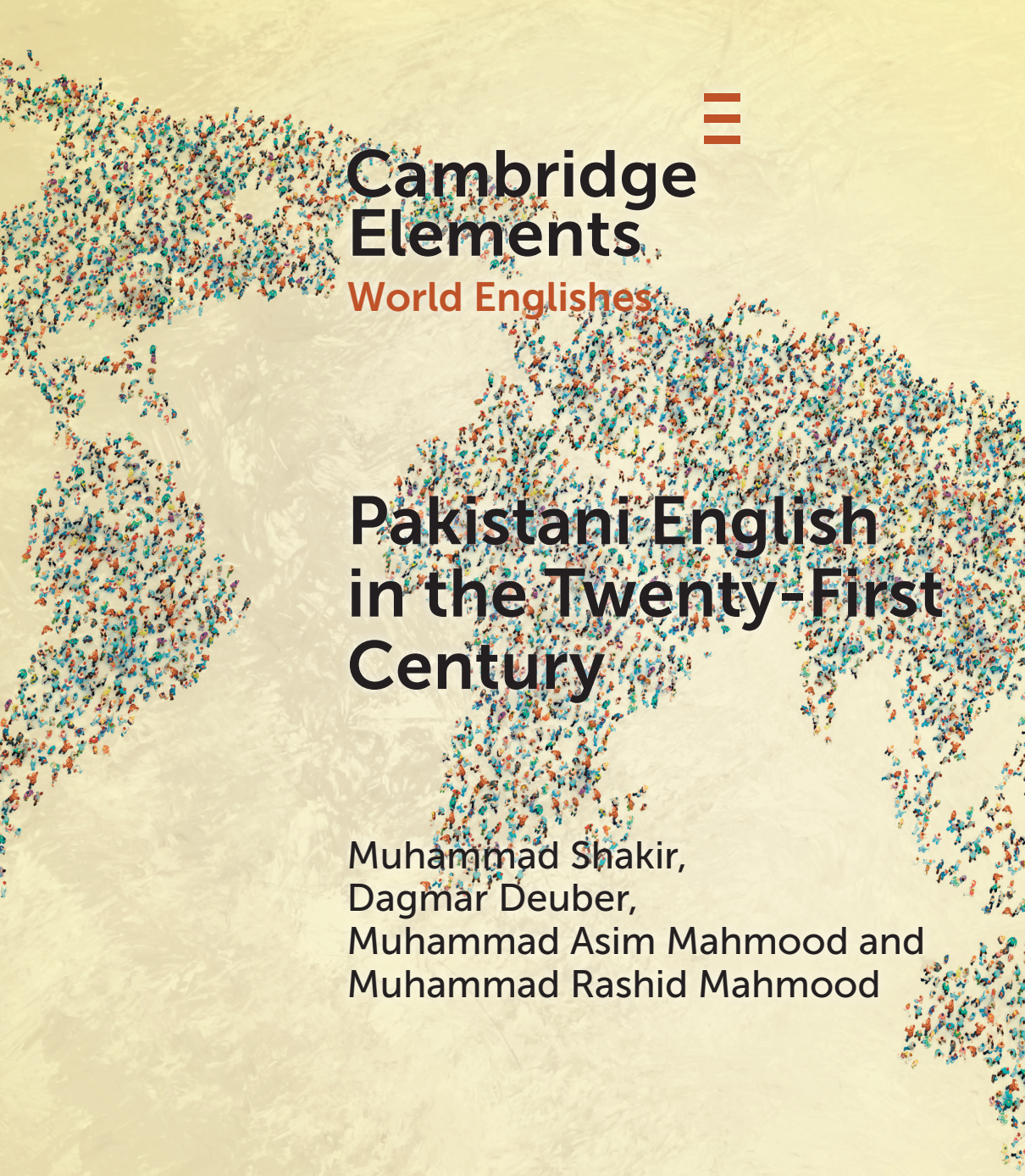


Cambridge Elements

World Englishes

Pakistani English in the Twenty-First Century

Muhammad Shakir,
Dagmar Deuber,
Muhammad Asim Mahmood and
Muhammad Rashid Mahmood

ISSN 2633-3309 (online)
ISSN 2633-3295 (print)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Cambridge Elements

Elements in World Englishes
edited by
Edgar W. Schneider
University of Regensburg

PAKISTANI ENGLISH IN THE TWENTY-FIRST CENTURY

Attitudes, Features, and Uses

Muhammad Shakir
University of Münster
Dagmar Deuber
University of Münster
Muhammad Asim Mahmood
Government College University Faisalabad
Muhammad Rashid Mahmood
Riphah International University Faisalabad

CAMBRIDGE UNIVERSITY PRESS

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781009665292

DOI: 10.1017/9781009665308

© Muhammad Shakir, Dagmar Deuber, Muhammad Asim Mahmood and Muhammad
Rashid Mahmood 2026

This publication is in copyright. Subject to statutory exception and to the provisions
of relevant collective licensing agreements, no reproduction of any part may take
place without the written permission of Cambridge University Press & Assessment.

When citing this work, please include a reference to the DOI 10.1017/9781009665308

First published 2026

A catalogue record for this publication is available from the British Library

*A Cataloging-in-Publication data record for this Element is available from the
Library of Congress*

ISBN 978-1-009-66529-2 Hardback
ISBN 978-1-009-66533-9 Paperback
ISSN 2633-3309 (online)
ISSN 2633-3295 (print)

Additional resources for this publication at doi.org/10.1017/9781009665308

Cambridge University Press & Assessment has no responsibility for the persistence
or accuracy of URLs for external or third-party internet websites referred to in this
publication and does not guarantee that any content on such websites is, or will remain,
accurate or appropriate.

For EU product safety concerns, contact us at Calle de José Abascal, 56, 1°, 28003
Madrid, Spain, or email eugpsr@cambridge.org

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the
Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Pakistani English in the Twenty-First Century

Attitudes, Features, and Uses

Elements in World Englishes

DOI: 10.1017/9781009665308
First published online: August 2026

Muhammad Shakir
University of Münster

Dagmar Deuber
University of Münster

Muhammad Asim Mahmood
Government College University Faisalabad

Muhammad Rashid Mahmood
Riphah International University Faisalabad

Author for correspondence: Muhammad Shakir, muhammad.shakir@uni-muenster.de

Abstract: English is an important part of the linguistic repertoire of Pakistanis and is used in government, top level judiciary, higher education, business, and the internet. This Element, covers brief history, attitudes, a feature description, and uses of English on the internet by Pakistanis. The study of attitudes shows that Pakistanis mostly look outwards for standards and norms. The analysis of various features at phonological, orthographic, lexical, and grammatical levels shows that Pakistani English has distinct features that can be categorised in three ways: contact-induced features (e.g. reduplication), the result of conservativism in the variety (e.g. *which* vs *that* relatives), and influenced by American and British English (e.g. spellings). English is also frequently used on platforms like Facebook, Twitter, and Instagram. The study concludes that English in Pakistan is well-established and stabilised with evolutionary characteristics of 'nativisation' on Schneider's Dynamic Model.

Keywords: world Englishes, Pakistani English, language variation, corpus linguistics, South Asia

© Muhammad Shakir, Dagmar Deuber, Muhammad Asim Mahmood and Muhammad Rashid Mahmood 2026

ISBNs: 9781009665292 (HB), 9781009665339 (PB), 9781009665308 (OC)
ISSNs: 2633-3309 (online), 2633-3295 (print)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Contents

1 Introduction 1

2 Attitudes towards English, Urdu, and Regional Languages in Pakistan 4

3 Phonology, Lexis, and Grammar 20

4 PakE on the Internet 41

5 Conclusion 60

Appendices 65

References 90

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

1 Introduction

1.1 Pakistan

Pakistan is located in South Asia, as shown in Figure 1. On the eastern side, it has a long land border with India, while on the western side, it shares mountainous border areas with Afghanistan and towards the south with Iran. China is located on the northern side of Pakistan across the Himalayan mountains. According to the 2023 digital census, Pakistan has a population of 241 million people (Anonymous 2023: 124). Administratively, the country has four provinces (Balochistan, Khyber Pakhtunkhwa (KPK), Punjab, and Sindh) and a federal capital territory – Islamabad. Additionally, there are two autonomous regions (Azad Jammu and Kashmir (AJK) and Gilgit & Baltistan), which, according to United Nations Security Council (UNSC) resolutions, are disputed territories with India. Punjabi, Pushto, Sindhi, Saraiki, Urdu, and Balochi are the main languages, along with many smaller ones (Anonymous 2023: 138). According to Ethnologue, there are sixty-eight indigenous languages spoken in Pakistan (Anonymous 2025); many very small and endangered languages are spoken in the north-west of the country in Gilgit & Baltistan (GB) and KPK.

1.2 English and English Language Teaching in Pakistan

Like other postcolonial countries in South Asia, Pakistan also inherited the English language from the British after its independence in 1947. Throughout Pakistan's history, it has remained an important language as we briefly explain.

