictors including phonetic variables (e.g. final sibilancy of possessor), semantic variables (e.g. animacy of possessor/possessum), syntactic variables (e.g. length of possessor/possessum), and pragmatic variables (e.g. discourse accessibility of possessor/possessum). We then applied the method of Multifactorial Prediction and Deviation Analysis with Regressions/Random Forests to the data to explore (i) how genitive choices in Singapore English differ from those in British English and, after a methodological interlude, (ii) how genitive choices changed over time in Singapore English. We conclude with some important recommendations regarding diachronic studies of structural nativization and their theoretical implications in models such as those of Moag (1982) or Schneider (2003, 2007).

**Keywords:** genitive alternation, Singapore English, diachrony, probabilistic grammar, MuPDAR

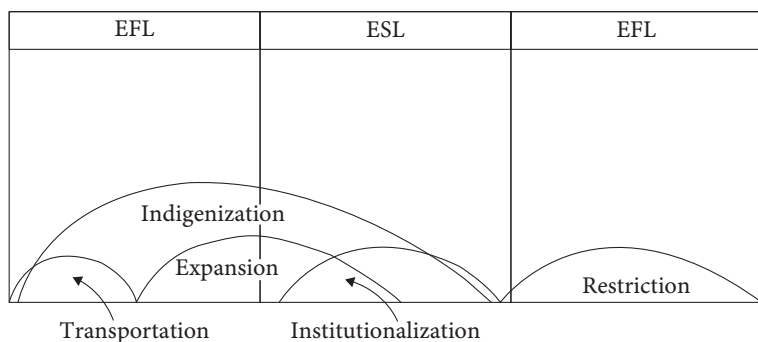
## 1. Introduction

The introductory sections offer theoretical as well as methodological perspectives on the study of diachronic developments in World Englishes. Section 1.1 focuses on relevant diachronic models – particularly Moag (1982, 1992) and Schneider (2003, 2007) – representing the evolution of (some native, but mainly) non-native

varieties of English world-wide, whereas Section 1.2 provides an overview of how this evolution has so far been tackled methodologically.

### 1.1 Theoretical considerations

Although Chinese, French, Spanish, Malay – among others – have diversified into additional national varieties outside the languages’ original homelands, the spread of English across national boundaries all around the globe and the resulting number of English speakers has up to this point not been matched by any other language. The process of transporting or transplanting English to a new territory is generally well-documented with historical evidence and only with a few exceptions as in the cases of Australia or what is the United States today, this re-rooting of English can be considered a linguistic by-product of an economically-driven quest for natural resources and trade monopoly of the British crown via the British East India Company. Despite the dearth of historical linguistic evidence for the respective diachronic diversification of English into national varieties, scholars in the World Englishes paradigm (e.g. Strevens’s (1980) world map of English, Kachru’s (1985) three-circle model, McArthur’s (1987) circle of World English) have repeatedly dedicated themselves to the static depiction of the outcomes of this diversification process, but Moag (1982, 1992) – inspired by Hall’s (1962) work on the cyclic evolution of pidgin languages – is the first explicitly dynamic model on New Englishes describing the “Life cycle of non-native Englishes” illustrated in Figure 1.



**Figure 1.** Moag’s (1982) life cycle of non-native Englishes

Five consecutively ordered, but generally overlapping processes constitute Moag’s (1982, 1992) life cycle: (a) transportation – English is brought “into a new environment for purposes of a more or less permanent nature” (Moag 1982: 271); (b) indigenization – the structurally layered and initially slow localization of the historical

input to create a distinct regional variety; (c) expansion in use and function – the spread of English to domains formerly reserved for local languages; (d) institutionalization – via the localization of English literature, English language teaching and the media and (e) restriction of use and function – the abandoning of the newly emerged English in favor of another legislated official language often as a symbol and result of political independence. This last process, however, is not applicable to all New Englishes, but, according to Moag's (1982: 283) judgment in the early 1980s, "may be in the cards for Malaysia, the Philippines, and perhaps even India". Moag's (1982: 271) cycle alludes to changing roles of "English-speaking aliens and some segment of the local population in the domains English occupies and in the norm orientation for the newly emerging variety. With a focus on linguistic structures and their developments, initial lexical borrowings and later grammatical innovations broadly characterize the succession through the life cycle, which is accompanied by a continuous spread of English usage from public to private social domains (cf. Moag 1982: 273). The life-cycle model is illustrated with cases in point from English in the Pacific region with a particular focus on Fiji. In a nutshell, Moag's (1982, 1992) life cycle suggests that emerging varieties of English develop from English-as-a-foreign-language (EFL) to English-as-a-second-language (ESL) varieties and then potentially return to EFL status facilitated by post-independence language planning promoting local national languages in official settings to the detriment of the status of English.

Schneider, the proponent of the probably most influential model in World Englishes so far, acknowledges and comments on the relationship between Moag's life cycle and his (2003, 2007) dynamic model of the evolution of postcolonial Englishes:

[f]or an earlier, comparable model, similar in some respects though different in others and considerably more constrained in its applicability, see Moag (1992) in relationship to the situation in Fiji. Moag distinguishes four overlapping phases, called 'transportation,' 'indigenization,' 'expansion in use and function,' and 'institutionalization,' sometimes followed by a fifth phase, 'restriction of use and function.' Perhaps the most important difference between Moag's idea and the present model is that he believes that in the end English typically tends to revert to a foreign-language status. (Schneider 2007: 319)

While it would be insightful to explore in which areas the scope of Moag's (1982, 1992) model is more restricted than that of Schneider's, Schneider's above statement must be considered too humble since it understates some central conceptual advances his evolutionary model presents (e.g. more rigorous focus on speech communities in a given territory, fully elaborated social as well as linguistic characteristics for each developmental stage, etc.). Still, the implication that noteworthy parallels across the two models exist is certainly correct. Schneider's (2003, 2007)

model also operates on five diachronic phases, i.e. (1) foundation – settlers bring English to a new territory; (2) exonormative stabilization – English conforms to (mostly British English) non-local norms, (3) nativization – English develops local characteristics, (4) endonormative stabilization – local characteristics of English become the norm and (5) differentiation – (regional) dialects of a given postcolonial English emerge, which – except for the last phase – successfully advance particular, roughly conceptualized facets from Moag (1982, 1992) and Hall (1962) using a distinct terminology. With regard to the succession of these phases, Schneider (2003, 2007) adds one source of universal appeal to his model (see Schneider 2014 for an overview of the reception of the model and its applications), which is his unique proposition that

the difference between phases 3 and 4 is commonly given symbolic expression by substituting a label of the ‘English in X’ type by a newly coined ‘X English.’ The former marks the dialect as just a variant without a discrete character of its own; while the latter credits it with the status of a distinct type, set apart from and essentially on equal terms with all others. (Schneider 2007: 50)

The ensuing checklist-like operationalization of independent variety status of any postcolonial English (cf. Schneider 2007: 56) was and certainly still is a welcome invitation to researchers to examine the evolutionary status of the variety they are concerned with. The evolutionary status so established is generally considered an indication with regard to which characteristics the variety scrutinized can be seen as a full-fledged variety of English on a par with other World Englishes. Further, the descriptive parameters for developmental stages, which are, however, only systematically applied and adapted to each stage by Schneider (cf. 2007: 56), generally conform with Moag (1982, 1992), but by no means completely coincide. To name just one of several examples, both models include perspectives on the indigenous and the settler communities in a given territory, but Schneider stresses the need to take the attitudes of the two strands of communicative perspective into account, while this is not taken into consideration by Moag (1982, 1992). The descriptive characteristics are presented as follows:

[a]t each of these stages, manifestations of four different parameters can be observed and will be pointed out, with a monodirectional causal relationship operating between them: (1) Extralinguistic factors, like historical events and the political situation, result in (2) characteristic identity constructions on the sides of the parties involved. These, in turn, manifest themselves in (3) sociolinguistic determinants of the contact setting (conditions of language contact, language use, and language attitudes), which, consequently, cause specific (4) structural effects to emerge in the form(s) of the language variety/-ies involved.

(Schneider 2007: 30–31)



The above statement presents additional conceptual progress in that Schneider (2003, 2007) is more explicitly a linguistic model than Moag (1982, 1992). Schneider (cf. 2007: 30–31) argues that the three mainly sociohistorical parameters, i.e. extralinguistic factors, identity constructions and sociolinguistic determinants, eventually find reflection in linguistic/structural effects. In essence, Schneider (2003, 2007) implies that linguistic structures are indicators of varietal progress in his evolutionary cycle.

Given this salience of structural investigations for the determination of evolutionary progress, it is unexpected that an operationalization of this determination does not figure more prominently in models in World Englishes. Moag (1982, 1992) – maybe not surprisingly – does not offer comments in this regard since structural effects do not take center stage in his developmental cycle, and corpus-linguistic data for World Englishes were by-and-large absent at the beginning of the 1980s. True, more recent models on Englishes world-wide such as Mair (2013) do in fact formulate empirically testable model assumptions. According to Mair (2013: 260), “the hub of the ‘World System of Englishes’ is Standard American English”, whose international leading role he (cf. 2013: 261–262) expects to be reflected in Americanisms being “a massive presence in practically all other varieties, including British English. Criticisms will be found in American usage, but to a far lesser extent.” In contrast, however, a similar or maybe even more rigorous empirical operationalization of how to determine evolutionary status via structural analyses is not available in Schneider (2003, 2007). In this light, Schneider (2004: 227) suggests that

[t]he most promising road to a possible detection of early traces of distinctive features is a principled comparison of performance data collected along similar lines, i.e. systematically elicited corpora. [...] [T]he International Corpus of English project (Greenbaum 1996) promises to provide a uniquely suitable database for such comparative investigations [...].

This proposition establishes an implicit link between the dynamic model, which – although the term has been used by others (cf. Kachru *passim*) – features structural nativization as the key notion in structural varietal developments, and corpus-linguistic investigations. While Schneider (2004) does not explicitly relate corpus-linguistic results to evolutionary stages in the dynamic model, the study nevertheless sets the methodological tone for future corpus-linguistic applications of his evolutionary model; in particular for the assumption that – in the absence of authentic historical corpus data – synchronic corpus-linguistic cross-varietal differences can be interpreted as structural representations of diachronic change in World Englishes. Corpus linguists have since related structural differences of postcolonial Englishes (often in comparison to their historical input varieties) to



evolutionary status in the dynamic model with the help of a number of different methodologies, which will be summarily presented in the next subsection.

## 1.2 Methodological considerations

Corpus linguists who were concerned with the structural pillar of Schneider's model have focused on two interrelated aspects in the modeling of the evolution of postcolonial Englishes: (i) how to study the diachronic evolution of varieties postulated in Schneider's model, and (ii) how to do this quantitatively.

As for the former, much work (e.g. Mukherjee 2008; Bernaisch 2015; Edwards 2016) has adopted the logic that we can make comparisons between data from a source variety (e.g. British English) and a target variety (e.g. Singaporean English) such that the amount and the nature of the differences found will allow us to determine the position of the target variety with regard to Schneider's five evolutionary stages, i.e. how (much) the target variety has emancipated itself from the source variety, or become nativized. For example, Mukherjee and Gries (2009) analyzed collostructional routines in Asian Englishes and established the lexicogrammatical distinctness of Indian and Singapore English when compared to British English, which is easily reconcilable with their relatively advanced evolutionary statuses. Still, it is important to point out that a number of studies as documented in Collins (ed., 2015) and Noël et al. (eds, 2014) have already engaged in truly diachronic corpus studies of World Englishes with a particular focus on lexicogrammatical features. A comparatively large number of these studies, however, relied analytically on frequency-based comparisons of surface structure choices (with or without statistical modelling), whose results and conclusions are unlikely to do full justice to the complexity of the data (see Gries & Deshors 2014: Section 3.1 for a discussion of the problems of such analyses). In contrast, the present paper adopts a multifactorial statistical research design to simultaneously control for various factors potentially influencing the structural choice concerned and thus enabling more detailed descriptions.

As for the latter, three different methodological levels can be distinguished in how the above logic has been applied: First, zero-/monofactorial frequency studies, in which the above kind of comparison is made on the basis of observed absolute or relative frequencies and cross-tabulation and tested via chi-squared or likelihood ratio tests (cf. e.g. Hoffmann, Hundt & Mukherjee 2011; Collins 2012; Huber 2012; Bernaisch 2015). Second, multifactorial classification/regression modeling where the dependent variable is, for instance, a constructional choice and where the independent variables are predictors known or hypothesized to affect, or at least be correlated with, the constructional choice as well as Variety, a variable that allows to

contrast the varieties represented in the data (e.g. the historical source variety/-ies as well as the target variety/-ies and maybe other varieties included for the sake of comparison; cf. e.g. Mukherjee & Gries 2009; Bernaisch, Gries & Mukherjee 2014; Biewer 2015; Deshors 2017); crucially, all predictors of the first kind need to be allowed to statistically interact with Variety so as to determine which predictors' effects differ between varieties.

Third, in recent years Gries and colleagues developed and then used an extension of the second approach called MuPDAR(F), short for *Multifactorial Prediction and Deviation Analysis with Regression/Random Forests* (see Gries & Adelman 2014; Gries & Deshors 2014, 2015; Wulff & Gries 2015; Deshors & Gries 2016; Gries & Bernaisch 2016; Heller, Bernaisch & Gries 2017; Wulff, Gries & Lester forthcoming). This method involves the following three steps:

- fitting a regression/random forest  $R(F)_1$  that predicts the choices that speakers of the source/reference level (typically, native speakers of the reference variety) make with regard to the phenomenon in question;
- applying the results of  $R(F)_1$  to the other/target speakers in the data (typically, learners or speakers of institutionalized second-language varieties) to predict for each of their data points what the native speaker of the source/reference variety would have done in their place;
- fitting a regression/random forest  $R(F)_2$  that explores how the other speakers' choices differ from those of the speakers of the source/reference variety: predictors that are significant in this regression are ones that help understand where the target variety speakers make choices that are not those of the source/reference variety.

In addition, the logic of this has also been used to explore similarities between indigenized varieties and to determine epicenter status in cases where multiple candidate varieties are available (see Gries & Bernaisch 2016; Heller, Bernaisch & Gries 2017).

In this paper, we are using the MuPDAR approach to explore Singaporean English and its emancipation/nativization from British English; as the vehicle of our analysis, we are using the genitive alternation exemplified in (1), i.e. a well-known, ubiquitous, and extensively researched alternation whose well-known characteristics at least in native English serve as a good benchmark:

- (1) a. the President's speech      *s*-genitive:    possessor's possessum  
       b. the speech of the President    *of*-genitive:    the possessum *of* possessor

The next section discusses the specifics of our *MuPDAR* analysis in detail.

## 2. A MuPDAR analysis of SinE 1990 with reference to BrE 1990

### 2.1 Methods

#### 2.1.1 *Generating and annotating the concordance data*

For the present case study, interchangeable genitives were extracted from two components of the International Corpus of English (ICE; Greenbaum 1991), ICE-SIN and ICE-GB, to represent genitive choices in Singapore English (SinE) and British English (BrE), respectively. Due to the high frequency of genitives, a sample of 10% of all text files was taken (i.e., file 1, file 10, ..., file 490, file 500). This sampling strategy both preserves the proportions of the corpus design and facilitates the annotation of previous choice (see below), which records the variant used in the preceding interchangeable genitive; extra corpus material as marked by either <X> ... </X> or marked by Xs in text unit markers was excluded from the analysis.

During the extraction, genitives were regarded as interchangeable if they could, in principle, be phrased in the respective other variant. The criteria that we applied are in accordance with previous variationist studies on the genitive alternation (most recently, Heller et al. 2017; also see Rosenbach 2014 for an exhaustive overview). In essence, this meant excluding all genitives that are either appositive (*the city of London*), partitive (*the high number of students*) classifying (*the old children's book*), double (*the painting of Pete's*), idiomatic/fixed (*Valentine's Day*) or that do not have a definite possessum (*a friend of my brother*).