Until 1971, the use of English continued in various government bodies, the military, and education (Abbas 1993: 150). Recommendations were made to promote Urdu as the medium of instruction in primary and secondary schools during this time period, but these efforts were thwarted due to various political issues (Shamim 2008: 237–238). Thus, a two-stream system (elite private English-medium and public sector Urdu-medium schools) continued during this era. After the separation of East Pakistan (now Bangladesh) in 1971, during the first relatively stable democratic period until 1977, regional languages like Sindhi, Saraiki, and Punjabi were recognised and promoted through government bodies like Institute of Folk Heritage (Abbas 1993: 150). However, severe linguistic riots between Urdu-speaking *muhajirs* ('immigrants') from India living in Karachi and the Sindhi-speaking population occurred in Sindh in this period. This established the status of English as a neutral and 'functional language' instead of a colonial legacy (Abbas 1993: 150). During the second martial law in 1978–1989, the military dictator promoted Urdu and Urdu-medium instruction.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308



Figure 1 Geographical map of Pakistan[1]

The National Language Authority was established at the federal level to develop and standardise Urdu (Shamim 2008: 238). Schools (mainly the public sector ones) were instructed to shift to Urdu-medium instruction. At the same time, Pakistan received massive teacher training grants from British and American agencies (British Council and USAID, respectively) to promote English-medium instruction (Abbas 1993: 150–151). Thus, the Urduisation project of this era did not succeed much.

After 1988, English language teaching was progressively extended by introducing the language as a compulsory subject from grade three and then, in the third martial law period (1999–2008), from grade one (Shamim 2008: 238–239). Subsequent education policies, for example, National Education Policy 2017 (Anonymous 2017), aim for English teaching as a compulsory subject from grade one. English is the medium of instruction in the majority of elite and non-elite private schools. Public sector schools (usually until grade ten) in Punjab were converted to English-medium instruction in 2009 (Bashir & Batool 2017: 14), though Urdu-medium textbooks are also available for students, and the question papers in examinations are bilingual (i.e. English and Urdu). In KPK's

[1] Source: https://upload.wikimedia.org/wikipedia/commons/thumb/d/de/PAK_AU_T1.svg/960px-PAK_AU_T1.svg.png

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

public schools, math and science textbooks are only available in English. Public sector schools in other provinces (including those under the federal government) also offer English-medium textbooks and exams. Most education from grade eleven onwards is in English medium.

The co-official status of English in government, the judiciary, the military, and higher education has not only remained largely unchanged, as noted by Haque (1983), but also strengthened due to more exposure to English through media like the internet. In the mid-2010s, the Supreme Court ruled that Urdu should be implemented as the official language, as is prescribed in the constitution (Raj 2017). However, a decade after this ruling, not much has changed in practical terms. The higher courts now more often publish Urdu translations of their judgements. Urdu also remains an option on government portals and mobile applications. Apart from these, English (along with Urdu and regional languages) is a very normal part of billboards and hoardings in markets and bazaars.

Pakistani youth – 50% Pakistanis are under thirty-five years of age (Anonymous 2023: 110) – have increasingly chosen to move abroad for study and work. The students (including masters and doctoral students) seek scholarships like German Academic Exchange Service, Ministry of Education, Culture, Sports, Science and Technology (Japan), Erasmus, Fullbright, and so forth (Qamar & Khurshid 2018). The job seekers are mostly skilled workers (e.g. IT professionals, engineers). Both of these groups prefer to move to Europe, English-speaking countries (Australia, Canada, New Zealand, UK, US), and more recently to China and South Korea as well. This trend has given birth to a host of locally and internationally owned private institutions that teach English and prepare students for language tests like International English Language Testing System and Test of English as a Foreign Language (Memon & Umrani 2016). Thus, the motivations to teach and learn the English language have also increased.

1.3 Previous Research and the Present Study

Pakistani English (PakE) has been studied since the 1990s. There are book length studies like Rahman (2015), first published in 1991, and several research papers from the same decade, for example, Baumgardner (1995). Since the late 2000s, several researchers have published on various aspects of PakE based on a corpus linguistics approach (Muhammad Asim Mahmood 2009; Rashid Mahmood 2009; Shakir & Deuber 2019). Despite these specialised research studies, a concise and descriptive account of PakE is still needed in the twenty-first century for the benefit of the research community in World Englishes and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

students interested in learning about PakE. In the present Element, we precisely aim to do the same by providing descriptive accounts of the attitudes towards English and learning English, a brief description of the features of the variety at various levels (phonology, lexis, and grammar), and a brief look at various uses of English by Pakistani internet users in various online domains.

The Element is structured as follows. The second section presents a study of attitudes towards English, Urdu, and regional languages based on a large survey conducted in each of the four provinces and two regions of Pakistan. The subsequent section, that is, number three, describes various features of PakE. The fourth section is dedicated to documenting the uses of English on the internet, especially on the social media platforms Facebook, Instagram, and Twitter. The last section concludes the Element by reflecting on the findings of the previous sections and by discussing the implications of these findings in the light of the Dynamic Model (Schneider 2007) and the Extra- and Intra-territorial Forces (EIF) Model (Buschfeld 2020).

2 Attitudes towards English, Urdu, and Regional Languages in Pakistan

The aim of the section is to identify and report on uses and attitudes towards various regional languages, Urdu – the national language, and English at the end of the first quarter of the twenty-first century. We start the section with a review of a few selected and relevant studies about language attitudes in Pakistan from the 1990s to the 2020s. Then we present the methodology and data collection procedure of our own study. Finally, we present the results and discussion.

2.1 Previous Relevant Research on Language Attitudes in Pakistan

Attitudinal research on English and other languages in Pakistan dates back to the 1990s or even earlier. One of the earliest studies is that of Baumgardner (1995), based on several questionnaire surveys conducted between 1989 and 1992. His participants include English newspaper journalists, English teachers, and university or college students from Lahore (the provincial capital of Punjab) and Islamabad (the federal capital of Pakistan). The respondents – especially teachers – most probably include people from many different parts of the country, as some of the surveys were conducted at teacher training courses (Baumgardner 1995: 262). Thus, the conclusions made by Baumgardner can be considered representative of general perceptions about the English language in Pakistan in the 1990s.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Baumgardner's (1995) surveys include questions related to perceptions about PakE as a variety, as well as about the acceptability of various grammatical structures that – as per the author's observation – are the hallmarks of the local English variety. The two questions related to perceptions about (Pakistani) English as a variety are: 'What variety of English do you speak?' and 'What variety of English do you think Pakistanis should learn?' (Baumgardner 1995: 263). In the case of the former, about one-third (33%) of the respondents report that they speak 'Pakistani English' and about 10% say they speak a mixture of PakE and another variety (British or American English (AmE)). On the question 'which variety should be spoken', about 80% of the participants advocate for an external model of English speech instead of the local variety (British English (BrE) followed by AmE). Hence, one can see that only a small proportion of the participants acknowledges the existence of a 'Pakistani English', and a very high majority still wants to learn a foreign standard instead of the local variety that they may be speaking.

In the past three decades, Pakistani universities have increasingly introduced the departments of 'English Linguistics' or 'Applied Linguistics' along with the already present 'Department of English' where traditionally English literature would be taught. Additionally, linguistics and its subdisciplines have also been introduced to purely literary programs (i.e. bachelors or masters in English Literature). This has led to an increased awareness of language variation and varieties of English, including PakE. Studies whose participants include students of such programs, for example, Jabeen, Mahmood, and Rasheed (2011) and Parveen and Mehmood (2013), report a higher awareness about PakE. Jabeen, Mahmood, and Rasheed's (2011) 100 participants include 50 university students, among them 25 students of MA English, while Parveen and Mehmood (2013: 659) report that all of their 25 participants, who are MA English students at University of Education in Lahore, have conceptual knowledge of World Englishes and sociolinguistics. The participants in both studies acknowledge the existence of 'Pakistani English'. In their view, the local variety is the best way to express creativity and communication by Pakistanis who do not have contact with foreigners (Jabeen, Mahmood, & Rasheed 2011: 116). However, learning 'Standard English' is also important because it is the means of 'prestige and success' in the world (Parveen & Mehmood 2013: 662). A more recent study of attitudes towards AmE, BrE, and PakE has been conducted using the matched guise technique (MGT) followed by semi-structured interviews by Gul and Channa (2023). Their results from the MGT show that university students in Balochistan have highly positive attitudes towards a BrE accent (Gul & Channa 2023: 8). The thematic analysis of interviews indicates a negative attitude towards a nativised Pakistani accent. Many of them feel an inferiority complex

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

due to their non-native accent (Gul & Channa 2023: 10–11). Gul and Channa's (2023) study shows that students – even from disciplines unrelated to English studies – have some awareness of a Pakistani way of speaking English, but the attitudes towards the accent remain largely negative and unchanged since the 1990s.

In addition to the studies mentioned previously in Section 2.1, many studies have looked at the motivations of learning English and attitudes to the English language among Pakistani students of various levels. Kahlid (2016) and Abbas and Iqbal (2018) are two studies that have been conducted in Lahore, Pakistan. Khalid's (2016: 5) participants are sixteen-year-old male students of English-medium public and private schools. Abbas and Iqbal (2018: 205) include a larger population of 445 university students from five different universities in Lahore, where students from other rural and urban areas of Pakistan also come for study. Both studies include questions related to motivations and attitudes towards English as well as towards local languages (Urdu, or Urdu and Punjabi). Khalid (2016: 8) finds that Pakistani learners have positive attitudes towards English and learning this language, and these are not at the expense of Urdu. Instead, English learning is mainly for utilitarian and instrumental reasons. In Kahlid's (2016: 12) view, Pakistani students are additive bilinguals where English is not a threat to the first language, Urdu. Similarly, Abbas & Iqbal (2018: 209) conclude that English is considered important as a global language and their participants are motivated to learn it due to instrumental reasons (i.e. better jobs, economic advantage, and importance for career). Urdu and Punjabi, on the other hand, hold 'integrative motivation' as sources of shared community feelings, making friends, and a strong community connection.