In sum, 4178 interchangeable genitives were extracted from the two corpora, 3162 from ICE-GB and 1016 from ICE-SIN (Table 1). In both varieties, the *of*-genitive is the most frequent variant, accounting for 62.93% of interchangeable genitives in BrE, and 70.08% in SinE.

**Table 1.** Overview of our data

	<i>of</i> -genitive	<i>s</i> -genitive	Total
British English (BrE)	1990	1172	3162
Singaporean English (SinE)	712	304	1016
<b>Total</b>	<b>2702</b>	<b>1476</b>	<b>4178</b>

After the extraction, all instances were annotated for the following linguistic constraints that are well-known to govern the choice between the *s*-genitive and *of*-genitive: possessor animacy, possessor definiteness, final sibilancy of the possessor, possessor givenness, possessor thematicity, length difference of the possessor and possessum phrase, previous choice, and type-token ratio of the immediate context.

- Possessor animacy is widely considered to be the most important constraint of genitive choice and is reported to be highly significant in every study of which we know. The higher a possessor is on the animacy scale (e.g. *the mayor* < *the administration* < *the plan*), the more likely it is to be realized as an *s*-genitive (Rosenbach 2014: 232).
- Possessor definiteness, similarly, but less strongly so, increases the likelihood of an *s*-genitive realization (e.g. *the mayor* < *some teacher*).
- On the other hand, if a sibilant (one of [s], [z], [ʃ], [tʃ], [ʒ], and [dʒ]) is present at the end of the possessor phrase (as in *peaceful coexistence*, for example), an *s*-genitive realization is less likely due to the phonological conflict of the sibilant and the subsequent genitive marker's (Zwicky 1987).
- Possessor givenness distinguishes whether the current possessor has been mentioned in the previous context or not. Not all studies found givenness to make a significant difference (e.g. Hinrichs & Szmrecsanyi 2007), but if they did (e.g. Grafmiller 2014 in Model 2), *s*-genitives are usually more likely if the possessor has been mentioned before.
- In a similar fashion, high levels of possessor thematicity, the degree to which the possessor in question constitutes a major topic of the text measured here as the relative frequency of mentions of the possessor head lemma throughout the corpus text, also facilitate *s*-genitive use (Osselton 1988).
- Length difference, here defined as length of the possessor phrase minus the length of the possessum phrase in number of characters, allows us to measure the effect of end-weight (Behaghel 1909) in genitive choice. End-weight is usually the second most important constraint in the genitive alternation (see Rosenbach 2005 for an in-depth comparison of the effects of animacy and syntactic weight on the genitive alternation); it causes speakers to use the variant that places the longer constituent after the shorter one. We therefore expect to see more *of*-genitives with positive values (i.e. when the possessor is longer than the possessum) and more *s*-genitives with negative values.
- Previous choice enables us to measure the degree of persistence in genitive choice. Szmrecsanyi (2006) has compared two types of persistence, (i) the influence of a previous mention of the genitive marker *of*, and (ii) the influence of the choice made in the previous interchangeable case. He concluded that in the genitive alternation, the latter (i.e.  $\alpha$ -persistence) is “vastly more powerful” (Szmrecsanyi 2006: 107) than the former, which is why we focus on  $\alpha$ -persistence here.
- Finally, type-token ratio (TTR) of the immediate context (i.e.  $\pm 100$  words) is a measure of lexical density; it has been shown that the *s*-genitive is preferred in lexically dense environments (e.g. Hinrichs & Szmrecsanyi 2007).

A summary of the predictors and their corresponding levels (and, where useful, their frequencies) is provided in Table 2.

### 2.1.2 Statistical analysis

In order to prepare the above data for a MuPDAR analysis, we did some initial exploration of the data. This included tabulating and plotting the data to determine whether variables needed to be transformed or variable levels needed to be conflated, etc.

Table 2. Annotation scheme

Predictor	Levels	Example (where applicable)
Possessor animacy (Animacy)	<i>animate1</i> <i>animate2</i> (animals) <i>collective</i> <i>inanimate</i>	<i>her husband</i> <i>the reef fishes</i> <i>the society</i> <i>the agreement</i>
Possessor definiteness (Definiteness)	<i>definite</i> <i>indefinite</i>	<i>the president</i> <i>poverty or prosperity</i>
Final sibilancy of the possessor (FinSib)	<i>absent</i> <i>present</i>	<i>the government</i> <i>its dominance</i>
Possessor givenness (Givenness)	<i>given</i> <i>new</i>	– –
Possessor thematicity (Thematicity)	0 to 100	–
Length difference (LengthDiff)	–155 to 216	Positive: <i>the [absence] of [a viable alternative]</i> Negative: <i>[China]’s [long-term-prospects]</i>
TTR	0.471 to 0.980	Low TTR: <i>In the [course] of [our conversation] I told him that our Singapore company manufactured picture tubes in Singapore. I told him that the TV sets from Singapore would have picture tubes made in Singapore (ICE-SIN:S2A-061)</i> High TTR: <i>Though news of Windows 3.1’s release may bring plenty of Windows enthusiasts together for some hearty discussion, but your ranking of this product over version 3.0 is really very much dependent upon [the individual]’s [requirements, needs and computing environment]. (ICE-SIN:W2B-031)</i>
Previous choice (PrevChoice)	<i>none</i> vs. <i>of</i> vs. <i>s</i>	–

As a result of this exploration,

- the variable Animacy was recoded to only three levels: *humanimate* (conflating *animate1* and *animate2*), *collective*, and *inanimate*;
- the variable LengthDiff was logged to the base of 2 such that positive values were just logged (i.e., a value of 16 became 4) whereas negative values were converted to positive ones, then logged, and then multiplied with  $-1$  (i.e., a value of  $-16$  became  $-4$ );
- the variable Thematicity was changed to Thematicity<sup>0.25</sup> (because of its extreme skew);

In addition, for  $R_1$  on the BrE90 data we decided to fit all numeric predictors as orthogonal polynomials to the second degree (to allow for the possibility that a numeric predictor's effect is not a straight, but a curved line); also, we added varying intercepts by file to the regression model – since this already led to occasional convergence problems, a more complex random-effects structure was not explored. The regression model we used therefore included all predictors but, for simplicity's sake, no interactions.

The results were quite encouraging: the model was highly significant ( $LR = 1422.2$ ,  $df = 46$ ,  $p < 0.001$ ), accounted for a fairly large proportion of the data ( $R^2_m = 0.54$ ,  $R^2_c = 0.59$ ), and came with a good classification accuracy (accuracy = 82.5%,  $C = 0.9$ ); we therefore proceeded with the analysis and applied the regression model to the SinE90 data (without random effects).

There, too, the model performance was good with a high accuracy (80.2%) and good classificatory power ( $C = 0.86$ ). Thus, in our final step, we created a variable BrElike which captured for every SinE90 choice whether it was what a BrE90 speaker (speaker of BrE in the 1990s) would have been predicted to use and used it as a dependent variable in  $R_2$ . To identify what might be the best model, we employed the following stepwise model-checking procedure:

- our first model involved all main and random effects from  $R_1$ ;
- we then checked (i) all predictors in the model for how much their deletion from the model would improve *AIC* (a widely used statistic to compare regression models) compared to the current model and (ii) all predictors not in the model as well as all their pairwise interactions for how much their addition to the model would improve *AIC* compared to the current model;
- we then deleted or added the predictor that would result in the best improvement of *AIC*, but only if this did not lead to too high degrees of overdispersion or multicollinearity (as operationalized by  $VIFs \geq 15$ ).

The final model we arrived at was again highly significant ( $LR = 240.01$ ,  $df = 38$ ,  $p < 0.001$ ), accounted for a fairly large proportion of the data ( $R_m^2 = 0.55$ ,  $R_c^2 = 0.56$ ), and came with a good classification accuracy (accuracy = 83.7%,  $C = 0.83$ ); incidentally and as is obvious from the above  $R^2$ -values, the varying intercepts contributed virtually nothing to the model: a model with them led to less than 0.5% changed predictions. In the following section, we will interpret selected results from this  $R_2$ .

As in many previous MuPDAR analyses, we are not discussing all significant interactions here to save space and for an important additional reason to be discussed below; the significant interactions we leave out are those we do not return to in Section 3: Animacy  $\times$  LengthDiff, Thematicity  $\times$  FinSib, Givenness  $\times$  Definiteness, Givenness  $\times$  PrevChoice, TTR  $\times$  Definiteness, and Thematicity  $\times$  LengthDiff; the visualizations we provide are plots of predicted probabilities of BrE-like choices (as lines for numeric predictors and as predicted means for categorical predictors on the  $x$ -axis, both with 95% confidence intervals).

## 2.2 Results of $R_2$

### 2.2.1 *FinSib* $\times$ *Animacy*

We begin by discussing the effect of *FinSib*  $\times$  *Animacy*, shown in Figure 2 with *FinSib* on the  $x$ -axis and the levels of *Animacy* represented in different colors of the lines and points. The results show that in the unmarked case of the possessor not ending in a sibilant (i.e. the left part of the plot), *Animacy* does not matter much and the SinE90 speakers behave like the BrE90 speakers most of the time. However, when the possessor does end in a sibilant (i.e. the right part of the plot), then the SinE90 speakers are very BrE90-like only with inanimate possessors, but deviate much more from BrE90 speakers with humanimate and collective possessors, i.e. possessors that are humans/animates or can be metaphorically seen as humans (i.e. the collectives). To fully appreciate these findings, it is instructive to explore  $R_1$  for guidance on what the choices of the BrE speakers are or, in an alternation as well studied as this one, to consider the large amount of previous work.  $R_1$  for BrE in the 1990s predicts *s*-genitive usage if the possessor is humanimate, both when a final sibilant is present or absent. Example (2), which contains a humanimate possessor that ends in a final sibilant, thus constitutes a representative example of 1990s BrE genitive genitive choice.

- (2) Brent councillor Bill Duffin, who followed the case closely, said that he was particularly delighted with the judge's decision. (ICE-GB, W2C-011)

Since Figure 2 shows that speakers of 1990s SinE are more likely than not to deviate from the BrE choice of an *s*-genitive with humanimate possessors with a final sibilant and opt for the *of*-genitive instead, i.e. say/write *he was particularly delighted with the decision of the judge*.

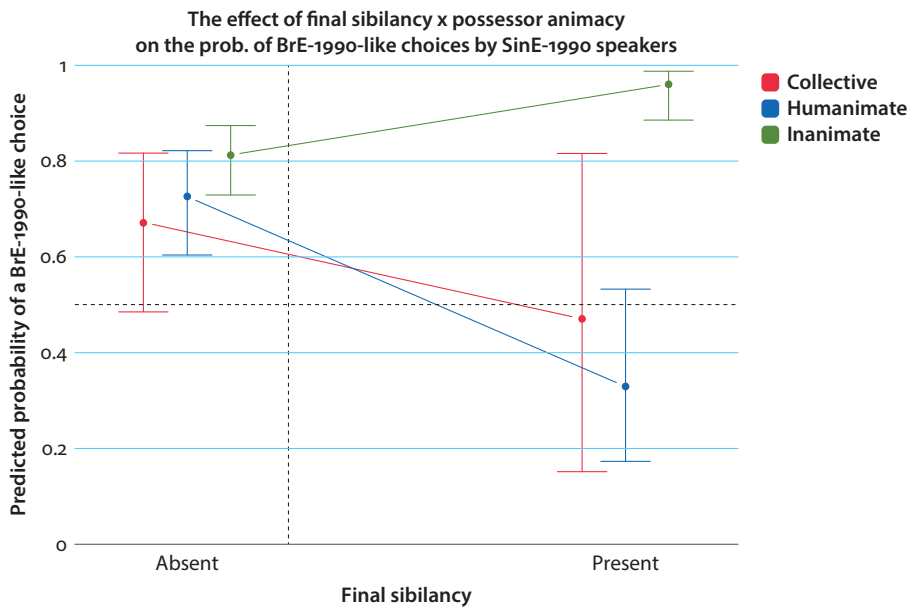


Figure 2. Predicted probabilities of BrE-like choices by SinE speakers: FinSib  $\times$  Animacy

### 2.2.2 *LengthDiff* $\times$ *FinSib*

The next significant interaction to be discussed is *LengthDiff*  $\times$  *FinSib*, which is represented in Figure 3. *LengthDiff* is shown on the *x*-axis (with a vertical dashed line representing its median; the predicted probability of SinE90 speakers making the BrE90-like choice is on the *y*-axis (with a horizontal dashed line at 0.5); the red and turquoise lines and confidence bands represent the predicted probabilities for the combinations of length differences and the pre-/absence of final sibilants; the red and turquoise points around  $y = 0$  and  $y = 1$  represent the non-BrE90-like and BrE90-like choices made by the SinE90 speakers respectively.

Figure 3 shows that, in the unmarked case – i.e. when the possessor does not end in a sibilant – then the SinE90 speakers make BrE90-like choices in particular when the possessor and the possessed differ in length, i.e. when *LengthDiff* gives a strong short-before-long cue, but when *LengthDiff* is around 0, then SinE90 speakers differ from the BrE90 speakers most. However, in the marked case of possessors with final sibilants, SinE90 speakers behave like BrE90 speakers only when the possessor is longer than the possessed (by using *of*-genitives) but they switch to *s*-genitives later than the BrE90 speakers, namely only when the possessum becomes much longer than the possessor. In other words, the presence of a sibilant seems to override the length-based recommendation unless the latter becomes really strong.





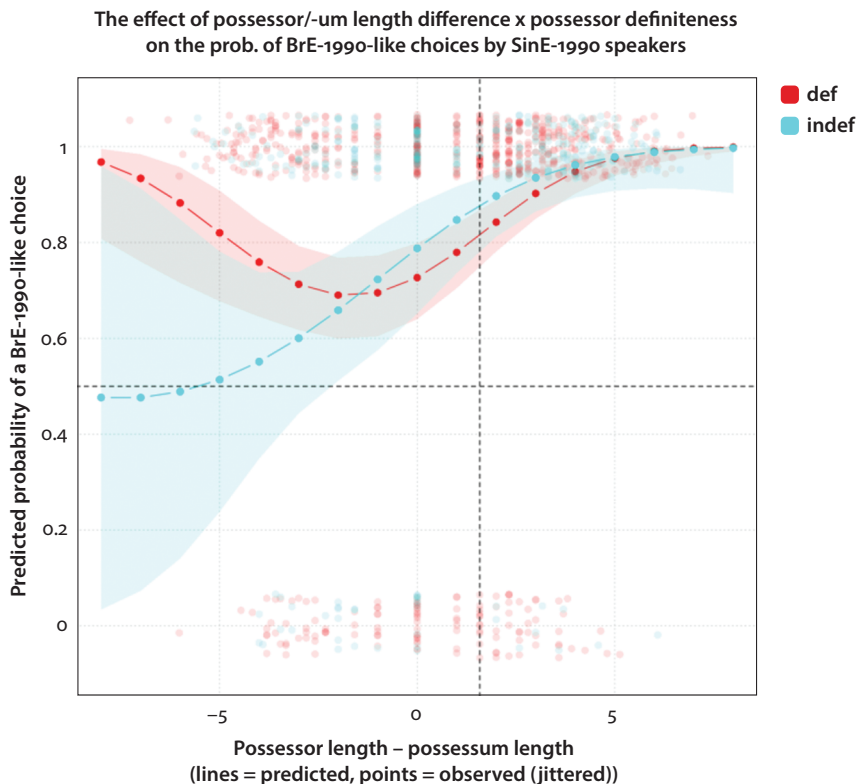
**Figure 3.** Predicted probabilities of BrE-like choices by SinE speakers:  
LengthDiff  $\times$  FinSib

### 2.2.3 *LengthDiff* $\times$ *Definiteness*

Figure 4 shows the corresponding (and incidentally extremely similar) results plot for LengthDiff  $\times$  Definiteness. When the possessor is definite, the SinE90 speakers make BrE90-like choices in particular when the possessor and the possessum differ in length, but not when they are about equally long. However, when the possessor is indefinite, SinE90 speakers again only behave like BrE90 speakers when the possessor is longer than the possessum. Put differently, only when the possessum is considerably longer than the possessor do the SinE speakers respond to short-before-long and switch to the overall less frequent *s*-genitive.

### 2.2.4 *TTR* $\times$ *Animacy*

The final effect to be discussed in this section is TTR  $\times$  Animacy as shown in Figure 5. Even disregarding for the moment the very wide confidence bands, this effect is hard to interpret. Essentially, it shows that SinE90 speakers use collective

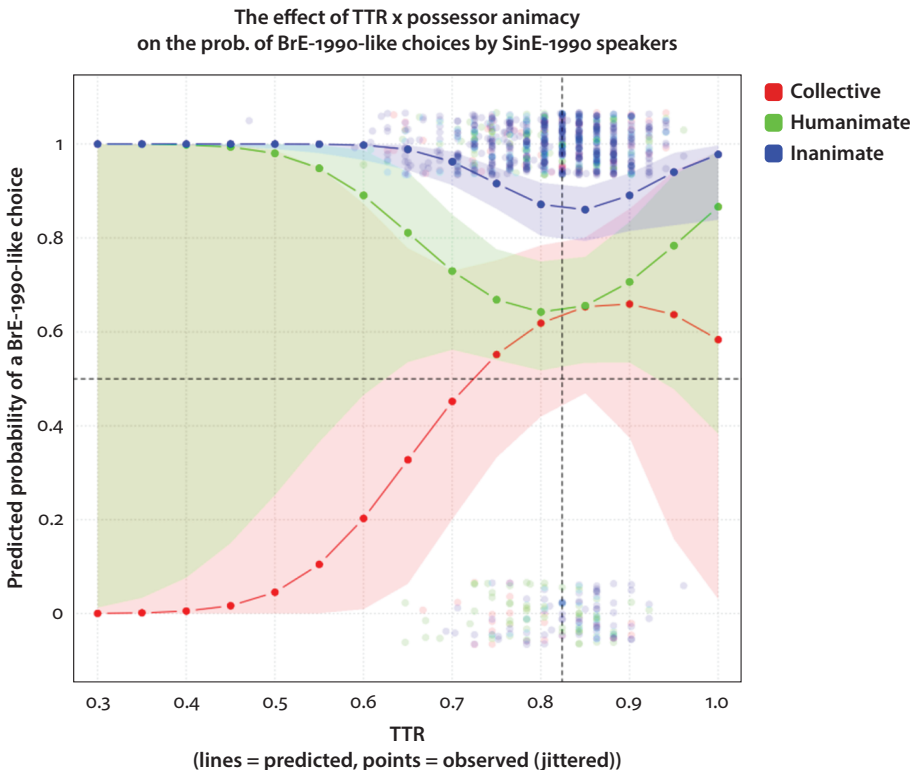


**Figure 4.** Predicted probabilities of BrE-like choices by SinE speakers:  
LengthDiff  $\times$  Definiteness

possessors most like BrE90 speakers in lexically complex texts whereas they use humanimate and inanimate possessors least like BrE90 speakers in texts of average lexical complexity; it is hard to make sense of how this might come about – however, as we will see below, one may not have to ...

### 2.3 Discussion

The analysis as discussed so far provides some clear evidence for how SinE90 speakers' genitive choices differ from those that BrE90 speakers would make. Maybe most notably, they point to different degrees of sensitivity of SinE90 speakers to the tendency of short-before-long, which seems to be stronger in BrE90 speakers but can be rendered less important to the SinE speakers when other factors – such as FinSib or Definiteness – are at their marked/less frequent levels (*present* and *indefinite* respectively). Also, we find that SinE90 speakers are most BrE90-like



**Figure 5.** Predicted probabilities of BrE-like choices by SinE speakers: TTR × Animacy

with inanimate possessors even though Animacy interacts with other predictors. In addition and on a slightly more abstract/methodological level, we find that, for this case at least, speaker-specific variability is surprisingly negligible and that it is necessary to abandon the assumption that numeric predictors' behavior is best modeled with a straight line, which is embedded in most regression analyses in the past but not supported here at all.

A 'normal' paper on structural nativization/indigenization would now probably launch a discussion of how these factors testify to SinE's emancipation from the historical source variety, given the significant differences in the genitive choices of BrE90 and SinE90 speakers. However, we will not pursue this route – rather, we believe it is incumbent upon us to call into question two related working assumptions that most studies of this type – and that includes our own, as we wish to highlight emphatically – have been making, which are conceptually quite similar to the sociolinguistic notion of 'apparent time'.

First, nearly all corpus-linguistic studies of structural nativization in general and of Moag's/Schneider's model in particular have assumed that the diachronic process

of nativization/emancipation of one variety (such as SinE) from another (such as BrE) can be studied or modeled on the basis of synchronic data. In sociolinguistics, the assumption is made that even if we sample language data synchronically, i.e. at one point in time, we can still study diachronic processes by sampling speakers from different age groups. In a similar vein, studies of structural nativization often rely on synchronic corpus data (e.g. from the International Corpus of English, where, however, information on speaker age is only rarely available) and claim to study the diachronic process of nativization because (i) BrE is the *historical* source variety of SinE (i.e. they are the analogue to older speakers in apparent-time sociolinguistic studies) and (ii) the varieties studied are from different stages of Schneider's evolutionary model (i.e. they are the analogue to differently younger speakers in apparent-time sociolinguistic studies).

Second, this first assumption implies the assumption that language patterns are relatively stable after adolescence, which is what allows the different age groups from a synchronic corpus to 'stand in' for real diachronic data. In synchronic structure-oriented studies of postcolonial Englishes, this stability is usually implicitly and generally for the lack of feasible empirical alternatives assigned to the historical input variety BrE since it is assumed that the historical source variety to which postcolonial Englishes are compared has not undergone (substantial/significant) changes during the time period in which the postcolonial English in question 'has been nativizing'. For instance, the historical input variety to what is current SinE is of course not current BrE but the BrE from the 19th century; that in turn means any comparison between current SinE and current BrE profiling structural differences as structural nativization in SinE kind of *has* to rely on the assumption that current BrE is not significantly/substantially different from BrE then.

As with the discussion surrounding apparent-time sociolinguistics and the widespreadness of these two assumptions, they are controversial, to put it mildly, and given the many studies that have shown differences between, say, Brown and Frown or LOB and FLOB, for many phenomena (e.g., Mair 1995, 2002; Hinrichs & Szmrecsanyi 2007), it does not even seem reasonable anymore to try and sell the above two assumptions as 'reasonable null hypotheses' – on the contrary, each and every study that documents reliable differences between, say, LOB and FLOB, adds more evidence against these two assumptions that are so fundamental to most corpus-based studies of structural nativization.

True, via the integration of contemporary BrE corpora as reference data, synchronic studies of structural nativization by design theoretically account for diachronic changes within BrE, but interpretations of findings generally do not consider diachronic variability of BrE a (noteworthy) factor. Synchronic structural differences between contemporary BrE and a given postcolonial English tend to be viewed as manifestations of structural nativization in this postcolonial English,

while the alternative interpretation – stability in a postcolonial English and development in BrE – is only rarely considered. Also with SinE – a postcolonial English in the phase of endonormative stabilization or beyond (see Schneider 2007: 155–261) with structural particularities on all linguistic levels – this interpretation of said synchronic structural differences as the result of nativization processes can be adequate for a number of products of structural nativization, i.e. first and foremost in cases where the assumption of BrE stability is valid. Based on e.g. the Singaporean and British components of the International Corpus of English, the *kena* passive (cf. Fong 2004: 98) instantiating a categorical difference in voice realizations between the historical input and its postcolonial English is undoubtedly a result of structural nativization because this option is not available in present-day BrE or, in other words, because the absence of the *kena*-passive is stable in BrE.

However, when the focus shifts from categorical choices to factor-driven quantitative preferences of structures (e.g. *of*- vs. *s*-genitive) available both in the historical input and postcolonial variety, what is sorely needed is the analog to real-time sociolinguistics, i.e. studies of structural nativization that are not based on ‘apparent/simulated’ diachronic data, but actual diachronic data. Two potential scenarios to trace structural nativization seem empirically desirable: (a) modeling the structural development of this feature over a certain period of time up to the present day via a diachronic corpus of the postcolonial English concerned and then identifying whether and which developments in the postcolonial English are or are not compatible with a present-day BrE model of said feature or (b) producing diachronic models for both the postcolonial English and BrE to detect at what points certain factors and resulting structural choices converged or diverged. The next section following approach (a) is devoted to be the first study to exemplify the protocol we submit is required to put structural nativization studies onto a (more) solid empirical footing.

### 3. A diachronic MuPDAR analysis of SinE

In this section, we discuss a truly diachronic analysis of how SinE90 has changed over the past and how those results relate to the results from the previous section on how SinE90 differs from BrE90. Specifically, we add to our above corpus data from additional data representing SinE from the (late) 1950s and the 1960s; these are then annotated in the same way as the SinE90 data and subjected to two analogous MuPDAR analyses, namely SinE50 → SinE60 and SinE60 → SinE90. The period thus covered captures – according to Schneider (2007: 155) – two evolutionary developments in SinE, i.e. the transition from the phase of exonormative stabilization to nativization (approximately 1945–1970s) and that from nativization towards

endonormative stabilization. A sociocultural perspective on the diachronic span of the data thus led us to assume that noteworthy structural changes – also with regard to the genitive alternation – reflecting said evolutionary progress should have occurred in the periods we chose to study. In the next section, we describe our methods, but we keep this part brief given how it overlaps with that of Section 2.1.

### 3.1 Methods

#### 3.1.1 *Generating and annotating the concordance data*

The genitives in question were extracted from a preview version of the *Historical Corpus of Singapore English*; this corpus of historical Singapore English (see Hoffmann et al. 2012, Hoffmann 2013) will eventually feature written texts representing Standard Singapore English (as opposed to Singlish) from 1951 to 2011 sampled in 10-year intervals. The dataset will be based on four major text categories – informative prose (general and academic), imaginative prose, newspapers and speeches – with potential additions from non-public material such as school essays, letters, computer-mediated communication, etc. Although there is evidence that certain constraints of genitive choice might weigh differently in certain sub-genres (Grafmiller 2014), there is no evidence for aggregate differences between spoken vs. written language in previous cross-varietal investigations of genitive choice (Szmrecsanyi et al. 2016; Heller et al. 2017). Therefore, we felt confident in assuming that differences in corpus compilation would not systematically affect our results in this case. Our searches and annotations were performed in the same way as above; the new data set's composition is shown in Table 3; to arrive at a reasonable sample size, all modes/registeres had to be included.

Table 3. Overview of our data, second case study

	<i>of</i> -genitive	<i>s</i> -genitive	Total
SinE 1950	354	120	474
SinE 1960	970	345	1315
SinE 1990	712	304	1016
<b>Total</b>	<b>2036</b>	<b>769</b>	<b>2805</b>

#### 3.1.2 *Statistical analysis*

On the whole, the statistical analyses were performed in the same way as above, the only difference being that now there were two MuPDARs: SinE50 → SinE60 and SinE60 → SinE90. We therefore do not reiterate the description of the variable transformations and the model-fitting parameters but turn to the overview results right away.

The first diachronic MuPDAR study, SinE50 → SinE60, yielded a highly significant model ( $LR = 285.95$ ,  $df = 46$ ,  $p < 0.001$ ), accounted for a very large proportion of the data ( $R^2_m = 0.88$ ,  $R^2_c = 0.91$ ), and came with a good classification accuracy (accuracy = 90.5%,  $C = 0.95$ ); we therefore proceeded with the analysis and applied the regression model to the SinE60 data (without random effects). There, too, the model performance was good with a high accuracy (83.3%) and good classificatory power ( $C = 0.86$ ). Thus, in our final step, we created a variable SinE50like which captured for every SinE60 choice whether it was what a SinE50 speaker would have been predicted to use and used it as a dependent variable in this MuPDAR's  $R_2$ . That  $R_2$  model was arrived at in the same way as in Section 2.1.2 above and yielded a highly significant final model ( $LR = 302.36$ ,  $df = 36$ ,  $p < 0.001$ ) with solid  $R^2$ s ( $R^2_m = 0.55$ ,  $R^2_c = 0.57$ ) and accuracies (accuracy = 86.2%,  $C = 0.86$ ).

The second diachronic MuPDAR study, SinE60 → SinE90, yielded a highly significant model ( $LR = 685.58$ ,  $df = 46$ ,  $p < 0.001$ ), accounted for a large proportion of the data ( $R^2_m = 0.6$ ,  $R^2_c = 0.67$ ). This model, too, came with a good classification accuracy (accuracy = 88.2%,  $C = 0.93$ ) and we proceeded by applying this  $R_1$  to the SinE90 data. There, too, the model performance was good (accuracy = 80.5%) with good classificatory power ( $C = 0.86$ ). Thus, in our final step, we created a variable SinE60like which captured for every SinE90 choice whether it was what a SinE60 speaker would have been predicted to use and used it as a dependent variable in this MuPDAR's  $R_2$ . The final model from that  $R_2$  then was highly significant ( $LR = 217.81$ ,  $df = 35$ ,  $p < 0.001$ ) and predicted the SinE90 choices well ( $R^2_m = 0.57$ ,  $R^2_c = 0.58$ , accuracy = 82.2%,  $C = 0.82$ ).

The above summaries of the two MuPDAR analyses show that there are significant effects but, as usual, the most relevant results are of course the significant effects that show how (i) SinE60 speakers' choices differ from SinE50 speakers' choices – the first (truly) diachronic comparison – and how (ii) SinE90 speakers' choices differ from SinE60 speakers' choices – the second diachronic comparison. In an attempt to highlight in particular how the real diachronic analysis differs from the apparent diachronic analysis, we found it useful to adopt the language of classification tasks, i.e. the notions of true/false positives and true/false negatives. Specifically, the possibilities are that

- BrE90 → SinE90 returns an effect which is also found in at least one of SinE50 → SinE60 or SinE60 → SinE90; this would be a true positive (*tp*); note that *true positive* may at this point be too flattering a label since an effect may be found as specified, but that does not imply, as we will see, that it is the same effect;
- BrE90 → SinE90 returns an effect which is found in neither SinE50 → SinE60 nor SinE60 → SinE90; this would be a false positive (*fp*);

- BrE90 → SinE90 returns no effect, but at least one of SinE50 → SinE60 or SinE60 → SinE90 finds it; this would be a false negative (*fn*);
- none of the three MuPDARs finds an effect, i.e. some effect never makes it through the model selection process; these would be true negatives.

We are not going to say much about true negatives but will of course discuss examples from each of the other three possible outcomes. Table 4 shows the effects we will discuss and in which MuPDAR they were found.

**Table 4.** Overview of effects to be discussed

Effect	SinE50 → SinE60	SinE60 → SinE90	BrE90 → SinE90	Type
3.1.3 LengthDiff × FinSib	X		X	<i>tp</i>
LengthDiff × Definiteness		X	X	<i>tp</i>
FinSib × Animacy	X	X	X	<i>tp</i>
TTR × Animacy			X	<i>fp</i>
Definiteness × Animacy	X	X		<i>fn</i>
LengthDiff × PrevChoice	X			<i>fn</i>
TTR × PrevChoice		X		<i>fn</i>

### 3.2 Results part 1: ‘Positives’

In this section, we revisit the first four results of Table 4.