Lastly, we review two additional studies on motivations and attitudes of undergraduate learners towards English that include participants from the South of Pakistan, from the provinces of Balochistan and Sindh. Soomro's (2016: 42) PhD study includes 300 university students aged eighteen to twenty-one from one university each from Balochistan and Sindh. Rasool & Winke's (2019) 229 participants are from three universities in Quetta, Balochistan, with an age range of eighteen to twenty-five and an almost equal gender distribution. Soomro's (2016: 105–109) participants are motivated to learn English as it helps internationally (e.g. admission at a foreign university, moving abroad), provides access to a wider range of information, and is a source of entertainment due to English language media. Similar findings are made by Rasool & Winke (2019: 58), that is, learning English is important to succeed academically and professionally. Undergraduates from both provinces also feel that English is not a threat to their mother tongue and local culture (Soomro 2016: 67). Instead, it is

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

an additional language which does not affect their mother tongue that has its own place in their lives (Soomro 2016: 73). Both of these studies reach similar conclusions regarding the motivations and attitudes to English learning and the language itself, and its relation with local languages.

The discussion on the selected studies on language attitudes shows that recognition of variation in English, especially with regard to PakE, may have increased since 1990s. However, highly positive attitudes towards 'Standard' and BrE accents remain largely unchanged. Additionally, Pakistani learners are motivated to learn English for better career and job, as well as due to the international and global power associated with the English language. At the same time, most learners of English do not see any effect of English on their mother tongues (or the national language, Urdu), showing positive attitudes to learning and using English. The studies, however, do not cover all provinces and language groups in Pakistan. They generally include questions related to domains of language use associated with English, Urdu, and regional languages (including mother tongues), but the situations made available by the internet (e.g. social media, email) have not been included in these studies. This section, hence, aims to cover most of the regions and provinces of Pakistan, and the main language-speaking groups to record and report their attitudes towards English, Urdu, and regional languages and the usage patterns associated with these languages.

2.2 Data Collection and Demographic Details

The data for this study have been collected using a questionnaire created using Microsoft Forms (shown in Appendix A). It was designed by following previous similar studies, for example, Khalid (2016); Labade, Lange, and Leuckert (2020). The first five questions are related to background information: age, gender, education, occupation, and place of residence. The sixth question asks whether the participants moved from somewhere else in Pakistan. This question was deemed ambiguous at the time of analysis, because it does not specify if the move of residence is happening between two provinces or within the same province. Hence, anyone who selected *yes* in the response to this question was excluded from the final analysis to keep only native participants of a given region in the study. Questions seven and eight are related to mother tongue and any additional (regional, national, or other) language learned by the participants. Both questions include an option for English, as well as an option to add their own reply if the language was not already listed. Question nine asks about language proficiency in English, Urdu, the main regional language (mother tongue) and any additional regional language participants may have learned.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

A response scale was provided with the options being 'none', 'beginner', 'intermediate', 'fluent', or 'advanced'.

Questions ten, eleven, and twelve are related to the use of three languages (mother tongue, Urdu, and English) in various situations: family, friends, colleagues, strangers (shopkeeper, bazaar), social media (personal), social media (professional), and email (personal or professional). The response to these three questions is recorded on a scale with options 'never', 'sometimes', 'often', 'usually', and 'always'. The last question consists of a set of nine statements related to motivations and attitudes towards learning English (1, 2), attitudes about PakE (3, 4), attitudes towards the use of Urdu (5, 6, 7), attitudes towards the use of mother tongue (8), and identifying with the language community who speak the mother tongue (9). Responses were to be given on a five point Likert scale with the options 'I strongly disagree', 'I disagree', 'Indifferent', 'I agree', or 'I strongly agree'.

The data were collected in the first half of 2025 in multiple stages. The sampling approach can be described as convenience sampling. In the first step, colleagues (i.e. university lecturers and professors) all over Pakistan were approached by the first and the third authors. They were requested to forward the online survey link to their colleagues and university students. This strategy helped in data collection from Punjab, Sindh, AJK, and to a certain extent from Balochistan. The second strategy was to post in various Facebook groups affiliated with Pakistani universities in different provinces, for example, Bacha Khan University Charsadda – BKUC. In parallel, the first author hired freelance assistants through Upwork.com from various provinces and regions of Pakistan. Their place of residence was confirmed from their Upwork.com profiles. Each freelancer was given the task to get fifty respondents by disseminating the survey link to their connections (college or university students, school/ college/university teachers, or other professionals) in their native province (or region), for example, Balochistan, GB. In some cases, more than fifty responses were collected from only one freelancer. For example, in the case of GB, a university lecturer working as a part-time freelancer was hired, who helped in data collection from his students and colleagues at a university in GB and in AJK. This way, two or more freelancers per province or region were used to ensure a diverse pool of participants. In total, more than 1,500 responses were collected. After removing problematic cases, for example, participants who had moved (see the discussion on the sixth question), we aimed to select 200 random cases from each province/ region. The final data table consists of 1,092 individuals.

The demographic details related to age, gender, and residence are shown in Figure 2. The figure shows that apart from Punjab and Sindh, other provinces

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

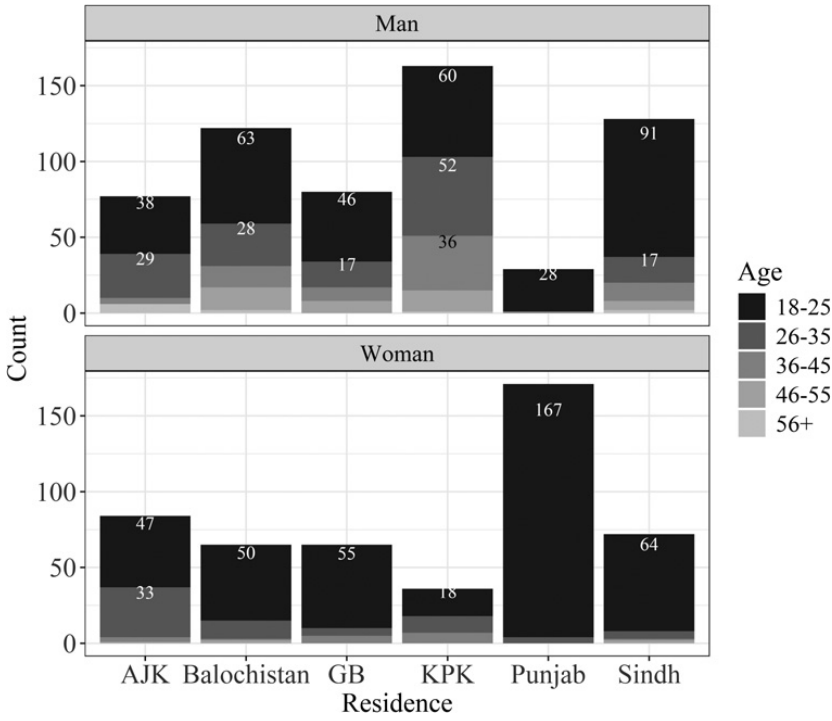


Figure 2 Demographic details with respect to age, gender, and residence

(KPK and Balochistan) and regions (AJK and GB) did not reach the target of 200 individuals, as many responses were discarded (see the previous paragraph). The lower panel shows that the highest number of women is in Punjab, followed by AJK and Sindh. Balochistan and KPK have a higher number of male participants than females. The top section in each bar shows that the participants are mainly eighteen to twenty-five years old in most provinces, followed by the age group of twenty-six to thirty-five. Older age groups are more prominent in males and in KPK, followed by Balochistan, Sindh and other provinces and regions. Other demographic details (not shown in the figure) are related to profession, education, and mother tongue. In terms of profession, the majority of the young participants are students, followed by teachers of any subject (mainly in KPK and Balochistan), and small proportions of journalists (in GB) and doctors (in AJK). Younger participants (as in Punjab) have ten years or twelve years of education. Older participants (e.g. twenty-six to thirty-five) indicate their education as fourteen years or sixteen years of education. These levels are present in all provinces. MPhil and PhD levels of education are mainly present in AJK, Balochistan, and KPK. The main languages spoken by the participants include Balochi, Balti, Hindko, Pahari/

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Pothwari, Pashto, Punjabi, Saraiki, Shina, Sindhi, and Urdu, along with some less spoken ones like Hazargi from Balochistan. Some respondents from AJK and Punjab (apart from the regional mother tongues) also indicate Urdu as their mother tongue, with the proportion of women being higher than men. The same is the case with Sindh, but that is expected due to the *mohajir* population of Karachi – Urdu-speaking Muslim immigrants from India who came to Karachi after the 1947 partition. Lastly, the main additional languages learnt by the participants are English and Urdu, followed by regional languages like Sindhi.

2.3 Data Analysis and Findings

In this subsection, we summarise the main trends in terms of language proficiency, language use in different situations, and responses to different statements in the last question of the survey. We construct ordinal regression models using the *polr* function from the package *MASS* (Venables & Ripley 2002) in the R programming language (R Core Team 2025). The final model is selected using the *stepAIC* function from the same library with the option *direction*='*both*'. For marginal effects plots we use the *marginaleffects* package by Arel-Bundock, Greifer & Heiss (2024) with the option *newdata*= '*balanced*' to account for unequal group sizes in several demographic variables. We also use the *car::Anova* (Fox & Weisberg 2019) function and the *emmeans* (Lenth 2025) package to construct analysis of deviance tables and find out significant contrasts based on the regression model objects (see Appendix B onwards).