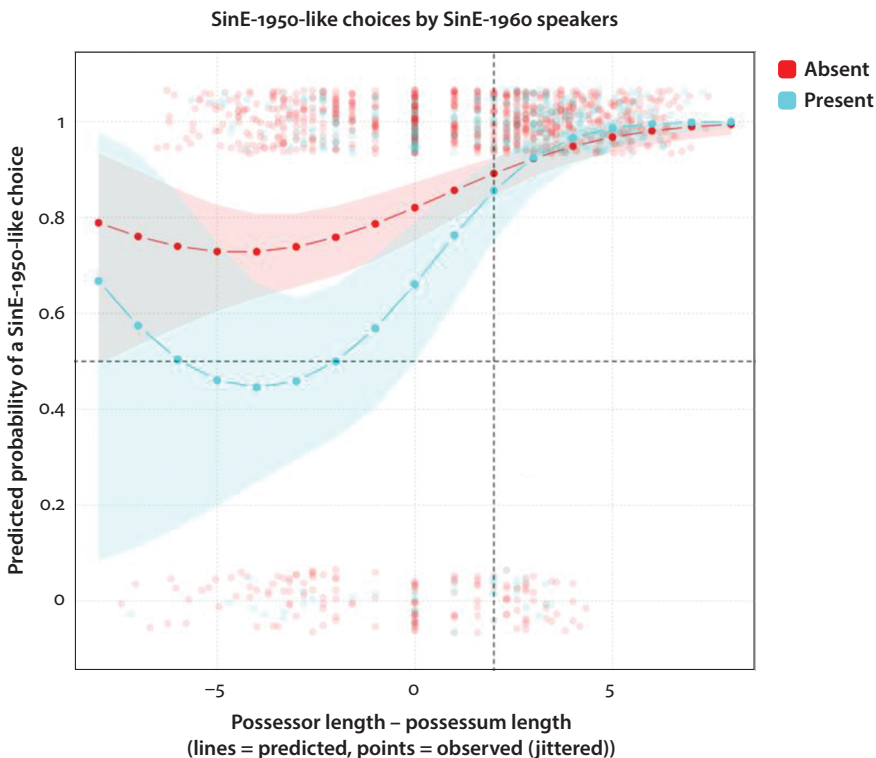
#### 3.2.1 *LengthDiff* × *FinSib*

Figure 6 shows the result of *LengthDiff* × *FinSib* found in  $R_2$  of SinE50 → SinE60, which is indeed very similar to the same effect of BrE90 → SinE90 in Figure 1. However, even this is not necessarily good news for the apparent-time MuPDAR because it indicates that the BrE90 → SinE90 analysis lacks temporal resolution and it is the real-time MuPDAR that returns the more precise location (in time) of the effect. Put differently, the presence of this effect in the apparent-time MuPDAR does not permit one to assume when an effect hypothesized to be at work diachronically has taken place.

#### 3.2.2 *LengthDiff* × *Definiteness*

Figure 6 shows the result of *LengthDiff* × *Definiteness* found in  $R_2$  of SinE60 → SinE90; while the confidence bands are wide especially on the left, it is again clear to see that the overall trends are similar.



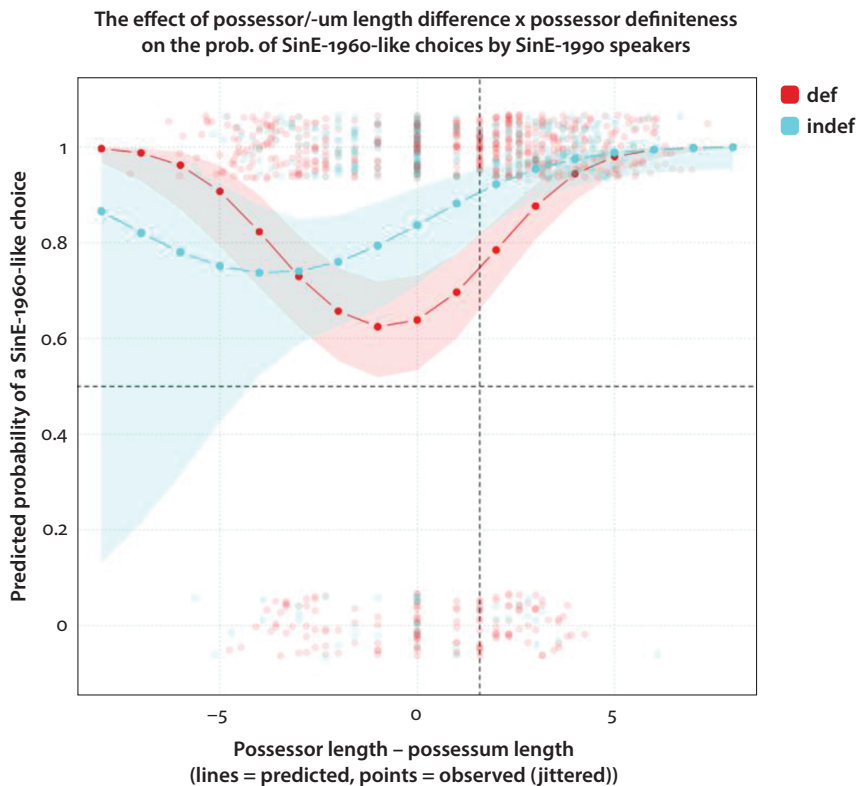


**Figure 6.** Predicted probabilities of SinE-1950-like choices by SinE-1960 speakers: The interaction of LengthDiff  $\times$  FinSib

As in the previous section, this suggests that much of what the BrE90  $\rightarrow$  SinE90 MuPDAR found for this effect is indeed only a development that the apparent-time MuPDAR could not pinpoint more precisely. Put differently, some, but not all, of how the SinE90 speakers differ from the BrE90 speakers in Figure 4 is actually how they differ from their own past, the SinE60 speakers; that also means that the discrepancy between Figure 4 and Figure 7 may well be due to how both SinE varieties involved here differ from BrE90 and/or how BrE has changed over time.

### 3.2.3 *FinSib* $\times$ *Animacy*

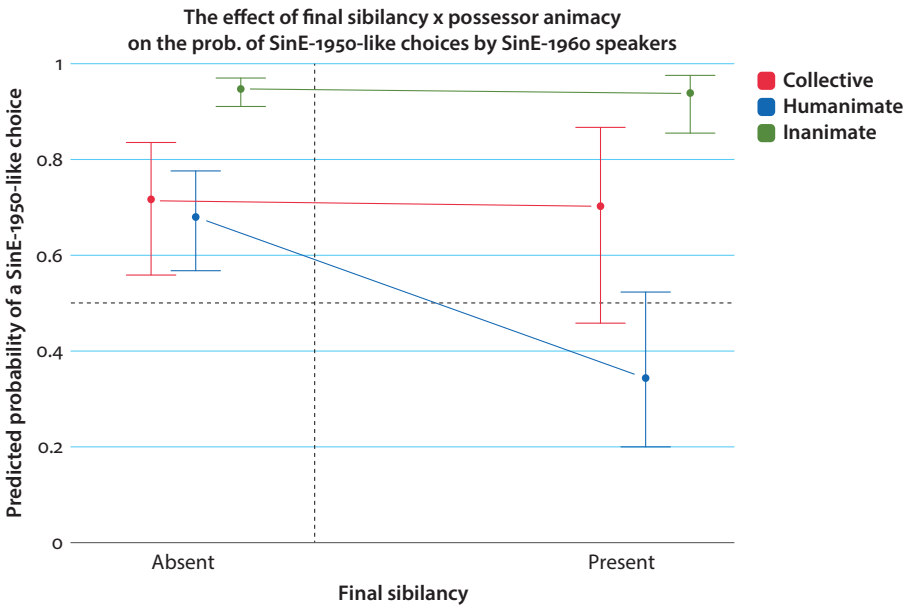
The above two effects have demonstrated that sometimes an apparent-time analysis can yield results that are similar to the real-time analysis, but show up in only a part of the time. Let us now turn to an effect where both diachronic time periods include the same interaction as the synchronic analysis; consider Figure 8 and Figure 9 for the different ways in which *FinSib*  $\times$  *Animacy* is manifested in the data.



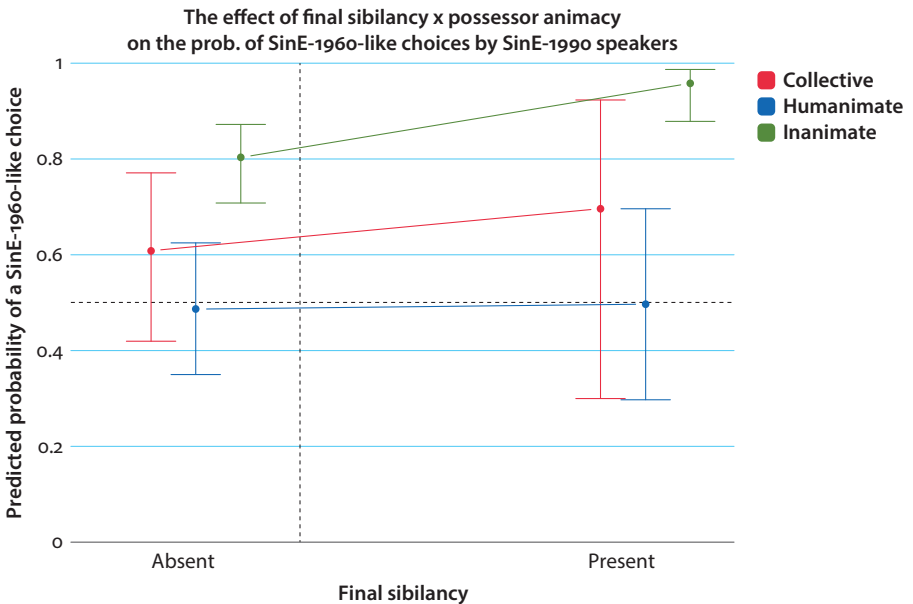
**Figure 7.** Predicted probabilities of SinE-1960-like choices by SinE-1990 speakers: The interaction of LengthDiff  $\times$  Definiteness

In the SE50  $\rightarrow$  SE60 analysis in Figure 8, we find that SE60 speakers behave pretty much like SE50 speakers for inanimate possessors no matter whether the possessor ends in a sibilant or not; a similar irrelevance of FinSib is found for collective possessors although at a lower level of SE50-likeness. However, with humanimate possessors, SE60 speakers are very different from SE50 speakers when the possessor ends in a sibilant: revisiting the data we see that the SE50 speakers have a much higher proportion of *s*-genitives with humanimates ending in a sibilant (38.6%) than the SE60 speakers (25.8%); interestingly, this difference for humanimate possessors between SinE50 and SinE60 speakers was also already reflected in the apparent-time MuPDAR in Figure 2.

In the SE60  $\rightarrow$  SE90 analysis in Figure 9, we find that SE90 speakers behave very much like SE60 speakers for inanimate possessors: compare the blue line to that of Figure 2. However, the other two kinds of possessors differ more from what we found in the apparent-time MuPDAR. Without discussing the results in



**Figure 8.** Predicted probabilities of SinE-1950-like choices by SinE-1960 speakers: The interaction of FinSib  $\times$  Animacy



**Figure 9.** Predicted probabilities of SinE-1960-like choices by SinE-1990 speakers: The interaction of FinSib  $\times$  Animacy

more detail, it does seem as if the apparent-time MuPDAR is a ‘hybrid’ of sorts of the earlier two ones: the apparent-time MuPDAR gets the collectives’ trend right, but it only happens in SinE from 1960–1990, and it gets the humanimates’ trend right, but that one only happens in SinE from 1950–1960. In sum, while the apparent-time MuPDAR was able to discover the same interactions as the two real-time MuPDARs – which is why we generously called these *true positives*, it is also clear that the real-time MuPDAR provides a more fine-grained resolution on what is happening and when.

### 3.2.4 *TTR × Animacy*

The last effect is interesting because it is a false positive: it was found in the apparent-time analysis BrE90 → SinE90, but the real-time analyses SinE50 → SinE60 and SinE60 → SinE90 do not support it; while it is always hard to interpret the complete absence of an effect, it does suggest that whatever BrE90 → SinE90 picked up is a truly synchronic difference, but not one that is based on SinE changing over time in its evolutionary stages we analyzed – the exact nature of this would require analyses going beyond the scope of this already lengthy paper.

## 3.3 Results part 2: ‘Negatives’

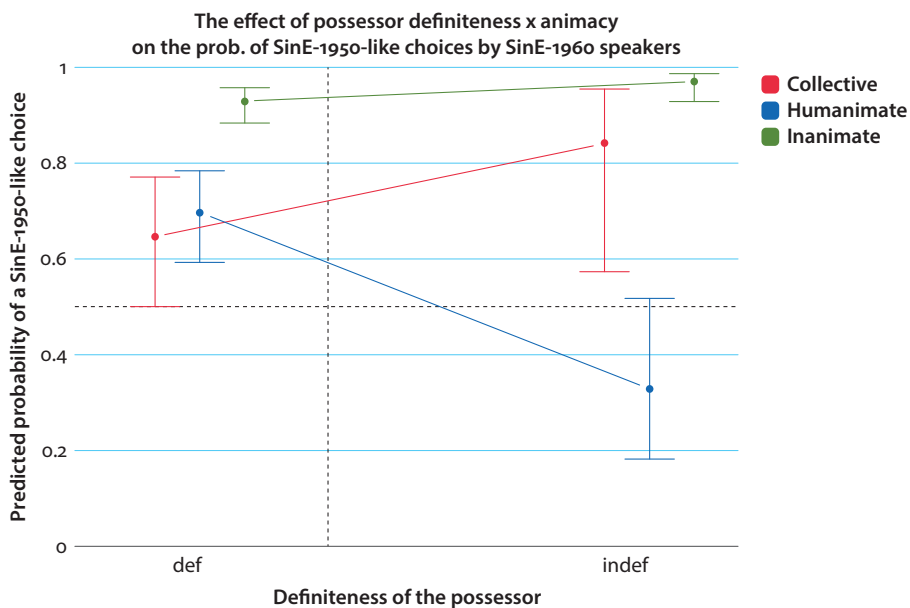
Let us now turn to some diachronic changes that the apparent-time analysis did not detect, i.e. the false negatives.

### 3.3.1 *Definiteness × Animacy*

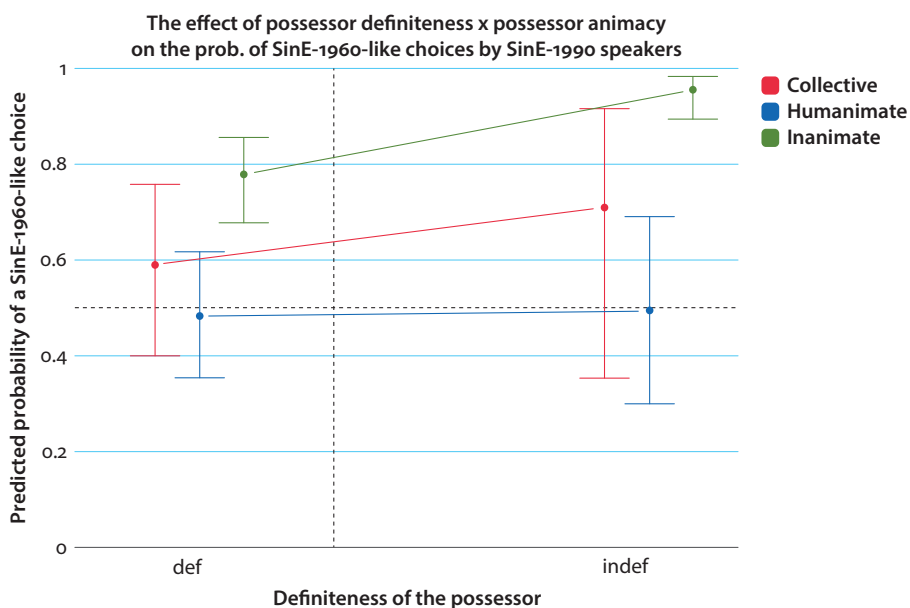
Figure 10 and Figure 11 visualize the interaction Definiteness × Animacy, which was obtained in both real-time analyses.

Comparing the results, it is clear that some diachronic change is discernible: SinE50 and SinE60 speakers behave quite similarly with regard to inanimate possessors and do so regardless of definiteness, but SinE60 and SinE90 speakers are similar only with indefinite inanimate possessors, not with definite ones. Similarly and what is probably the most pronounced change, consider the changes in how humanimate possessors are used: SinE60 speakers use definite humanimate possessors fairly much as the SinE50 speakers do, but are considerably more different with indefinite humanimate possessors – however, SinE90 speakers use humanimate possessors at chance level compared to SinE60 speakers, but regardless of definiteness, which plays no role with them.

Revisiting the original data, we can see that all SinE speakers use *of*-genitives with indefinite humanimate possessors notably more often and particularly so in the SinE60 and SinE90 data, but much less so in the SinE50 data.



**Figure 10.** Predicted probabilities of SinE-1950-like choices by SinE-1960 speakers: The interaction of Definiteness  $\times$  Animacy

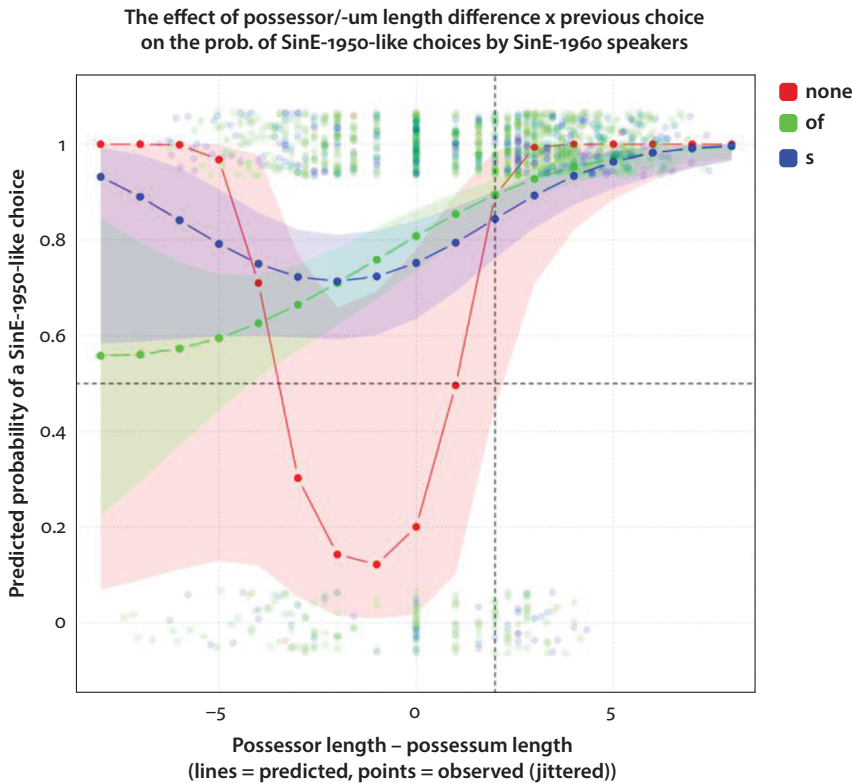


**Figure 11.** Predicted probabilities of SinE-1960-like choices by SinE-1990 speakers: The interaction of Definiteness  $\times$  Animacy

While there is undoubtedly more to discuss here, the above does already highlight that the apparent-time analysis fails to uncover patterns in the data that the real-time analysis did see.

### 3.3.2 *LengthDiff* × *PrevChoice*

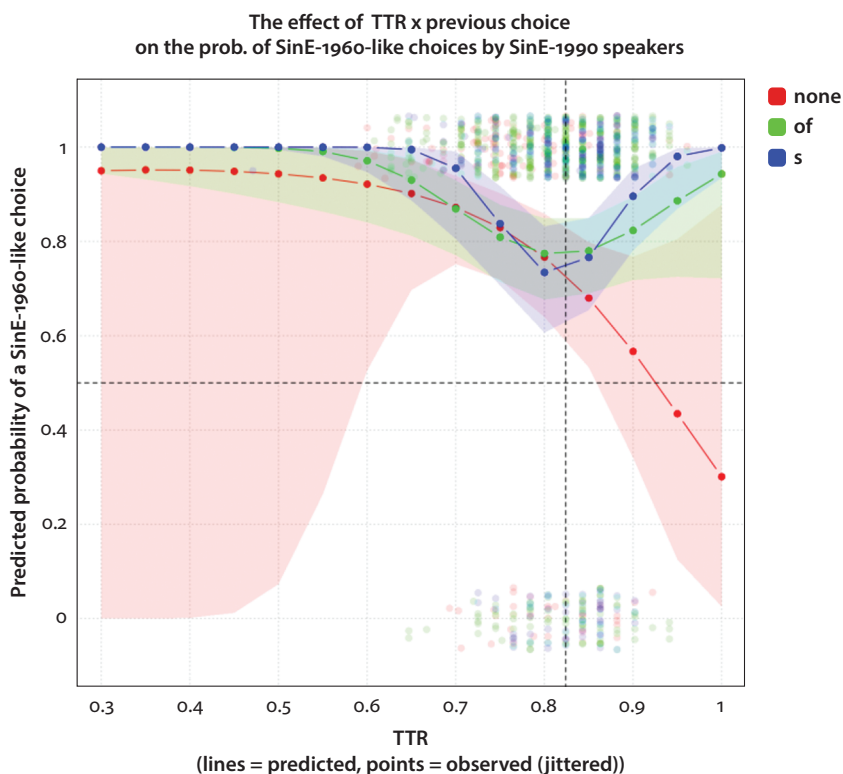
Figure 12 is yet another false negative, the interaction *LengthDiff* × *PrevChoice* obtained in the first real-time analysis but not in the apparent-time one. The results show that, in the absence of priming (in the rare cases when there is no previous choice), SinE60 speakers conform to short-before-long very much like the SinE50 speakers – it is only when *LengthDiff* makes no strong prediction that they differ considerably. A similar pattern emerges when the previous choice was an *s*-genitive although, there, SinE60 speakers are now much closer to the SinE50 speakers – no doubt in part due to the priming. However, when the previous choice was on *of*-genitive, things are different and SinE60 speakers behave less and less like SinE50 speakers the longer the possessum becomes relative to the possessor.



**Figure 12.** Predicted probabilities of SinE-1950-like choices by SinE-1960 speakers: The interaction of *LengthDiff* × *PrevChoice*

### 3.3.3 $TTR \times PrevChoice$

The final effect to be discussed briefly is shown in Figure 13, the interaction  $TTR \times PrevChoice$ , which was only obtained in  $SinE60 \rightarrow SinE90$ . When there is some priming from a previous choice,  $SinE90$  speakers behave a lot like  $SinE60$  speakers but somewhat less so when texts are of average lexical complexity, but when there is no priming,  $SinE90$  speakers behave differently from  $SinE60$  speakers in the most lexically complex texts. (It is worth pointing out that this may be related to the fact that the TTR values in  $SinE90$  are a bit higher on average than those in  $SinE60$ , but that difference is so small that it seems practically negligible: difference between means: 0.02, difference between medians: 0.015, both on the TTR scale from 0.39 to 1.



**Figure 13.** Predicted probabilities of  $SinE$ -1960-like choices by  $SinE$ -1990 speakers: The interaction of  $TTR \times PrevChoice$

#### 4. Concluding remarks

To recap, we have performed one MuPDAR analysis that is closely related in spirit and assumptions to how most of the field has been conducting its corpus-based analyses of structural nativization/emancipation of varieties from a historical source variety and we discovered a variety of differences between BrE90 and SinE90. However, we then proceeded to discuss more explicitly than is usually done two central assumptions that underlie virtually all those analyses and that prove to be highly problematic in indigenized-variety research, as they have in fact been in sociolinguistics where a similar problem/conflict – apparent-time vs. real-time research – has been discussed extensively. These assumptions are that (i) diachronic processes can be reasonably enough approximated by synchronic data with certain sampling characteristics and that (ii) the historical source variety changes so little in the time period under consideration that its changes relative to the target variety can be dismissed from consideration (despite much evidence testifying to how BrE has changed over time).

Based upon this logic, we then proceeded to do the indigenized-variety equivalent of real-time analyses and performed two MuPDAR analyses tracking changes within SinE over time and we have seen that the apparent-time analysis produces true positives (though without the added finer temporal resolution of the real-time analysis!), but also false positives (effects that the real-time analysis cannot confirm) and false negatives (effects that only the real-time analysis reveals). Our focus here was methodological so we did not discuss each of the obtained effects in great detail, but it seems clear to us that the results are ‘mixed’ enough to raise serious concerns regarding what seems to be the state of the art in corpus-based indigenized-variety research relating to evolutionary models of the Schneider type, but also more general. This has two central implications.

First, we do not mean to imply that Moag’s or Schneider’s model(s) are flawed. They are abstract sociolinguistic models with largely sociolinguistic classifications and – although they feature structural or lexicogrammatical indicators of evolutionary processes – they do not bear responsibility for how corpus linguists, with their structural or lexicogrammatical interests, decide to operationalize their claims and interpret corpus-based findings. That being said, it would certainly be useful if such models were formulated with a degree of precision that makes it (more) straightforward to arrive at falsifiable operationalizations to test their claims, not to mention predictions.

Second, we also do not mean to imply that all non-real-time analyses of structural nativization are on the wrong track, and we remind the reader that due to the general lack of diachronic data for World Englishes we ourselves have been involved in analyses of the type we warn of here. That being said, it is clear that the



assumptions underlying apparent-time analyses of the type that have been done so frequently are not obviously tenable and that, when tested, the results from such analyses do not obviously get confirmed – on the contrary. Thus, if the field wants to (begin to) make better-founded claims about whether, when, how, and why structural nativization happens, it needs to face the inconvenient facts that (i) the methodological shortcuts we all have been relying on so far are treacherous, to say the least, and that (ii) real diachronic data are required for analysis.

With regard to how to conduct real-time analyses of structural nativization, we also want to emphasize that we do not mean to imply the procedure(s) adopted are the only tenable ones or the obvious best ones – rather, the strongest claim we wish to make with regard to our specific methodological choices – two MuPDARs for three time periods – is that it yields results good enough to caution us. However, other approaches are conceivable and need to be explored. For instance, while we chose to do separate MuPDAR analyses for SinE50 → SinE60 and SinE60 → SinE90, this is not the only (and certainly not the simplest way to proceed). Immediately obvious alternatives would be the following two:

- one overall multifactorial regression on the whole SinE data set with Variant: *of vs. s* as the dependent variable and all independent variables as well as Time: 1950 vs. 1960 vs. 1990 and all their, say, pairwise interactions as predictors;
- a MuPDAR approach of the type {SinE50 & SinE60} → SinE90, i.e. one where R1 is fit on the combined 1950s and 1960s SinE data and R2 is fit with a predictor that separates SinE50 and SinE60.

In other words, we are trying to (i) raise a greater awareness of the fact that nearly all previous structural nativization research is based on the same two assumptions that make apparent-time studies in sociolinguistics risky and (ii) promote some kind of real-time analysis that avoids those problematic assumptions. Consequently, we hope our contribution is that of a – we believe, much needed – wake-up call, one that will inform a hopefully large set of more precise and rigorous contributions to indigenization.

All the above notwithstanding, we do also think that the approach outlined here has a lot of merit and potential. First, to the extent that the results are robust, the way in which the diachronic MuPDAR approach was able to pinpoint the time period at which differences between successive points of time of the same variety can be observed seems to be a promising additional tool to see when processes compatible with nativization take place (to use the most careful language possible). Second, more comprehensive comparisons – different varieties ‘crossed with’ different time periods – may help shed light on how both varieties in question, here BrE and SinE, change over time.

More diachronic studies in the field of World Englishes will also encourage detailed models of language change in postcolonial settings. Moag's (1982, 1992) and Schneider's (2003, 2007) models are invaluable points of departure for further advancing our understanding of the interplay of progressive and conservative forces in postcolonial Englishes. Still, when trying to relate the findings of the present paper to said models, it becomes all the more obvious that their nature is rather sociocultural than more strictly sociolinguistic. Both models assume that lexical innovations occur at earlier stages of varietal development than grammatical ones. More precisely, Schneider (cf. 2007: 56) reserves lexicogrammatical innovations for the phase of nativization, while novel lexical forms can already emerge in the phases of foundation or exonormative stabilization. In the light of the corpus-based evidence presented here, both models would suggest that SinE has already developed local structural flavors to a considerable degree and investigations of Singaporean history, speaker identities, codification processes, attitudes, etc. would show that SinE should (at least) be classified as an advanced endonormatively stabilized postcolonial English (cf. Schneider 2007: 155). More structurally inclined models of diachronic change in World Englishes would, however, seek to complement these sociocultural findings by addressing questions with regard to e.g. agents of language change, the speed of language change across modes and different genres or – as elaborated in the next paragraphs – the equation of evolutionary progress with structural divergence from a historical input variety.

With a view to future studies, it was argued in Section 1.1 that – in the dynamic model of postcolonial Englishes (Schneider 2003, 2007) – past evolutionary progress and current status is evident from the structural profile of a given postcolonial English. A continuation of this line of thought implies that the structural distinctiveness of a postcolonial English will increase as it progresses through Schneider's (2003, 2007) developmental cycle because sociohistorical and/or sociolinguistic advancement is assumed to be reflected in variety-specific/variety-preferential linguistic choices (cf. Schneider 2007: 30–31). In other words, the Dynamic Model rests on the assumption that more evolutionary progress means more structural difference from a historical input variety, which is British English in most cases.

With the availability of diachronic corpus data for postcolonial Englishes, this model assumption is (maybe finally) empirically testable. In this paper, we focused on the congruence (and its absence) of structural findings in real-time compared to apparent-time corpus studies in World Englishes using data from Singapore from the 1950s, 1960s and 1990s and from Great Britain from the 1990s. With complementary British English datasets from the 1950s and 1960s (e.g. via adapting the methodology put forward in Gries & Bernaisch 2016 or Heller, Bernaisch & Gries 2017 for synchronic regional varieties to diachronic scenarios), studies to come will

be in a position to show specifically for Singapore English whether its progress from phase 3 (nativization) to phase 4 (endonormative stabilization) in the 1970s (cf. Schneider 2007: 155) is indeed marked by an increase in structural distinctiveness and – more generally – whether postcolonial Englishes structurally converge with or diverge from their historical input variety as social and sociolinguistic configurations historically re-adjust.

## Acknowledgement

We thank the compilers of the Historical Corpus of Singapore English (Sebastian Hoffmann, Andrea Sand & Peter Tan) for granting us early access to the data as well as Nina S. Funke for her support with coding the data; also, we are grateful to the reviewers and the editor for their feedback and to Sebastian Hoffmann for additional discussion.