2.3.1 Language Proficiency in English, Urdu, and Regional Languages

Overall language proficiency for regional languages, Urdu, and English is shown in Figure 3 in descending order of proficiency. Unsurprisingly, the participants indicate the highest proficiency for the regional languages that are mother tongues, followed by Urdu and English. Additional regional languages have the lowest proficiency. For English proficiency 'native' and 'fluent' get a combined probability of 43%, while intermediate proficiency has a 32% probability. The final model included gender, residence, education, occupation, mother tongue, and an interaction of gender, occupation, and mother tongue with the four languages as predictors. Marginal effects based on different demographic variables (not shown in the figure) show that education has a positive impact on language proficiency in English, Urdu, and both types of regional languages. Men report slightly higher language proficiency for English as compared to women, but the comparison is non-significant. The proficiency in an additional regional language is significantly better in men as compared to women irrespective of residence or mother tongue, which can be attributed to

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

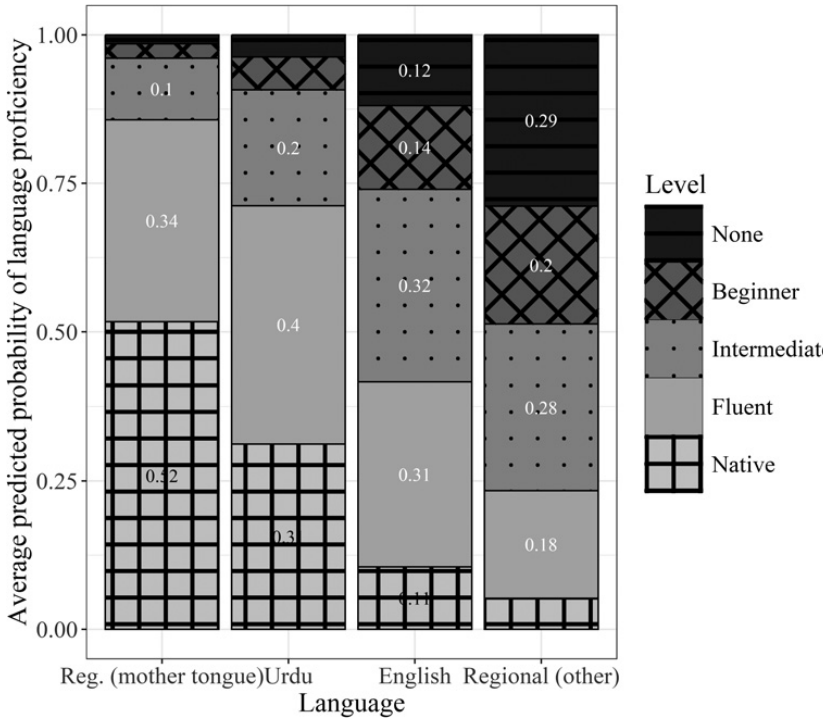


Figure 3 Language proficiency in English, Urdu, and regional languages

culturally more independent roles assigned to males, who are more likely to travel across Pakistan and learn other regional languages. Teachers and other professionals report higher proficiency in English, Urdu, and mother tongues in comparison to students. In terms of mother tongues, the speakers of Pahari/Pothwari, Punjabi, and Urdu indicate a higher proficiency in Urdu. The speakers of Balochi, Pashto, Sindhi, and smaller languages, on the other hand, report lower proficiency in Urdu, showing significant differences from some language-speaking groups (see Appendix B). English proficiency remains unaffected by mother tongue-based groups.

2.3.2 Use of Mother Tongue in Different Situations

We use a separate ordinal logistic regression model for the situational uses of the mother tongues because a lower number of situations are available for this question due to a spelling error in the two situations related to social media. The final model for situational use of mother tongues includes gender, residence, education, occupation, mother tongue, and situation as predictors of usage frequency, while gender, residence, and mother tongue are also included as

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

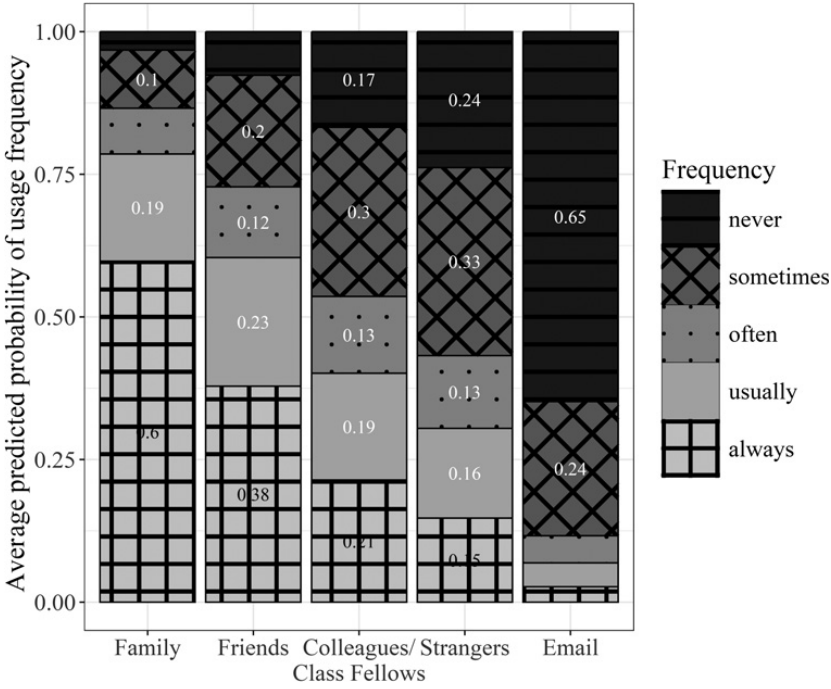


Figure 4 Usage frequency of mother tongue in different situations

interactions with situation. We do not include interactions of grouping variables like age and gender due to zero occurrences in certain categories, which lead to warnings by the *MASS:: polr* function. The overall usage frequency in Figure 4 shows that mother tongue is more likely to be used with family with a 79% combined probability of 'always' and 'usually'. The usage frequency decreases as the face-to-face (F2F) situations become more formal, as can be seen with higher probability of 'never' (24%) and 'sometimes' (33%) in the case of strangers (including with shopkeepers and in bazaar). The only internet situation included here is emails, and the participants are highly likely to select 'never' and 'sometimes' (a combined probability of about 89%). The use of the mother tongue on social media for responses where the spelling mistake mentioned at the beginning of this paragraph was corrected before the survey distribution also shows that a combined probability of 70% (personal social media) and 78% (professional social media) for the two options 'never' and 'sometimes'. Thus, overall the mother tongues are more likely to be used with family and friends.

Different participant groups also use the mother tongues significantly differently from each other. In other words, gender, residence, and mother tongue

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

have significant interactions with the five situations (see Appendix C for the contrasts, but we do not include the relevant marginal effects plots due to space restrictions). Apart from family-related situations, women are significantly less likely to use the mother tongue in all other situations when compared to men. These gender-related differences are generally stable across residence and mother tongue-based groups, with no reverse trends found in any subgroup. Residents of AJK, Balochistan, and Punjab are less likely to use the mother tongue with colleagues and class fellows. KPK respondents stand out as they are more likely to use the mother tongue with friends, colleagues/class fellows, and strangers (including with shopkeepers and in bazaar): they select 'always', 'often', or 'usually' in 50% or more cases. Pashto (the main language of KPK and parts of Balochistan) and Sindhi (especially the main language of rural Sindh) speakers are more likely to use the respective mother tongue with colleagues/class fellows and strangers as compared to the other regional languages, but these contrasts are not significant. Urdu native speakers are obviously more likely to use their mother tongue (i.e. Urdu) with friends, colleagues, and strangers. Punjabi and Saraiki speakers are significantly different from Urdu speakers in most of these situations, that is, instead of their mother tongues they most probably use Urdu with friends, colleagues/class fellows, and strangers.

2.3.3 Use of English and Urdu in Different Situations

The responses for situational uses of Urdu and English were pooled in one model to enable better comparisons, where language and situation were used as interaction effects for all other demographic grouping variables. The Analysis of Variance (ANOVA) table based on the model is available in Appendix D, while the marginal effects plot of overall usage for the two languages is shown in Figure 5. As it is clear from the panel 'Urdu', the Urdu language is most likely to be used with colleagues/class fellows, a rather formal situation. The probability assigned to 'always', 'usually', and 'often' in friends and family-related situations is higher than the situations involving strangers (which includes shopkeepers and bazaar as well). The use of mother tongue (especially by Pashto and Sindhi speakers as mentioned in Section 2.3.2) in the situations related to strangers may be one reason of an overall lower use of Urdu in this situation (see Section 2.3.3, paragraph 3). In the three internet-related situations, the use of Urdu decreases as the formality increases. Hence, Urdu is more likely to be used in personal social media posts, followed by professional posts, and least likely in emails. This is opposite and significantly different as compared to the usage of English in these three situations. This usage also includes Roman

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

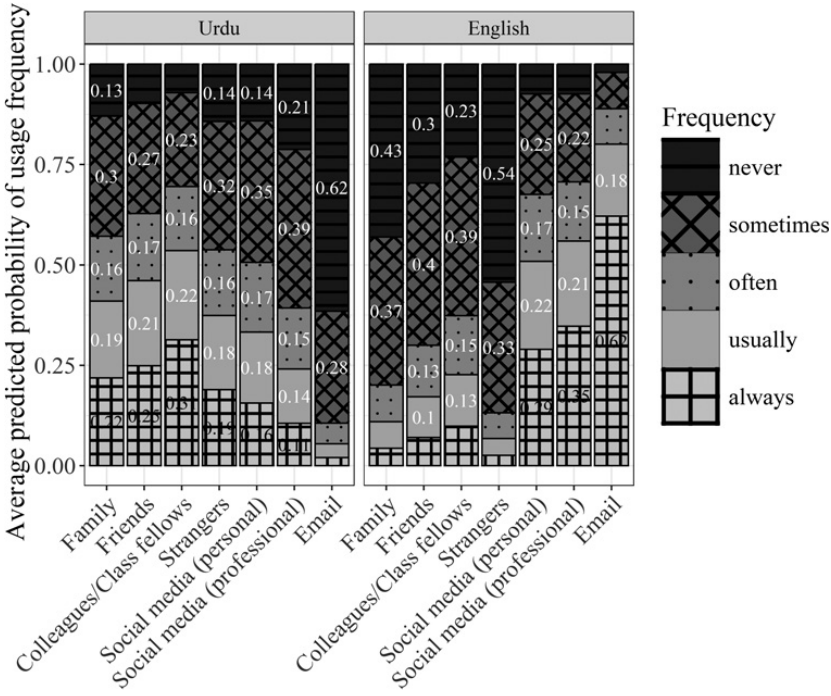


Figure 5 Usage frequency of Urdu and English in different situations

Urdu – a practice of writing Urdu in Roman letters. This kind of Urdu is more likely to be mixed with English words, as it is easier to spell English words in Roman letters in comparison to the native Perso–Arabic script of Urdu.

The use of Urdu among different subgroups also shows variation. The eighteen to twenty-five year olds are significantly more likely to use Urdu in emails in comparison to all other age groups, with men leading women in this age group as well, as discussed later in this paragraph. This ever so slightly higher likelihood of using Urdu could be due to the use of Roman Urdu among the younger generation. However, nothing conclusive can be said based on the current data. A more meaningful contrast is between males and females, where women are more likely than men to use Urdu in more formal F2F situations (friends, colleagues/class fellows, and strangers including with shopkeepers and in bazaar). The first two contrasts (i.e. friends and colleagues/class fellows) are also significantly different in both groups (see Appendix D). On the contrary, males are significantly more likely to employ Urdu in the internet-related settings of professional social media and emails. The residents (especially women) of AJK, Punjab, and in some cases those of GB are more likely to use Urdu in F2F situations, while those of Balochistan and KPK are among the least likely to employ Urdu especially with colleagues/class fellows and strangers (including in

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

bazaar and with shopkeepers). This is in line with the higher likelihood of using the mother tongue by these groups in similar situations, as noted in Section 2.3.2. Similar usage patterns can be assigned to relevant language-speaking groups: a higher likelihood to use Urdu in different F2F situations (e.g. friends, colleagues/class fellows) can be assigned to Punjabi and Pahari/Pothwari speakers (the residents of Punjab and AJK, respectively), while Balochi (one of the main ethnolinguistic groups from Balochistan) are less likely to use Urdu with friends and strangers. Some of these subgroups are significantly different from other subgroups, as shown in the three-way contrasts in Appendix D.

These trends of regional and mother tongue-based groups generally show the historical patterns of (dis)favouring the use of Urdu. The residents of Punjab (Punjabi, Saraiki, and partially Pahari/Pthowari speakers in these data) have been more open to adopt Urdu, even from the era of the British India (Singh 1997). Punjabis giving lesser prestige to Punjabi in comparison to Urdu is a well-documented phenomenon, for example, Abbas & Iqbal (2018: 210). Urdu has been the official language of AJK since 1968 (Anonymous 2002). This factor and many similarities between Punjab and AJK (e.g. the Pahari/Pthowari language-speaking group in these data, their geographical proximity) may also be the cause of similar results here. The historical relationship of Pashto and Sindhi-speaking communities with Urdu has not been as smooth. Pashto has been associated with Pashtun nationalism and a closer relationship of such movements with Afghanistan (where a sizeable Pashto speaking population resides). Even in the British Raj, the imperial officers preferred Urdu on Pashto in a bid to amalgamate the then North West Frontier Province (NWFP) better with the rest of British India (Rahman 1995: 156). After the creation of Pakistan and the inclusion of NWFP (now KPK) in the country, a strong nationalist movement involving Pashto and Pashtun identity has remained throughout Pakistan's history. On the other hand, Sindhi has been used in formal situations (e.g. courts) since even before 1947 (Rahman 1999: 25). After the partition, the influx of Urdu-speaking Muslim immigrants (*muhajirs*) from central India in urban Sindh increased the nationalistic sentiment of local Sindhi-speaking population. Sindhi is taught in primary schools, Sindhi and Urdu are co-official languages of the provincial government, and Sindhi is one of very few regional languages in Pakistan in which a master's degree can be obtained (Rahman 1999). Due to these historical and political reasons, Pashto and Sindhi speakers (and the corresponding provinces of Pakistan) are more likely to prefer their respective mother tongues instead of Urdu even in formal situations.

Overall, the use of English is more likely with friends and colleagues/class fellows in F2F situations (Figure 5, English). One reason could be that the interlocuters in situations related to family and strangers (including in bazaar and with shopkeepers) are less likely to be proficient in English (e.g. parents,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

other elders, shopkeepers); hence mother tongue or Urdu and more appropriate choices. English is much more likely to be used on the internet, where a probability of only 25% or lower is associated with 'sometimes' and 'never'.

Just like mother tongues and Urdu, various demographic groupings also show variation in (dis)preference of English in the seven situations. In terms of age, younger participants (eighteen to twenty-five) assign higher probability to 'always' and 'usually' in relation to the use of English in F2F situations related to family, friends, and colleagues/class fellows. This age group is significantly different from the age groups of twenty-six to thirty-five and thirty-six to forty-five at least in family-related situations. This shows an overall lowering barrier of formal language use: English might be taking the place of Urdu, and Urdu of the regional languages (mother tongues) in at least some parts of the society (also indicated by a comparatively higher use of Urdu in formal F2F situations by women as mentioned in Section 2.3.3, paragraph 2). Apart from that, on the internet the use of English increases with age. This is most probably because of a higher level of education that also increases with age in these data (see below in this paragraph). In terms of gender, women are ever so slightly more likely to use English in F2F situations of friends and colleagues/class fellows. Men indicate a very small preference for English usage with strangers (also significantly different from women; see Appendix D), but this contrast requires more in depth investigation (e.g. interviews with those who indicated positive responses). The residents of Punjab (in all three internet settings) and GB (in personal social media) are less likely to use English on the internet. Punjab is also significantly different from some other provinces and regions (e.g. Sindh) in email and professional social media. One reason may be a higher percentage of younger participants belonging to Punjab. Education does not show significant contrasts, but generally higher education (just like age) is associated with a higher likelihood of English usage, for example, MPhils and PhDs grouped in the education group of eighteen year plus. Lastly, Balochi speakers stand out in comparison to most other language speaking groups by showing a high dispreference for English in most F2F situations.

2.3.4 Attitudes towards Various Statements about English, Urdu, and Regional Languages

The attitudes of the Pakistani respondents about various statements on language use and language identity are shown in the form of (dis)agreement probability in Figure 6. Like the previous models, this one also included all demographic variables and one variable related to the nine questions as an interaction effect with all other predictors. The ANOVA table based on the model is shown in Appendix E. The overall comparison in Figure 6 shows that Pakistanis are

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

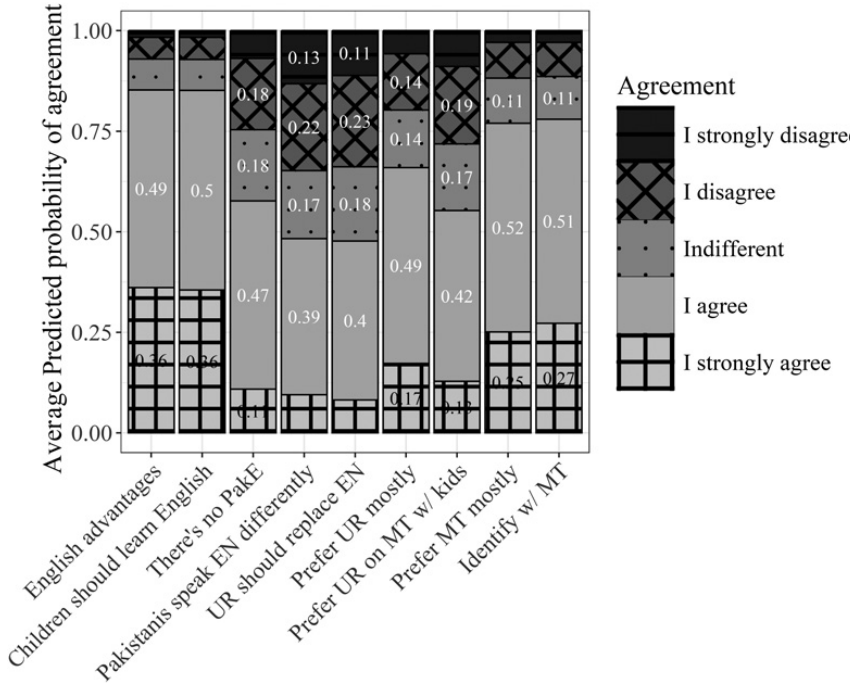


Figure 6 Attitudes towards various statements about languages

highly in favour of advantages associated with English and children learning English at school (86% probability for the first two positive statements). About 25% probability is assigned to 'disagree' and 'strongly disagree' in response to the third statement, the negative statement about PakE, which shows that most of the participants are not aware of PakE as a variety – a conclusion similar to that of Baumgardner (1995). The more general statement that 'Pakistanis speak differently' is more likely to be agreed with by a little less than 50%, which shows some general perception about the distinctiveness of the local speech. Statements 5–7 are related to the use of Urdu and show varying attitudes towards Urdu as the official language, and its use with kids and in different situations. These responses are explored in more detail in the next paragraph. The last two statements about the mother tongues show that the participants generally have highly positive attitudes about their mother tongues, which are expected from most language speaking groups.