## References

- Behaghel, O. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25(1): 110–142.
- Bernaisch, T. 2015. *The Lexis and Lexicogrammar of Sri Lankan English* [Varieties of English Around the World G54]. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g54>
- Bernaisch, T., Gries, S. Th. & Mukherjee, J. 2014. The dative alternation in South Asian Englishes: Modelling predictors and predicting prototypes. *English World-Wide* 35(1): 7–21. <https://doi.org/10.1075/eww.35.1.o2ber>
- Biewer, C. 2015. *South Pacific Englishes: A Sociolinguistic and Morphosyntactic Profile of Fiji English, Samoan English and Cook Islands English* [Varieties of English Around the World G52]. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g52>
- Collins, P. 2012. Singular agreement in *there*-existentials: An intervarectal corpus-based study. *English World-Wide* 33(1): 53–68. <https://doi.org/10.1075/eww.33.1.o3col>
- Collins, P. (ed.). 2015. *Grammatical Change in English World-Wide* [Studies in Corpus Linguistics 67]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.67>
- Deshors, S. 2017. Structuring subjectivity in Asian Englishes: Multivariate approaches to mental predicates across genres and functional uses. *English Text Construction* 10(1): 132–163. <https://doi.org/10.1075/etc.10.1.o7des>
- Deshors, S. & Gries, S. Th. 2016. Profiling verb complementation constructions across New Englishes: A two-step random forests analysis to *ing* vs. *to* complements. *International Journal of Corpus Linguistics* 21(2): 192–218. <https://doi.org/10.1075/ijcl.21.2.o3des>
- Edwards, A. 2016. *English in the Netherlands: Functions, Forms and Attitudes* [Varieties of English Around the World G56]. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g56>
- Fong, V. 2004. The verbal cluster. In *Singapore English: A Grammatical Description* [Varieties of English Around the World G33], L. Lim (ed.), 75–104. Amsterdam: John Benjamins. <https://doi.org/10.1075/veaw.g33.o6fon>

- Grafmiller, J. 2014. Variation in English genitives across modality and genres. *English Language and Linguistics* 18(3): 471–496. <https://doi.org/10.1017/S1360674314000136>
- Greenbaum, S. 1991. ICE: The International Corpus of English. *English Today* 7: 3–7. <https://doi.org/10.1017/S0266078400005836>
- Greenbaum, S. (ed.). 1996. *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon.
- Gries, S. Th. & Adelman, A. 2014. Subject realization in Japanese conversation by native and non-native speakers: Exemplifying a new paradigm for learner corpus research. In *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, J. Romero-Trillo (ed.), 35–54. Cham: Springer.
- Gries, S. Th. & Bernaisch, T. 2016. Exploring epicenters empirically: Focus on South Asian Englishes. *English World-Wide* 37(1): 1–25. <https://doi.org/10.1075/eww.37.1.01gri>
- Gries, S. Th. & Deshors, S. 2014. Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora* 9(1): 109–136. <https://doi.org/10.3366/cor.2014.0053>
- Gries, S. Th. & Deshors, S. 2015. EFL and/vs. ESL? A multi-level regression modeling perspective on bridging the paradigm gap. *International Journal of Learner Corpus Research* 1(1): 130–159. <https://doi.org/10.1075/ijlcr.1.1.05gri>
- Hall, R. Jr. 1962. The life cycle of pidgin languages. *Lingua* 11(1): 151–156. [https://doi.org/10.1016/0024-3841\(62\)90021-9](https://doi.org/10.1016/0024-3841(62)90021-9)
- Heller, B., Bernaisch, T. & Gries, S. Th. 2017. Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes. *ICAME Journal* 41: 111–144. <https://doi.org/10.1515/icame-2017-0005>
- Heller, B., Szmrecsanyi, B. & Grafmiller, J. 2017. Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English. *Journal of English Linguistics* 45 (1): 3–27. <https://doi.org/10.1177/0075424216685405>
- Hinrichs, L. & Szmrecsanyi, B. 2007. Recent changes in the function and frequency of Standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics* 11(3): 437–474. <https://doi.org/10.1017/S1360674307002341>
- Hoffmann, S. 2013. The Corpus of Historical Singapore English – Practical and methodological issues. UCREL Corpus Research Seminar. <<http://ucrel.lancs.ac.uk/crs/attachments/UCRELCRS-2013-03-26-Hoffmann-Slides.pdf>> (26 June 2017).
- Hoffmann, S., Hundt, M. & Mukherjee, J. 2011. Indian English – An emerging epicentre? A pilot study on light verbs in web-derived corpora of South Asian Englishes. *Anglia* 129(3–4): 258–280. <https://doi.org/10.1515/angl.2011.083>
- Hoffmann, S., Sand, A. & Tan, P. 2012. The Corpus of Historical Singapore English – A first pilot study on data from the 1950s and 1960s. Paper presented at ICAME 33, University of Leuven, Belgium.
- Huber, M. 2012. Syntactic and variational complexity in British and Ghanaian English. Relative clause formation in the written parts of the International Corpus of English. In *Linguistic Complexity. Second Language Acquisition, Indigenization, Contact*, B. Kortmann & B. Szmrecsanyi (eds), 218–242. Berlin: De Gruyter.
- Kachru, B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the World: Teaching and Learning the Language and Literatures*, R. Quirk & H. G. Widdowson (eds), 11–20. Cambridge: CUP.
- Mair, C. 1995. Changing patterns of complementation and concomitant grammaticalisation of the verb *Help* in present-day British English. In *The Verb in Contemporary English*, B. Aarts & C. Meyer (eds), 258–272. Cambridge: CUP.

- Mair, C. 2002. Three changing patterns of verb complementation in Late Modern English: A real-time study based on matching text corpora. *English Language and Linguistics* 6(1): 105–131. <https://doi.org/10.1017/S1360674302001065>
- Mair, C. 2013. The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34(3): 253–278. <https://doi.org/10.1075/ewww.34.3.01mai>
- McArthur, T. 1987. The English languages? *English Today* 3(3): 9–13. <https://doi.org/10.1017/S0266078400013511>
- Moag, R. 1982. The life cycle of non-native Englishes: A case study. In *The Other Tongue*, B. B. Kachru (ed.), 270–288. Oxford: Pergamon Press.
- Moag, R. 1992. The life cycle of non-native Englishes: A case study. In *The Other Tongue*, 2nd edn, B. B. Kachru (ed.), 233–252. Urbana IL: University of Illinois Press.
- Mukherjee, J. 2008. Sri Lankan English: Evolutionary status and epicentral influence from Indian English. In *Anglistentag 2007 Münster: Proceedings*, K. Stierstorfer (ed.), 359–368. Trier: Wissenschaftlicher Verlag Trier.
- Mukherjee, J. & Gries, S. Th. 2009. Collostructional nativisation in New Englishes: Verb-construction associations in the International Corpus of English. *English World-Wide* 30(1): 27–51. <https://doi.org/10.1075/ewww.30.1.03muk>
- Noël, D., van Rooy, B. & van der Auwera, J. (eds). 2014. Diachronic approaches to modality in World Englishes. *Journal of English Linguistics* 42: 3–88. <https://doi.org/10.1177/0075424213512655>
- Osselton, N. 1988. Thematic genitives. In *An Historic Tongue: Studies in English Linguistics in Memory of Barbara Strang*, G. Nixon & J. Honey (eds), 138–144. London: Routledge.
- Rosenbach, A. 2005. Animacy versus weight as determinants of grammatical variation in English. *Language* 81(3): 613–644. <https://doi.org/10.1353/lan.2005.0149>
- Rosenbach, A. 2014. English genitive variation – The state of the art. *English Language and Linguistics* 18(2): 215–262. <https://doi.org/10.1017/S1360674314000021>
- Schneider, E. W. 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79(2): 233–281. <https://doi.org/10.1353/lan.2003.0136>
- Schneider, E. W. 2004. How to trace structural nativization: Particle verbs in world Englishes. *World Englishes* 23(2): 227–249. <https://doi.org/10.1111/j.0883-2919.2004.00348.x>
- Schneider, E. W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511618901>
- Schneider, E. W. 2014. New reflections on the evolutionary dynamics of World Englishes. *World Englishes* 33(1): 9–32. <https://doi.org/10.1111/weng.12069>
- Stevens, P. 1980. *Teaching English as an International Language: From Practice to Principle*. Oxford: Pergamon Press.
- Szmrecsanyi, B. 2006. *Morphosyntactic Persistence in Spoken English. A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110197808>
- Szmrecsanyi, B., Grafmiller, J., Heller, B. & Röthlisberger, M. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37(2): 109–137. <https://doi.org/10.1075/ewww.37.2.01szm>
- Wulff, S. & Gries, S. Th. 2015. Prenominal adjective order preferences in Chinese and German L2 English: A multifactorial corpus study. *Linguistic Approaches to Bilingualism* 5(1): 122–150. <https://doi.org/10.1075/lab.5.1.05wul>

- Wulff, S., Gries, S. Th. & Lester, N. Forthcoming. Optional *that* in complementation by German and Spanish learners: Where and how German and Spanish learners differ from native speakers. In *What does Applied Cognitive Linguistics Look Like? Answers from the L2 Classroom and SLA Studies*, A. Tyler & C. Moder (eds). Berlin: De Gruyter Mouton.
- Zwicky, A. 1987. Suppressing the Zs. *Journal of Linguistics* 23(1): 133–148.  
<https://doi.org/10.1017/S0022226700011063>



# Modeling World Englishes in the 21st century

## New reflections on model-making

Sandra C. Deshors and Gaëtanelle Gilquin

Michigan State University / Université catholique de Louvain

This chapter seeks to bring together all the contributions in the volume. We identify converging lines of argumentation and findings across the studies featured in the book and we show how the approaches they adopt and the linguistic patterns they uncover shed new light on globalized Englishes, the diversity of their uses and their emerging functions. Based on these studies, we discuss possible avenues for future research in the modeling process of World Englishes (WEs) and we make suggestions as to what contemporary theoretical models of WEs should look like in order to truly capture the developmental patterns of WEs in the 21st century. Generally, it emerges that theoretical models anchored in the ‘moment of communication’ are likely to reflect most effectively the intricate dynamics that lies behind the development of Englishes worldwide and that is stirred by linguistic, pragmatic, social, ideological and cultural forces, simultaneously.

**Keywords:** modeling World Englishes, core aspects of theoretical models, ‘communicative event’ approach, dynamics of World Englishes

### 1. Introduction

This book set itself the goal of assessing the capacity of theoretical models to reflect the uses, development and dynamics of World Englishes (WEs) in the 21st century. What the different contributions have underlined is that the unprecedented position of English in today’s world requires new models that account for the ever-expanding roles that it plays in an increasingly large number of contexts. A model like Kachru’s (1985) Three Circles of World Englishes, one of the first attempts to represent the English language in its global dimension, had the great merit of highlighting the non-monolithic nature of English and recognizing the status of non-native varieties of English (especially those in the Outer Circle), as did McArthur’s (1987) Circle of World English or Görlach’s (1990) Circle Model of English, for example. However, most of these traditional models have become



untenable for at least three reasons, having to do with the expansion of English, the advent of corpora and the progress in language modeling.

Starting with the expansion of English, it is clear that the English as we know it today is very different from the English as it was used in the 1980s or 1990s, when the traditional models of WEs were devised. As noted by Buschfeld et al. (this volume), English is “continuously moving, expanding and growing into new regions, functions and application domains”. The process of rethinking our theoretical models of WEs has unveiled new uses of Englishes which should be accounted for, such as “digital Englishes” (Friedrich & Diniz de Figueiredo 2016) or “global English slang” (Coleman 2014). Several of these recent uses are examined in the chapters of this volume and are shown to cross the boundaries established by traditional models and hence challenge the validity of these models. In addition, new language contact situations have emerged, both locally – cf. the use of English in Germany (Mair this volume) or in the Netherlands (Edwards this volume) – and globally – e.g. through the increasing permeability of territorial borders (Siemund this volume). Theoretical models, in order to be ecologically valid, need to be representative of these new linguistic and sociolinguistic realities (see van Rooy & Kruger this volume).

The advent of corpora, and in particular of comparable corpora representing different varieties of English (like the International Corpus of English), has also led to a questioning of traditional models. Corpora have shifted the focus away from the political and historical considerations that lay at the core of most of these models, onto language as such. This has brought about a different perspective, more linguistically oriented, and not always in sync with the political-historical perspective. Recent corpora, especially those representing new types of Englishes (such as the interactive online data collected by van Rooy & Kruger this volume), make it possible to shed light on WEs in the 21st century, while diachronic corpora (cf. Gries et al. this volume) reveal how English varieties have evolved through time. More generally, corpus studies like those brought together in this volume show that the linguistic realities of English and the links between varieties are far more complex than suggested by earlier models. They also underline the importance of having models whose assumptions have been tested empirically – or at least models formulated in such a way that their claims can be tested empirically (Gries et al. this volume).

Finally, the practice of language modeling as a whole has undergone drastic changes over the last few years. By using sophisticated statistical techniques and applying them to corpus data, researchers have been able to represent language usage with a level of refinement never attained before. It has thus become possible to group WEs according to how a certain linguistic phenomenon behaves in these varieties (e.g. Mukherjee & Gries 2009) or to identify patterns of development of certain English varieties (e.g. Gries et al. this volume). In this context, simplistic models like those proposed thirty years ago are obviously not up to standard. Besides, they

have been shown not to stand up to the test of statistical analysis, as illustrated for instance by Deshors's (2014) analysis of data representing English as a foreign language (EFL) and English as a second language (ESL). While necessarily aiming at a certain degree of generalization, theoretical models should also seek to capture the complexity of language and language users, taking into account the heterogeneity of speakers (Buschfeld et al. this volume) and the multiplicity of factors potentially affecting language use (Gries et al. this volume).

The contributors to this volume all aspire to better theoretical models, able to represent the dynamics of 21st-century English use worldwide. While most of them have built on existing models, they have shown how these models could be improved or, in some cases, replaced by other models. In what follows, we describe some of the main aspects of 21st-century modeling of Englishes that have emerged from the different contributions.

## 2. Core aspects of 21st-century modeling of Englishes

Rethinking the theorizing of WEs for the 21st century has helped identify aspects of English uses that are central to a modeling process aimed at capturing the dynamics of English in today's world. In this section, we focus on modes of communication (with special emphasis on computer-mediated communication), genres of expression, multilingual settings, and ideologies and identity construction.

### 2.1 Modes of communication and digital Englishes

The turn of the century has been characterized by the digital revolution. Computer-mediated communication (CMC) has become a major channel to interact with each other and has facilitated exchanges between individuals to such an extent that it can be said to be one of the driving forces behind globalization (see van Rooy & Kruger this volume). It also emerges as an important factor in how the language is shaping. On the one hand, the language used on the internet represents a mode of communication that differs from both written and spoken language (Simpson 2002). It has also led to the creation of many new genres, including emails, tweets or blogging, which "behave quite distinctly from the more traditional genres" (Laitinen this volume). On the other hand, the global dimension of CMC means that individuals from different linguistic backgrounds interact with each other, which "has produced unprecedented forms of language contact and code-switching and mixing" (Buschfeld et al. this volume). English, as the predominant language of this digital revolution, has been greatly affected by these linguistic changes: new modes of communication have led to new linguistic features; new genres have led

to new communicative functions; and new language contact situations have led to new varieties (including hybrid varieties, cf. van Rooy & Kruger this volume).

Existing traditional models of WEs have understandably not given a prominent place to these types of communication, since the digital revolution had hardly begun when the first models were developed. Current models, however, simply cannot ignore this factor. Due emphasis should be placed on the modes of communication, especially computer-mediated ones, and their effects on the (uses of the) English language. These models should also be flexible enough to accommodate the new digital varieties that are likely to emerge in the near future as a consequence of technological advances.

## 2.2 Genres of expression

While it is not new that different genres affect the way globalized Englishes are used, it emerges from the contributions in this volume that certain genres so far relatively unexplored are pivotal in understanding how World Englishes are developing. Many of those are not currently represented in standard corpora such as the International Corpus of English, which makes it necessary for researchers interested in these genres to collect their own data (cf. van Rooy & Kruger this volume). In his chapter, Mair considers both elite and non-elite domains, thus recognizing the importance of “grassroots” usage (Schneider 2016) for the development of English as a global language, next to the long-recognized elitist genres and educated varieties that have been central in traditional models (see van Rooy & Kruger this volume). Among non-elitist genres, Mair mentions popular music in German, which has witnessed an increasingly deep “Anglicisation” through heavily mixed English-German language practices. Van Rooy & Kruger (this volume) investigate online soapie forums, another non-elitist genre characterized by extensive mixing of multilingual repertoires. As for fanfiction writing, examined in Buschfeld et al.’s chapter, it is one of those genres that have developed thanks to the digital revolution (see Section 2.1) and have strongly favored the use of English.

Interestingly, the 21st century has not only seen the addition of new genres, never practiced before, but also the addition of existing genres to the repertoire of certain varieties of English. In particular, the varieties traditionally described as Expanding Circle Englishes have started to use English for certain (internal) functions that used to be fulfilled exclusively by local languages. This is very clearly the case in countries like Sweden, Finland or the Netherlands (see Edwards this volume). In this context, English as a Lingua Franca (ELF) should also be mentioned, as it can be used in many different discourse situations, which makes it possible to approach this variety from a multi-genre perspective, as advocated in Laitinen’s chapter.