Apart from the overall attitudes, there are also interesting differences among different demographic groups within the predictor variables. In terms of age, the youngest respondents (eighteen to twenty-five and twenty-six to thirty-five) are significantly different from the respective older groups in recognising the distinctive English speech of Pakistanis. Moreover, the same two groups also

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

have higher probability for '(strongly) agree' in response to 'I prefer Urdu in most situations', indicating that younger participants prefer Urdu over the regional mother tongues (eighteen to twenty-five are significantly different from forty-five years and older groups; the differences are slightly less pronounced between twenty-six to thirty-five and thirty-six to forty-five old women). Men and women significantly differ from each other on four statements: Women are more likely to (strongly) agree with the first two statements (advantages and learning English); They also prefer English by disputing 'Urdu should replace English' more often than men (male responses are slightly more varied across regions); Women indicate lesser preference for the mother tongue by disputing 'Prefer mother tongue mostly' more often than men. Other professionals (including journalists and doctors) are significantly less likely to agree with 'Pakistanis speak differently' and with 'Urdu should replace English' when compared with teachers. This shows that Pakistanis not involved in teaching are less aware of speech differences Pakistanis have, and that they favour English potentially due to its practical use in their respective fields. In terms of language speaking groups Balochi, Pashto, Sindhi, and small language speakers (e.g. Shina, and the ones categorised under Other) are more likely to dispute the statement 7 ('prefer to use Urdu over mother tongue with kids'), while Punjabi, Saraiki, and Urdu speakers are more likely to agree with it. Balochi and Pashto speakers are also more likely to agree with the statement 8 related to the use of the mother tongue (see significant contrasts in Appendix E).

2.3.5 Summary of Findings

The main trends related to age group, gender, residence, and mother tongue-based groups, and profession in three types of languages (regional mother tongues, Urdu, and English) can be summarised as follows.

The mother tongues are primarily used at home with family members, and are least likely to be used on the internet. KPK residents employ their mother tongue in a wider variety of situations. Pashto (spoken in KPK and Balochistan) and Sindhi (from Sindh) speakers prefer mother tongue even in formal situations like colleagues/class fellows where Urdu may be employed by other language speakers (e.g. Punjabi and Saraiki speakers from Punjab). Women are less likely to use mother tongue outside of home and family. In other words, men use mother tongue in more diverse situations. Most participants say they prefer mother tongue in most situations and also strongly identify with the relevant ethnolinguistic group. Balochi, Pashto, and smaller language speakers are more likely to do so.

The findings related to Urdu show that it is mainly associated with friends and work-related situations, and less likely to be used in professional social media and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

email. Respondents from AJK and Punjab and the ethnolinguistic groups in these areas (Pahari/Pothwari, Punjabi, Saraiki) indicate more preference for Urdu; for example, some report Urdu as their mother tongue, and others indicate a higher proficiency in Urdu. In contrast, the residents of KPK, Sindh, and Balochistan (speakers of Balochi, Pashto, Sindhi, etc.) indicate a lower proficiency in Urdu, are more likely to disprefer Urdu in formal F2F situations, and dispute statements related to the use of Urdu ('prefer to use Urdu over mother tongue with kids'). Women are more likely than men to employ Urdu in formal F2F situations. They also indicate Urdu as their mother tongue in AJK and Punjab more often than men. Younger participants indicate a higher preference of using Urdu over the regional mother tongues, even in family-related situations.

Lastly, the use of English mainly occurs in friends and work-related F2F situations and on the internet. The majority of participants does not agree with the statement that there is a PakE. However, a little less than 50% participants show awareness about the English speech differences of Pakistanis. Proficiency in English increases with age and education. Younger participants also indicate the use of English in some F2F situations. Most participants agree with the advantages associated with English and the need to teach English to kids, with women more likely to do so than men. Women also slightly prefer English in formal F2F situations compared to men, and disagree more often with the statement about replacing English with Urdu in official functions.

2.4 Discussion and Conclusion

We have presented the results of a language use and attitudes survey from more than 1,000 participants across different provinces and regions of Pakistan. The participants are not balanced across various demographic groups, especially with regards to gender, residence, and mother tongue speakers. Nevertheless, we hope to have uncovered and shown stable trends regarding the use of different languages and attitudes towards them across these demographic groupings. These trends can be broadly discussed under two main themes: the relationship, use of, and attitudes towards the regional mother tongues versus Urdu; the use of and attitudes towards English.

The detailed findings in Section 2.3 and their summary in Section 2.3.5 show that the use of regional mother tongues differs across various demographic groups, where certain groups favour the use of Urdu in place of regional mother tongues while others resist it. Younger participants, women, and residents of AJK, Punjab (Punjabi, Saraiki, Pahari/Pothwari speakers) generally favour lowering the barrier of the use of Urdu in more informal, intimate family-related situations. These findings are in line with the observations made by the three authors of Pakistani

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

origin in this study and the previous research and discussion presented in Sections 2.1 and 2.3.3. On the other hand, men, older participants, and the residents of Balochistan, KPK, and Sindh (along with relevant language-speaking groups) are more conservative and favour the use of mother tongues for the historical and political reasons discussed in Section 2.3.3. Two additional highly relevant demographic contrasts not included in these data but worth mentioning are: urban versus rural (where the urban class favours Urdu) and lower/working class versus upper social strata (where the middle class favours Urdu). In short, Urdu is increasingly taking the place of the regional mother tongues in certain demographics of Pakistani society.

In terms of perceptions about English, not many Pakistanis are aware of the concept of a 'Pakistani English', though many acknowledge that they speak English differently, which demands more policy level guidelines to increase such awareness. Nonetheless, Pakistanis are in favour of teaching English to school going kids and acknowledge the instrumental value associated with English in professional settings, a finding that also echoes previous research. The global power of English may as well be the reason that many participants disfavour the suggestion to replace English with Urdu as an official language. This has not happened in the second half of the twentieth century. It looks even less likely at the end of the first quarter of the twenty-first century, now that Pakistanis have been using the internet for the past two decades. English, being the language of technology, internet, and the graphical user interfaces (GUIs) of the software used by the majority of Pakistanis (Urdu or any regional language as the GUI language is an exception rather than a norm) most probably reinforces its hegemony even in non-internet domains. Also, the internet has given rise to many new situations and domains where English is highly likely to be used by Pakistanis, as is clear from higher likelihoods of English use in the three internet-related settings. This finding has been confirmed previously through corpus-based research (e.g. Shakir & Deuber 2019). However, previous attitudinal studies (see Section 2.1) on PakE generally do not list the internet as a significant context of English use.

To conclude this section, we have provided a comprehensive overview of English use by Pakistanis in various situations and their attitudes towards English and the various languages they speak in addition to English. In the next sections we further explore various features of PakE, and also focus on English use on the internet.

3 Phonology, Lexis, and Grammar

After presenting the attitudes of Pakistanis towards the English language, now we turn to a brief description of selected features of PakE. We focus on

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

distinctive phonological aspects, orthographic patterns, and selected lexical as well as grammatical features. In this process, we compare the variety with AmE and BrE as appropriate. In the following subsection we start with the phonological description (Section 3.1). The Section 3.2 describes the corpora used for the orthographic, lexical, and grammatical analyses, and the Sections 3.3, 3.4 and 3.5 in turn present the findings.

3.1 Phonological Features

The present study examines two complementary audio corpora: a controlled reading corpus of sixty university lecturers and MPhil candidates (see Appendix F for the reading passage), and a naturalistic corpus of fifty plus YouTube recordings of the public broadcaster Pakistan TV. A majority of these videos are political talk shows, interviews, morning shows, and so forth. The audios of the controlled reading corpus were listened to and analysed by the third author, and the presence or absence of distinctive PakE pronunciation features was marked. Afterwards example words were selected (e.g. *technology, better* for retroflexion) and audio clips containing these were extracted from the audios of the YouTube Pakistan TV corpus due to its studio quality data (one example word per unique speaker). Then, Google Gemini 3.1 Pro was used to obtain a narrow phonetic transcription of each word in the resulting 930+ clips of the 130+ speakers – the clip length was two to ten seconds and could contain multiple words. We performed a validation of Gemini's narrow transcription by selecting one example word (e.g. *think*) from five randomly selected speakers (eighty-three examples in total) and listened to the audio to confirm the presence or absence of the transcribed features. The validation results were mixed. While the human reviewer overwhelmingly agreed with Gemini on features like th-stopping, there was a large disagreement in the results related to retroflexion and monophthongisation. In the two cases involving disagreement, we provide evidence from previous acoustic studies (retroflexion) or conduct our own analysis to determine the presence or absence of the feature (monophthongisation). We show that the phonological characteristics of educated PakE arise primarily from systematic transfer from Urdu (the national language and usual medium of early schooling) and secondarily from regional languages such as Punjabi, Sindhi, and Pashto. The result is a highly homogeneous and an instantly recognizable accent of PakE.