This diversification of the English language in terms of genres and its diffusion into new contexts of use (see Buschfeld et al. this volume) should be taken into account in the modeling process, together with the linguistic effects of these changes. Importantly, none of these genres should be excluded on the grounds that they represent grassroots usage. On the contrary, all domains and all registers should be considered, both elitist and non-elitist, both standard and non-standard, as they all have something to say about the linguistic reality of WEs in the 21st century.

### 2.3 Multilingual settings

Although exact figures are lacking, for obvious reasons having to do with the difficulty of making accurate estimates, it is usually considered that more than half of the world population is multilingual (cf. Siemund this volume). In addition to inherently multilingual settings, globalization trends as well as “increased currents of voluntary and forced migration” (Mair this volume) have given rise to new language contact situations, in which the English language plays a central role. As Mair (*ibid.*) points out in relation to Germany, “[t]he English language has become a key component in the linguistic ecology of a country which has become increasingly diverse and multilingual internally and increasingly connected globally”.

These language contact situations have a considerable impact on the form of WEs. This impact can be very visible, as illustrated by code-switching, or it can be more subtle, as in the case of cross-linguistic influence from the mother tongue(s) / substrate language(s), which can materialize as frequency variation or differential preferences. While cross-linguistic influence has been investigated quite thoroughly for the Expanding Circle varieties, this is perhaps less true of the Outer Circle varieties. This can easily be explained by the usually more complex linguistic situation of the speakers of Outer Circle Englishes, the less detailed knowledge of most researchers about the substrate languages that may have influenced these English varieties, and the lack of rich metadata concerning the speakers whose production is included in current corpora. For the sake of simplicity, scholars may therefore have a tendency to “abstract away from (...) issues of language contact, happily comparing American English with Indian English, New Zealand English with Irish English, or Ghanaian English with Singapore English” (Siemund this volume). The ebb and flow of people and the resulting linguistic diversity of populations are unlikely to end soon. It is therefore crucial that, in the future, models of WEs should take up the issue of multilingualism more seriously and more consistently than has been the case so far.

## 2.4 Ideologies and identity construction

Identity construction is one of the core elements of Schneider's (2003, 2007) Dynamic Model. This underlines the close ties that exist between language and identity. By using a certain language variety, speakers can communicate their desire to be recognized as members of a given linguistic (and possibly cultural) community. As noted by van Rooy & Kruger (this volume), marked linguistic choices "become meaningful tools for identity expression". What seems to have changed over the last few years, however, is the nature of the identities that can be constructed through language. First, identity construction appears to be at work among a larger group of speakers, including people who originally used English for utilitarian purposes only, but who now see it as "an additional local language for creative self-expression and identity performance" (Edwards this volume). Second, identities tend to be less fixed and less stable than they used to be (van Rooy & Kruger this volume). Due to increased multilingualism (see Section 2.3), speakers often combine linguistic markers pointing to different identities, which yields hybrid types of identity (*ibid.*). Finally, it is not unusual for speakers nowadays to express an identity that does not match their cultural background, as illustrated by van Rooy & Kruger (this volume) for online soapie forums. This complexification of the linguistic identity and of its link with one's cultural background implies that more attention should be devoted to this issue in future language modeling.

## 3. Model-making in the 21st century: Looking forward

### 3.1 The communicative event as a possible 'focal point' for 21st-century models

Altogether, the various above-discussed aspects that characterize the dynamics of contemporary model-making in WEs research (i.e. modes of communication, genres of expression, multilingual settings, and ideologies and identity construction) clearly confirm the kaleidoscopic nature of World Englishes today. What is more, the contributions in the volume, collectively, point towards the need for scholars to begin to consider those characteristics simultaneously and to account for the possible effect(s) that those characteristics may have on one another as speakers are using the English language. Based on van Rooy & Kruger's (this volume) and Mair's studies (also this volume), this new and unifying approach to the development of WEs (unifying in the sense that it brings together various facets of Englishes worldwide) would provide a way of improving on existing theoretical frameworks and a

potentially fruitful approach to developing new ones. Put differently, the key to a better understanding of the development of Englishes today may lie in an understanding of how the various forces that cause English to develop interact and how, over time, the linguistic systems of Englishes reflect this interaction. As explained in van Rooy & Kruger (this volume) and as Mair (also this volume) suggests, the innovative aspect of this approach lies in that it would require analysts to grant communicative situations (or communicative events) a central place in the model-making process and to view such situations or events as catalysts for linguistic development. Based on Mair's and van Rooy & Kruger's contributions, it emerges that anchoring theoretical models within the frame of a communicative situation (rather than focusing on types of English speakers or developmental processes) not only provides a way to explore the functional forces that drive the development of WEs – an aspect that has so far remained uncharted – but also has the potential of offering a new 'take' on the development of WEs, namely one that brings in several factors known to influence the development of Englishes and considers those factors simultaneously at a given moment in time within a communicative event.

Zooming on such events, or "communicative space", to use Mair's (this volume) terminology, allocating such space a central part in the model-making process would add a dimension to the theorizing of WEs that has so far rarely been accounted for in studies of Englishes worldwide. Communicative space assumes communicative needs that, according to van Rooy & Kruger (this volume), trigger "interpersonal communication [that] provides a highly valuable perspective on developmental patterns". More specifically, the interactive, multilingual communicative dynamics, which takes the form of a negotiation process between speakers, may be a central factor in the general evolutionary processes of Englishes. It is in this context that van Rooy & Kruger observe that "intersubjective alignment [...] appears to be an important functional driver of use, contributing to the propagation of forms". What is crucial here, though, is that from a theorizing perspective, focusing on interpersonal communication and anchoring theoretical models in the communicative space incurs bringing together the linguistic, social, pragmatic, cultural, multilingual and communicative aspects of language use (and language development) that are known to influence language development and to considering those aspects at the same time in a unified fashion by accounting for their interconnectedness. Based on van Rooy & Kruger (this volume), this interconnectedness is a driving force beyond the structural development of Englishes. Indeed, according to the authors, speakers' knowledge of resources "includes knowledge of the sociolinguistic value of the elements and a sense of possible combinations with other elements". Van Rooy & Kruger add that

From here, they [speakers] select elements which they assume to be understood by other users that they communicate with, and which fulfil a desired function within the communicative context of an online forum. (van Rooy & Kruger this volume)

Finally, van Rooy & Kruger point out “[t]he selection of these resources emerges from the interplay between participants’ exploitation of the conventions of the register of interactive online communication, and the ways in which they express complex local and global identity alignments”. While, to some degree, the notion that speakers’ global alignment triggers developmental patterns echoes Buschfeld et al.’s (this volume) view that intra- and international forces influence the way Englishes develop, an emphasis on communicative spaces to model WEs provides a way to address Gries et al.’s (this volume) concern that sociocultural models (such as Schneider (2007) and Moag (1982)) are not enough to explain the linguistic changes of WEs and there is a need to focus more on structural development.

### 3.2 The 21st century speaker within a ‘communicative event’ approach

In the above-described context, the importance of allocating a central part of our theories to the speaker is not only because, within a communicative space approach, he/she is a crucial negotiator but also because the traditional EFL/ESL user has become, in the 21st century, a *global communicator* whose language-related profile (i.e. education, attitudes towards English, functional use of English in various areas of life) is, in itself, a contributing driving force behind the development of WEs. In this context, should communicative events be considered a window into the development of WEs, then factoring in the rich and highly complex profile of WEs speakers is necessary. Indeed, today, the profile of English speakers has changed dramatically. For instance, speakers are no longer restricted to one type of English input that would partly be determined by their location of residence and would thereby characterize the type of English they use. In our day and age, one can easily envisage a non-native English speaker to have developed, as a result of globalization, high geographical mobility and a complex language-related profile. More specifically, such a hypothetical speaker may come with, say, the profile of someone who started out by learning English as a foreign language but then traveled the world and became exposed to a range of English varieties (thereby blurring the line between an EFL and ESL status), and who is an active online blogger routinely communicating with speakers with various proficiency levels in English and from different native linguistic backgrounds. While this is not an uncommon profile for a 21st-century (non-native) English user, theoretical models with a focus on communicative events provide an opportunity to account simultaneously for the speaker’s linguistic history, his/her attitudes towards English and the specific



type of audience he/she is addressing during the communicative event. In this context, models anchored in the ‘moment of communication’ are likely to reflect, more than ever before, the intricate dynamics that lies behind the development of Englishes worldwide and that, as the contributions in this volume demonstrate, is stirred by linguistic, pragmatic, social, ideological and cultural forces, simultaneously. However, a major challenge for 21st-century model-making remains to determine how to empirically account for the mobility and the rich and complex language-related profile of English speakers. Finally, as we discuss in the next section, integrating these forces and how they interplay within communication situations bears important theoretical implications.

### 3.3 Implications of the ‘communicative event’ approach for the categorization of English varieties

Two aspects that are central to model-making in WEs are the categorization of English varieties and the notion of norm (i.e. against which (native) standard should EFL and ESL varieties be assessed to capture linguistic developmental patterns most efficiently). With regard to the former, the *categorization* of Englishes, existing literature shows that the categorization process is not, in itself, a straightforward exercise and models such as Kachru’s (1985) Three Circles, Streven’s (1980) world map of Englishes, McArthur’s (1987) Circle of World English and Mair’s (2013) world system of Englishes illustrate, with all their differences and among other models, the complexity of classifying Englishes effectively. With regard to Kachru’s model specifically, while, today, the Three Circles model remains widely used, Buschfeld et al. (this volume) confirm the danger of nation-bound, static categorizations. This echoes on-going discussions on the ENL-EFL-ESL continuum over the past few years and the question whether or not the three types of varieties constitute distinct categories or whether they can be placed on a continuum of nativeness (see Mukherjee & Hundt (2011) for a collection of studies on the issue; see also Rautionaho et al. (2018)). Collectively, throughout the present volume, the contributions provide empirical evidence supporting the urgent need to let go of nation-bound theoretical frameworks.

Relating this discussion to the above-mentioned *globalized speaker*, the limitations of the ENL-EFL-ESL categorization are quite clear in that not only the boundaries of the three types of Englishes have, by now, widely been reported to be fuzzy but also, as Laitinen (this volume) stresses, ELF is a variety of English that needs to be accounted for in our model-making effort, thereby making the tripartite categorization obsolete. Bringing in ELF to the broader discussion of the categorization of English implies an inevitable and conscious move away from nation-bound theoretical models, as Laitinen (this volume) recommends. In this



context, and in the spirit of including ELF in our attempt to best model Englishes in the 21st century, communicative space approaches, by virtue of their focus on functional forces (and therefore their alignment with ELF research), would offer a promising modeling option for an integrated treatment of ENL, EFL, ESL and ELF, and one that accounts for the pragmatics of communication. This is an important point for, as Gilquin (this volume) reminds us, “[w]hile general linguistic models abstracting away from individual variation are certainly useful for the bird’s-eye view they offer, we should not forget that ultimately, it is the language of *people* that these models seek to describe”.

This focus on the *people* is also echoed in Mair (this volume) and his notion that people are the driving force behind linguistic developments across WEs (cf. Mair’s distinction between elite and non-elite communities). Thus, to relate the current discussion to our earlier point on the usefulness of communicative space as an anchor for the theorizing process, such space is helpful because at its core is a negotiating speaker who is actively engaged with the process of linguistic development. Therefore, in this context, and beyond the question of the continuum, considering the contributions in the volume collectively, the question might arise whether, and to what extent, categorizing English varieties still has a place in contemporary modeling processes, as the individual profiles of globalized English speakers are becoming very hard to trace and as it is increasingly difficult to empirically control for the broad distinctions between EFL, ESL and ELF.

### 3.4 Implications of the ‘communicative event’ approach for the notion of norm

With regard to the second aspect central to the model-making process, the notion of *norm*, adopting unified theoretical models over nation-bound models or models focused on developmental processes raises the central question of what norm should be used and how scholars should control for it empirically. While traditionally British English has served extensively as the native yardstick to assess the degrees of development of ESL varieties from a historical norm, recent research increasingly shows how American-like linguistic patterns are gradually infiltrating non-native varieties. This has been shown in ESL varieties (specifically Singapore English; see Horch (2016) and Deshors & Gries (2016); see also Hundt (this volume) for possible traces of American English across a wider range of ESLs), but as Gilquin (this volume) shows, this infiltration is now also observed in EFL varieties that are closer to the UK than they are to the US. While future research should maintain a strong focus on assessing the influence of American English over different types of Englishes worldwide (as compared to British English), Gilquin’s (this volume)

and Hundt's (also this volume) contributions flag the danger of continuing to use a single native norm as a yardstick for the development of Englishes and question the usefulness of the notion of norm, as traditionally envisaged (i.e. one native English variety equals the normative standard). In other words, rather than assuming the norm to be a variety, perhaps in the 21st century the norm should be envisaged in terms of a collection of linguistic features. Put differently, as van Rooy & Kruger (this volume) ask, is the traditional norm shifting towards a common "core" as the normative standard? Importantly, this notion of a shifting norm, coupled with the above-discussed notion of the globalized speaker and communicative events as catalysts for linguistic change, raises the question whether, in our globalized world, the use of a 'static' normative standard (in the form of an English variety) – as opposed to a norm constantly *in the making* – constitutes a dangerous limitation for our modeling purposes. In other words, in light of the contributions in this volume, scholars should perhaps address the questions to what extent speakers are creating new norms as they are using the language, to what extent the norm has become (and therefore should be approached as) a collective norm (i.e. one that all speakers are contributing to/towards) and how this "standard core" (van Rooy & Kruger, this volume) can be accounted for both theoretically and methodologically.

### 3.5 Methodological implications

In light of the above discussion on possible future directions for the theorizing of WEs, a few methodological considerations are in order. Indeed, a possible shift towards theoretical models that account for communication events would bear significant methodological implications concerning the type of (meta)data that would need to be collected as well as the type of methodological approaches that would need to be adopted to process those data. Overall, it emerges from several contributions in the volume (e.g. Mair's, van Rooy & Kruger's and Gries et al.'s studies) that existing corpora of WEs may be limited to model the development of Englishes in the 21st century. Assuming the above-discussed focal shift towards communicative events as catalysts of linguistic change, data of a type that is not available from existing corpora would be required. That is, large-scale data of on-line communication reflecting speakers' multilingual situations, large-scale data including more diverse genres of expression than are generally explored such as pop culture and song writing (cf. Sections 2.1 and 2.2 of the present chapter). What is more, this above-mentioned focal shift would affect the type of annotation taxonomies that tend to be adopted in WEs research. For instance, while, over the past fifteen years or so the greater majority of studies of WEs has been quantitative and corpus-based, for the most part, those studies have investigated

WEs from a morpho-syntactic (and to a lesser extent lexical) perspective. A focal shift towards communicative events would necessarily require that scholars widen the scope of their annotation schemes so as to include information of a pragmatic and extra-linguistic nature. Ultimately, this focal shift would incur initiating new (corpus) data compilation projects with a view to diversifying our approaches to the development of WEs. Finally, returning to the notion of conducting studies aimed at assessing the interconnectedness of the linguistic, social, pragmatic, cultural, multilingual and communicative aspects involved in shaping World Englishes today, assessing this interconnectedness and the extent to which these diverse variables affect one another in speech production inevitably calls for the development of methodological techniques capable of handling richly annotated data. In that regard, Gries et al. (this volume) clearly illustrate the power of sophisticated statistical techniques and their potential for the model-making process. However, adopting such techniques will require analysts to develop necessary sophisticated practical and statistical knowledge of a higher standard than the current norm so as to ensure the development of theoretical models that truly reflect the complexity of the dynamics of WEs today. At a time when the development of English varieties worldwide involves a complex network of interconnected factors, it is crucial that analysts are equipped with state-of-the-art methodological tools that allow them to assess how the linguistic, social and pragmatic factors collectively shape Englishes worldwide. In the complex reality of the use of English today, the methodological approach we develop and the (statistical) technique we decide to apply have a crucial role to play in the model-making process of WEs, and the connection between, on the one hand, the model-making process and, on the other hand, the methodological designs scholars choose to adopt, should not be underestimated.