3.1.1 Consonantal Features

Th-stopping, retroflexion, de-aspiration, and rhoticity are the main consonantal features that reflect substrate influences and are present in educated speech. Th-stopping represents the most invariant marker of PakE: the interdental

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

fricatives /θ/ and /ð/, absent from Urdu, are overwhelmingly replaced by dental stops [t̪] and [d̪]. Gemini transcribed 71% cases of *think* (Audio 1_think.mp4. Audio of think. Audio file available at https://doi.org/10.1017/9781009665308) with [t̪] and 90% cases of *this* (Audio 2_this.mp4.. Audio of this. Audio file available at https://doi.org/10.1017/9781009665308) with [d̪] in the YouTube audio corpus, which is consistent with prior documentation of these features (Mahboob & Ahmar 2004; Rahman 1991; Shahid et al. 2024).

Retroflexion of alveolar /t/ and /d/ to [ʈ] (as in /ˈbe.ʈər/ (Audio 3_better.mp4. Audio of better. Audio file available at https://doi.org/10.1017/9781009665308)) and [ɖ] (as in /ɖuː/ Audio 4_do.mp4. Audio of do. Audio file available at https://doi.org/10.1017/9781009665308), respectively, has been confirmed through acoustic studies of PakE (Syed, Bibi, & Shafi 2024: 69). It occurs especially intervocalically and across morpheme boundaries. In our YouTube corpus, Gemini identified 5%–30% instances of /t/ and /d/ as retroflex stops depending on their position (start, middle, or end) in words like *do*, *top*, *end*, *fast*. A majority was identified as dental stops ([t̪] and [d̪]), which was deemed incorrect by the human reviewer. Since previous acoustic studies have confirmed this feature in PakE, the present results need further investigation.

The neutralisation of /v–w/ merges the English contrast in favour of a labial approximant [w] or [ʋ], reflecting Urdu phonology, which is found in 35% examples of *very* (Audio 5_very.mp4. Audio of very. Audio file available at https://doi.org/10.1017/9781009665308) in our data transcribed by Gemini.

The voiceless plosives /p/, /t/, and /k/ in positions where AmE (General American or GA) and BrE (Received Pronunciation or RP) demand strong aspiration are regularly de-aspirated, a finding supported by acoustic analyses of voice onset time in PakE (Yasir et al. 2021). The variety is fully rhotic, that is, postvocalic /r/ is always pronounced, while /l/ remains clear [l] in all environments; velarized [ɬ] does not occur often (Jadoon & Ahmad 2022). Gemini found a majority of the example words pronounced in this way, that is, de-aspirated [p] in *points* (Audio 6_points.mp4. Audio of points. Audio file available at https://doi.org/10.1017/9781009665308), rhotic [r] or [ɹ] in *far* (Audio 7_far.mp4. Audio of far. Audio file available at https://doi.org/10.1017/9781009665308), and clear [l] in *all* (Audio 8_all.mp4. Audio of all. Audio file available at https://doi.org/10.1017/9781009665308) – see Table 1 for a comparison with RP and GA.

3.1.2 Vocalic Features

Monophthongisation of closing diphthongs is among the most systematic vocalic processes: GOAT (/əʊ/ ~ /oʊ/) is almost invariably [oː] (e.g. Audio 10_social.mp4. Audio of social. Audio file available at https://doi.org/10.1017/9781009665308), while FACE (/eɪ/) is regularly [e] (e.g. Audio 9_education.mp4. Audio of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Table 1 Distinct consonantal features of PakE in comparison to RP and GA. Audio files available at https://doi.org/10.1017/9781009665308

Word	RP	GA	PakE
Think	/θɪŋk/	/θɪŋk/	/t̪ɪŋk/[Audio 1_think.mp4]
This	/ðɪs/	/ðɪs/	/d̪ɪs/[Audio 2_this.mp4]
Better	/ˈbetə(r)/	/ˈbetər/	/ˈbɛ.tər/ or /ˈbe.t̪ər/[Audio 3_better.mp4]
Do	/duː/	/duː/	/d̪uː/[Audio 4_do.mp4]
Very	/ˈveri/	/ˈveri/	/ ʋɛri/ or / ʋe.ri/[Audio 5_very.mp4]
Points	/pɔɪnts/	/pɔɪnts/	/pɔɪnts/ (no aspiration)[Audio 6_points.mp4]
Far	/fɑː(r)/	/fɑːr/	/faːr/[Audio 7_far.mp4]
All	/ɔːl/	/ɔːl/	/ɔːl/ (clear l)[Audio 8_all.mp4]

Note: RP and GA transcriptions are taken from Oxford Advanced Learner's Dictionary, PakE transcriptions are based on our impressionistic analysis and Gemini 3.1 Pro.

Table 2 Distinct vocalic features of PakE in comparison to RP and GA

Lexical set	Word	RP	GA	PakE
FACE	*Education*	/ˌedʒuˈkeɪʃn/	/ˌedʒuˈkeɪʃn/	/ɛdʒuˈkeːʃ(ə)n/[Audio 9_education.mp4]
GOAT	*Social*	/ˈsəʊʃl/	/ˈsəʊʃl/	/ˈsoːʃ(ə)l/[Audio 10_social.mp4]

education. Audio file available at https://doi.org/10.1017/9781009665308), though slightly more variable (see Table 2 for a comparison with RP and GA). Monophthongisation of these vowels has been observed in earlier empirical work on PakE as well (Jadoon & Ahmad 2022; Mahboob & Ahmar 2004; Shahid et al. 2024), and also matches typological descriptions of diphthong simplification in new Englishes, for example, Bobda (2007). The absence of low front /æ/ results in TRAP being raised or backed to [ɛ] or [a], which is also documented across studies of PakE vowel systems (e.g. Yasir et al. 2021).

For monophthongisation, Gemini identified most instances of the GOAT and FACE vowels as diphthongs, which was disputed in all example cases in the human review. Thus, a short analysis of relevant words (e.g. *so, make*) on the same data was run by calculating F1 and F2 values at onset 20% and offset 80% using WebMaus and the *praat-parselmouth* library in Python (Jadoul, Thompson, & De Boer 2018; Boersma & Weenink 2026). The results show that the Pakistani speakers monophthongise both GOAT and FACE vowels in a majority of cases (see Figure 19 in Appendix F).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

3.1.3 Prosodic and Suprasegmental Features

Three features are notable at the suprasegmental level. Firstly, syllable-timed rhythm characterises all 100+ sources across both audio corpora, with roughly equal syllable duration and minimal reduction of unstressed vowels. This contrasts sharply with the stress-timed rhythm of RP and GA and represents perhaps the most immediately salient characteristic of PakE, which has also been documented in previous perceptual and acoustic studies (Asgher, Kashif Shahzad, & Hanif 2023; Jadoon & Ahmad 2022; Rahman 1991; Shahid et al. 2024). Secondly, lexical stress shows a strong tendency towards penultimate prominence in polysyllabic words, which reflects Urdu patterns and is documented consistently across empirical investigations (e.g. Yasir et al. 2021). Thirdly, intonation is characterised by narrower pitch range and frequent level or rising terminals on declaratives, where RP and GA would employ falling pitch (Mahboob & Ahmar 2004).

3.2 Corpora Used for Spelling, Lexis, and Grammar

We use two corpora of PakE for the analyses in the remainder of Section 3. The first one is the Pakistani Written English (PWE) corpus compiled by Muhammad Asim Mahmood (2009) and Rashid Mahmood (2009). The corpus consists of approximately 2.1 million words of standard written and edited PakE published mainly in the 2000s. The corpus includes press-related registers (columns, editorials, letters to the editor, news) as well as magazine articles from local English magazines like The Mag. Other notable registers include various types of books (including English textbooks written and published locally), business letters, legal decisions, interviews, literary registers (novels, English short stories, translations of Urdu short stories), fatawa or opinions of Islamic scholars on Islamic law, research papers, PhD theses, and texts from various websites (e.g. introductions from the 'About us' pages). The corpus includes most of the registers of the Brown family corpora, but the proportions are different, as the overall size also indicates. We use the complete corpus for the study of spelling variation and lexis. However, for comparative analyses of grammatical features (e.g. passive voice), we only use three register categories – editorials (newspaper columns, editorials, that is, columns by the newspaper editorial staff, and letters to the editor), news reports and research papers – from PWE that have equivalent categories in the AmE and BrE corpora (see Section 3.2, last paragraph).

As a secondary, more contemporary and informal source of data, we use the South Asian Online Englishes (SAOnE) corpus compiled by Shakir & Deuber (2023). The corpus contains various Computer-Mediated Communication (CMC) registers – newspaper comments, tweets, text messages, web forums,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

and websites – where English is the main language and comprises comparable data from six countries (Bangladesh, India, Pakistan, Sri Lanka, UK, and US). Each register category includes one million words, except text messages, with only 500,000 words for the four South Asian (SA) countries. For the analyses of spelling variation and lexis, we utilise the Pakistani component of the corpus, which comprises approximately 4.5 million words. We also selectively use the other sub-components of the corpus to confirm the presence of a certain feature in the other five varieties of SAOnE. For the analyses related to grammatical features, we use the AmE, BrE, and PakE components from the 10% sample of SAOnE used by Shakir (2024) consisting of fifty text files each from comments, tweets, web forums, and websites. This dataset was manually spellchecked, which makes Part of Speech (POS) tagging and parsing more reliable.