#### 4. Concluding remarks

To briefly conclude, throughout the volume it has become clear that globalization is now an inevitable force that linguists have to reckon with as they continue to explore and model the developmental patterns of Englishes worldwide. Further, the contributions in the volume, taken together, have opened our eyes to the fact that 21st century theoretical models of WEs need to be, more than ever before, multi-faceted in nature and should incorporate factors of language change that traditional early models did not include. In other words, it has become clear that in order to paint a faithful picture of the developmental patterns of Englishes, future theoretical models should broaden their analytical scope so as to include explanatory factors that previous models would have considered independently of one another. Further, more than *just* affecting the English language structurally and functionally, the

contributions in this volume suggest that globalization is on its way to changing the way we, as scholars, conceive of the development of Englishes around the world and the factors that collectively stir linguistic change in a globalized and digitalized world. In light of this volume, a new generation of approaches to model-making in WEs research seems to have emerged – that is, approaches that reach beyond the mere structure of English varieties as linguistic systems and a primarily sociolinguistic perspective on developmental patterns and that, instead, aim at integrating pragmatic, social, attitudinal perspectives as well as modes of communication more tightly than ever before into the modeling process. Altogether, the different chapters provide ample evidence that in the 21st century model-making in WEs research has reached a turning point that promises not only a greater understanding of the driving forces that stir the development of World Englishes today, but also lead towards a better grasp of how those forces contribute to the broader processes of globalization and language change.

## References

- Coleman, J. 2014. Understanding slang in a global context. In *Global English Slang: Methodologies and Perspectives*, J. Coleman (ed.), 1–14. London: Routledge.
- Deshors, S. C. 2014. A case for a unified treatment of EFL and ESL. A multifactorial approach. *English World-Wide* 35(3): 277–305. <https://doi.org/10.1075/eww.35.3.02des>
- Deshors, S. C. & Gries, S. T. 2016. Profiling verb complementation constructions across New Englishes: A two-step random forests analysis of *ing* vs. *to* complements. *International Journal of Corpus Linguistics* 21(2): 192–218. <https://doi.org/10.1075/ijcl.21.2.03des>
- Friedrich, P. & Diniz de Figueiredo, E. H. 2016. *The Sociolinguistics of Digital Englishes*. London: Routledge.
- Görlach, M. 1990. The development of Standard Englishes. In *Studies in the History of the English Language*, M. Görlach (ed.), 9–64. Heidelberg: Carl Winter.
- Horch, S. 2016. Innovative conversions in South-East Asian Englishes. *International Journal of Learner Corpus Research* 2(2): 278–301. <https://doi.org/10.1075/ijlcr.2.2.07hor>
- Kachru, B. B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the World. Teaching and Learning the Language and Literatures*, R. Quirk & H. G. Widdowson (eds.), 11–30. Cambridge: CUP for The British Council.
- Mair, C. 2013. The world system of Englishes. Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34(3): 253–278. <https://doi.org/10.1075/eww.34.3.01mai>
- McArthur, T. 1987. The English languages? *English Today* 3(3): 9–11. <https://doi.org/10.1017/S0266078400013511>
- Moag, R. F. 1982. The life cycle of non-native Englishes: A case study. In *The Other Tongue*, B. B. Kachru (ed.), 270–288. Oxford: Pergamon Press.
- Mukherjee, J. & Gries, S. T. 2009. Collostructional nativisation in New Englishes: Verb-construction associations in the International Corpus of English. *English World-Wide* 30(1): 27–51. <https://doi.org/10.1075/eww.30.1.03muk>

- Mukherjee, J. & Hundt, M. (eds). 2011. *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap* [Studies in Corpus Linguistics 44]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.44>
- Rautioaho, P., Deshors, S. C. & Meriläinen, L. 2018. Revisiting the ENL-ESL-EFL continuum: A multifactorial approach to grammatical aspect in spoken Englishes. *ICAME Journal* 42: 41–78. <https://doi.org/10.1515/icame-2018-0004>
- Schneider, E. W. 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79(2): 233–281.
- Schneider, E. W. 2007. *Postcolonial English. Varieties Around the World*. Cambridge: CUP. <https://doi.org/10.1017/CBO9780511618901>
- Schneider, E. W. 2016. Grassroots Englishes in tourism interactions. *English Today* 32(3): 2–10. <https://doi.org/10.1017/S0266078416000183>
- Simpson, J. 2002. Computer-mediated communication. *ELT Journal* 56(4): 414–415. <https://doi.org/10.1093/elt/56.4.414>
- Stevens, P. 1980. *Teaching English as an International Language*. Oxford: Pergamon Press.

# Index

## A

- accessibility hierarchy 144–146
- African American Vernacular English 20, 43, 51, 53, 59, 72, 136, 150, 152–153, 190
- agreement 64, 110, 139, 147–150, 159, 161, 254, 276
- americanization 11, 187, 213–214, 220, 240, 244
- angloversals 5, 14, 139, 146–147, 151–152, 158, 162

## B

- British English 4, 10–11, 51–52, 112–113, 122–123, 125, 135–136, 158–159, 167, 187–190, 192, 213, 219, 240, 243, 245, 248–252, 275, 277, 290

## C

- categorization of Englishes 4, 289
- code-mixing 20, 184
- colloquialisms 96
- communicative event 281, 286–290
- comparative method 137
- computer-mediated communication 32–33, 40, 77, 80, 170, 263, 283, 294
- context of acquisition 187, 189
- corpus-assisted discourse studies (*see also* CADS) 164, 168, 185
- (CADS) 164, 168–169, 171, 182, 185
- corpus of English texts in Sweden 114

- corpus of English texts in Finland 114
- Cyprus English 15, 29, 41

## D

- de Swaan 5, 50–51, 53, 134–135
- diachrony 245
- digital Englishes 33, 38–39, 42, 282–283, 293
- Discourse analysis 163–164, 168, 183–184, 278
- Dynamic Model 4–5, 8, 12, 15, 17–19, 22–23, 32, 40–41, 43–44, 46, 55–56, 74, 78, 106, 163, 165, 188, 190–191, 238–239, 246–247, 249–250, 275, 286

## E

- EFL 1, 4, 9–12, 17–19, 21, 24, 26–29, 31–32, 42, 52, 109, 124, 127, 165–168, 174, 185, 190–216, 247, 277, 283, 288–290, 293–294
- ELF 7, 9, 44, 51, 71, 109–116, 118–131, 190, 192, 214, 284, 289–290
- English as a Foreign Language 17, 109, 163, 187, 189–190, 283, 288
- English as a Lingua Franca 43–44, 48, 71, 109–110, 114, 129–131, 136, 158, 184, 190, 284
- English as a Second Language 17, 109, 115, 129, 155, 163, 208, 283
- English in Bosnia-Herzegovina 15

- ESL 1, 4, 9–12, 17–19, 21, 24, 26, 28–29, 31–32, 42, 66, 109, 127, 165–166, 168, 190–209, 211–213, 215–218, 220–222, 225, 227, 238, 245, 247, 277, 283, 288–290, 293–294
- Extra- and Intra-territorial Forces Model 15, 23–24, 41

## F

- Facebook 15, 33–35
- fanfiction 15, 33–39, 42, 44, 284

## G

- Gambian English 64–65, 71, 74
- genitive alternation 11, 159, 245, 251–253, 263, 277
- genre 33, 61, 113–116, 118–119, 124–126, 129–130, 284
- genres 109, 113–116, 118–119, 122, 124–128, 130, 183, 196, 275–277, 283–286, 291
- German English 15, 46
- globalisation (globalization) 1–2, 4–7, 9–10, 12, 15–16, 19, 22–23, 29, 32, 38–41, 43, 45–46, 50, 55–56, 72–74, 77, 80–84, 106–107, 127, 130, 152–154, 182–185, 191–192, 206, 208, 210, 213, 214, 241, 239, 242, 243, 283, 285, 288, 292–293
- grammaticalization 139, 142, 152, 156, 158–159, 161–162, 277
- Greek English 15

## H

- hierarchy of individuation 143

## I

identity 10, 18–19, 21–22, 33, 38–39, 43, 72, 74, 80, 83, 93, 95, 97–100, 102–107, 163–164, 166, 181–182, 208, 215, 248–249, 278, 283, 286, 288, 294  
 identity construction 39, 43, 80, 104, 107, 163–164, 166, 181–182, 215, 278, 283, 286, 294  
 Igbo 65–69, 73, 154  
 interference 49, 68–69

## K

Kachru 3–5, 11, 13–15, 17–19, 32, 43, 45, 47, 49, 56, 72, 78, 80, 105, 107, 110, 154, 163, 165, 188, 190, 193, 203, 214, 244, 246, 249, 278, 281, 289, 293

## L

L3 acquisition 160  
 language contact 2–3, 8–10, 12–13, 18, 22, 28, 31–32, 34, 36, 56, 74, 107, 137, 139, 152, 154–156, 161, 184, 238, 248, 282–285  
 language ideologies 163–164, 167–169, 171, 180, 182–186  
 language universals 9, 133, 139–140, 159  
 learner English 42, 65, 109, 111, 113, 122–123, 130, 195  
 Learner Englishes 32, 42, 131, 184, 187, 213–214, 294

## M

macro variation 133  
 mandative constructions 11, 230–231, 241  
 mandative sentences 217, 219, 221–223, 225–226, 228, 238, 240, 242  
 micro variation 133  
 Moag 6, 11, 245–249, 260, 273, 275, 288  
 modal alternation 217  
 modes of communication 124, 283–284, 286, 293

multilingual individual(s) 129  
 multilingual settings 2, 56, 78, 128, 283, 285–286  
 multilingual speaker(s) 129, 138, 154, 157–158  
 multilingualism 9, 45, 50, 71, 77, 79, 99, 106–107, 114, 158, 183–184, 186, 285–286  
 multiplayer online games 7–8, 15, 34, 37–38  
 MuPDAR 11, 245, 251–252, 254, 256, 262, 264–267, 269, 273–274

## N

Namibia 8, 15, 26–28, 31–32, 40–41, 43, 55  
 negative concord 141, 147, 150–153  
 Netherlands 10, 12, 19, 22, 42, 154, 163–164, 166–170, 174–175, 178–179, 181–183, 185–186, 194, 199, 276, 282, 284  
 New Englishes 13, 43, 52, 107, 154, 184, 187, 213, 215, 243, 246–247, 276, 278, 293–294  
 Nigerian English 51, 103, 154–155, 190  
 nonstandard 9–10, 20, 37, 40, 51, 54, 59–60, 71–72  
 norm 3–4, 10–13, 18, 40, 52, 54, 65, 68, 78, 82–83, 107, 110, 136, 192, 205–206, 211–213, 241, 247–248, 289–292  
 Northern Subject Rule 147–149

## P

postcolonialism 15  
 probabilistic grammar 217, 245

## R

reflexive markers 140–141  
 relative clause formation 146, 277  
 repertoire 2, 77–78, 82–83, 91, 93, 95, 98, 101–106, 167, 284

## S

Schneider 1, 3–6, 8, 10–11, 15–23, 26, 32, 38, 41–43, 46, 52, 54–56, 74, 77–80, 105–106, 108, 110, 155, 159–166, 182–183, 188, 190, 192, 207–208, 218, 220, 228, 236, 238–239, 242, 245, 247–250, 260–262, 273, 275–276, 284, 286, 288  
 Singapore English 7, 11–12, 113, 119, 121, 137, 154–156, 158, 162, 207, 220, 239, 245, 250, 252, 263, 276–277, 285, 290  
 social network(s) 104, 128–129, 131  
 South Africa 77, 79, 82, 88, 95, 99, 101, 106–107, 194, 204–207, 215–216  
 South African English 9, 51, 87, 90, 94, 103, 108, 190, 204, 207  
 South-East Asia 8, 15, 26–27, 39, 192, 214  
 standard 1, 4–5, 10, 21, 28–29, 34, 42, 45, 48, 50–54, 59–61, 68, 71, 77, 79–81, 86–98, 102–105, 110, 112–113, 118–119, 121–122, 124, 126, 135–136, 141–150, 153–158, 161, 167–168, 175, 180, 185, 187–190, 192, 214, 221–222, 234, 239, 241, 249, 263, 277, 282, 284–285, 289, 291–293  
 subjunctive 217–232, 234–236, 238, 240–244

## T

theories of World Englishes 15, 191  
 Three Circles model 4, 19, 163, 165, 289  
 Transnational Attraction 5, 8, 10, 15, 22–23, 32, 38, 40, 55, 79, 164–165, 182  
 typological profiling 109, 111, 127, 131  
 typology 9, 14, 73, 130, 133, 136, 138–140, 158–159, 161–162

## V

vernacular universal(s) 14, 110,  
139, 146–147, 158–159, 161–162  
VOICE corpus 116, 119–120

## W

World Englishes 1–17, 20–21, 23,  
28, 34, 38, 41–44, 46–47, 50,  
52, 55–56, 70, 73–74, 77, 79, 81,  
106–108, 111, 127–130, 133, 135,  
137–141, 146–147, 150, 152–153,

157–158, 160–162, 164, 182–186,  
190–193, 196, 202, 211–215,  
217–222, 231, 236, 240–250, 273,  
275, 278, 281, 284, 286, 292–293

## World System of Englishes

2, 5, 8, 10, 13, 43, 45, 51, 53–54,  
71, 74, 79, 107, 110, 130, 133,  
139, 144, 146, 149–150, 152, 157,  
160, 187–188, 190–191, 199,  
202, 211, 214, 243, 249, 278,  
289, 293

## Written English as a Lingua

Franca (WrELFA) 114, 116,  
120, 122, 128



At a time when globalization and the advent of the internet have accelerated the spread and diversification of English varieties worldwide, this book provides a constructive assessment of the theoretical models that best account for the development and use of Englishes in the early 21st century. In this endeavor, the present book brings together cutting-edge contributions by leading scholars who explore the notion of linguistic globalization based on a wide range of ESLs, EFLs and ELF, synchronic and diachronic data, different methodological approaches (corpus-based, sociolinguistic, ethnographic), and a variety of data resources (social media, multiplayer online games, journalistic data, GloWbE, Corpus of Historical Singapore English, thematic blogs). Collectively, these studies serve as a springboard for future research on the globalization of Englishes and they contribute to a timely and necessary scholarly conversation on what constitutes adequate theoretical models of World Englishes in the 21st century.

“This volume makes an extremely multifaceted contribution to discuss current developments and challenges in the modelling of world Englishes in the 21st century. By bringing together leading experts in the field, the studies in this volume insightfully put forward necessary theoretical and methodological adaptations to current theoretical models of world Englishes in today’s globalized world. The volume is thought provoking, inspiring and extremely hard to put down. A must-read for every researcher interested in world Englishes.”

**Sandra Götz**, *University of Giessen*

“Changing realities require changing models of these realities, and where modifications of existing models no longer do, new models are being asked for. This volume makes an important first step in that direction by assessing, modifying and, partly, succumbing existing theoretical models of World Englishes and the uses to which they are put to make them fit for capturing the manifold complexities in the ongoing evolution of the global English language complex.”

**Bernd Kortmann**, *University of Freiburg*

“The volume is a powerful reminder to the World Englishes community that state-of-the-art description should be accompanied by state-of-the-art theorizing. It will help re-stimulate valuable discussion in this spirit.”

**Benedikt Szmrecsanyi**, *University of Leuven*

Cover illustration by Nadia de Lange

ISBN 978 90 272 0123 2



9 789027 201232

**John Benjamins Publishing Company**