For the two native varieties of English, we use the Crown (AmE) and China-Lancaster-Oslo-Bergen (CLOB) (BrE) corpora compiled by Xu and Liang (2009). Both corpora consist of 1 million words each and follow the structure of the Brown family corpora. The data were collected between the years 2009 and 2012. For comparisons with PWE, we use three categories: 'A' (press reportage or news reports), 'B' (press editorials), and 'J' (learned that mainly includes research publications). The corpora were selected because the compilers very kindly shared the text files with us, which enabled us to tag and parse the data for the present analyses, which would not have been possible with the AmE06 and BrE06 corpora available only through the CQPWeb platform of Lancaster University.

3.3 AmE versus BrE Spellings in PakE

AmE versus BrE spelling patterns have been previously described in detail, for example, Totti (2008). We adopted the spelling search patterns used by Deuber and Hänsel (2019: 74) based on Totti's description, excluding variants that required manual filtering using Keyword in Context (KWIC) (e.g. check vs cheque). The spelling patterns were extracted by using the spellcheckers (both for AmE and BrE) provided by the Python library *pyenchant* and a Python script (Merejkowsky 2025; Python Core Team 2025). The words on the given pattern were first searched and then English Word Net was used to check if both words refer to the same concept, if yes they were extracted. At the end, we obtained 333 spelling pairs for PWE and 390 pairs for SAOnE-PK, the difference being largely attributable to the latter's larger size. Table 3 presents the results of the main patterns, followed by miscellaneous patterns represented by a single item only. The last two rows show the total raw frequency for each variety (first with s–z variants, and then without).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Table 3 AmE versus BrE spellings in PWE and SAOnE (raw frequency)

Corpus (Variety) ➔	PWE (AmE)	PWE (BrE)	SAOnE-PK (AmE)	SAOnE-PK (BrE)
Pattern ↓	–	–	–	–
<-ense> vs <-ence> (e.g. defense vs defence)	59	414	358	513
<-gram> vs <-gramme> (e.g. program vs programme)	186	477	751	233
<-iz-> vs <-is-> (e.g. finalize vs finalise)	3099	1524	4053	1798
<-l-> vs <-ll-> (e.g. labeled vs labelled)	125	248	353	725
<-ll-> vs <-l-> (e.g. enrollment vs enrolment)	74	77	153	85
<-log> vs <-logue> (e.g. catalog vs catalogue)	5	2	13	196
<-or> vs <-our> (e.g. favor vs favour)	265	1182	856	1463
<-ter> vs <-tre> (e.g. meter vs metre)	141	428	588	522
<-yz-> vs <-ys-> (e.g. analyze vs analyse)	89	105	90	111
Cozy vs *Cosy*	5	5	8	6
Gray vs *Grey*	21	46	29	175
Tire vs *Tyre*	7	13	64	114
Skeptic vs *Sceptic*	8	14	9	20
Sum overall	4084	4535	7325	5961
Sum without s-z variants (Percentage)	891 (23.5%)	2901 (76.5%)	3174 (44%)	4046 (56%)

The results from PWE confirm Muhammad Asim Mahmood's (2009: 116–119) study based on the same corpus, albeit using a broader set of spelling patterns than the ones used here. He found that BrE spellings were more frequently used but only with a ratio of 1:1.66, that is, for each AmE spelled word there were 1.66 BrE variants used in the text. The higher frequency of AmE variants was partly due to the <-z-> versus <-s-> variants, as is also observable in the current results for both corpora (Table 3, row 5). As also noted by Deuber and Hänsel (2019: 49), the <-z-> variants are also used in BrE and in their data of Caribbean newspapers the frequency of these variants is much higher than the <-s-> variants. Setting aside the s/z alternation, we see that British conventions account for 76.5% in PWE and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

56.0% in SAOnE-PK (Table 3, last row). The categories <-ense> versus <-ence>, <-log> versus <-logue>, <-or> versus <-our>, and <-ter> versus <-tre> show a marked preference for BrE variants in either or both corpora. However, the dominance of BrE spelling decreases when SAOnE-PK is considered, which can be interpreted in two ways: (a) Informal English CMC and AmE defaults in smartphones promote AmE spellings, and (b) Younger Pakistanis (e.g. Gen Z) are more open to AmE spellings.

The corpus-based results are also confirmed by attitudinal studies of Pakistani students on AmE versus BrE spellings. For example, Ali, Ishtiaq, & Shahid (2020) conduct a survey with sixty-six students from Kohat University of Science and Technology and the National University of Modern Languages. They report a clear preference for American forms (68% of the participants), while the rest (32%) report a preference for British spellings (Ali, Ishtiaq, & Shahid 2020: 103). A 2024 survey of 500 Pakistani graduates from the University of Gujrat and GIFT University, Gujranwala, finds that 60% of respondents prefer American over British forms (40%), and attribute this to American media saturation and the US global position (Azam, Shaheen, & Jahan 2024: 1641). To sum up, it can be said that certain BrE spelling variants (e.g. <-our>) are still dominant in PakE, especially in formal and written English. Otherwise, AmE spellings are making their way in informal English communication (e.g. on the internet), and the younger generation generally prefers AmE spellings, especially the <-z-> variants.

3.4 Lexical Features

This subsection deals with neologisms (using suffixes *-ism*, *-ian*), reduplication, and cultural vocabulary (adjectives and nouns) in productive patterns (bigrams of Urdu + Urdu and Urdu + English words).

3.4.1 Productive Patterns with English Suffixes

The use of English suffixes to form new nouns has been reported by Rashid Mahmood (2009: 255–257) based on PWE. One of the unique patterns he observes is the suffix *-ism* to create neologisms like *adhocism*, *red-tapism*, *landlordism*, *militantism*, and *technocratism* that use English stems. There are also examples where the stem comes from Urdu, for example, *bradarism* (i.e. the practice of, e.g., voting collectively as a community based on caste). Additional examples of *-isms* from SAOnE-PK include:

(1) *fawadism*: behaving like a Pakistani cricketer, Fawad Alam, who focused on his performance instead of PR to get selected in the national team.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

(2) *bhuttoism*: The way of Zulfiqar Ali Bhutto and his daughter Benazir Bhutto, two assassinated members of a prominent Pakistani political family.

(3) *wadera-ism*, *jagirdari-ism*, *chaudhri-ism*: feudalism

(4) *patwarism*: Based on *patwari*, a negative political slang referring to supposedly *illiterate* followers or voters of Pakistan Muslim League (Nawaz), a local political party.

(5) *mullahism*, *mullaism*: The dominance of religious leaders

More recently it has been observed that Pakistani internet users employ the English suffix *-ian* to create adjectives that refer to the membership of an educational institute or a (political) organisation. The examples from SAOnE-PK (Tweets) include:

(6) *insafian*: a member or supporter of Pakistan Tehreek-e-Insaaf

(7) *youthian*: a negative slang to refer to the supporters of the same political party

(8) *xxxxx_nedian*: a student of the Nadirshaw Edulji Dinshaw (NED) university

(9) *ex-uetian*: a former student of University of Engineering and Technology (UET)

(10) xxxxxxxx*comsian*: a student of COMSATS university

Example (11) shows the use of *-ian* with the name of Iqbal (the national poet and philosopher of Pakistan) to refer to his philosophy.

(11) Thus, the *Iqbalian* 'self' is an entity in which the body and the soul have to work together (PWE, Books).

3.4.2 Reduplication

Reduplication is a very common process in Urdu and the regional languages of Pakistan. Word reduplication is used to create emphasis as well as to convey certain particular meaning. The present analysis has been conducted on PWE and SAOnE-PK to identify cases of reduplication and explore the influence of Urdu and the regional languages on the PakE communication. PWE has a very few cases of reduplication. These include adjectives (e.g. *bad bad*), adverbs (e.g. *far far*, *chupke chupke* 'secretly secretly'), and interjections (e.g. *dammit dammit*). The examples show that emphasis is the main reason for reduplicated lexical items, whether English or borrowed from Urdu. These items occur in the magazine articles, press reportage, and English short story sections of the corpus. SAOnE-PK has a larger variety of such items due to the informal CMC data and its larger size. In the remainder of this subsection, we describe various types of reduplication in this dataset.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Table 4 Reduplication in SAOnE: Part of speech and origin of the lexical items (Tokens and Types)

POS	English (shared with UK, US)	Non-words and interjections	English (unique to PK, South Asia)	Borrowed items (e.g. from Urdu)	Total
adjective	283 (35)	0	30 (28)	4 (4)	317 (67)
adverb	247 (30)	0	10 (9)	6 (6)	263 (45)
affix	7 (1)	0	0	0	7 (1)
discourse items	184 (26)	0	26 (10)	55 (24)	265 (60)
idiom	73 (12)	0	0	0	73 (12)
interjection	4 (3)	218 (35)	0	12 (2)	234 (40)
negation	51 (2)	0	0	8 (6)	59 (8)
noun	76 (15)	0	46 (26)	16 (9)	138 (50)
verb	21 (13)	0	29 (23)	0	50 (36)
Total	946 (137)	218 (35)	141 (96)	101 (51)	1406 (319)

Note: The cells show number of Tokens and Types (in brackets). Spelling variants (e.g. soooo, so) were considered separate types.

Table 4 presents an overview of the reduplicated lexical items in SAOnE-PK that occur in otherwise English sentences. We also checked if the reduplicative use was attested in the other regional components of SAOnE. Hence, the column 'English (shared with UK, US)' shows the items found in SAOnE-PK that were also attested at least once in the UK and/or US data along with the other three SA varieties. The number of tokens and unique types shown in the table, however, are only based on the SAOnE-PK data. This column has the largest number of reduplicated items in terms of tokens and types, with the main categories being adjectives (e.g. *deep deep*, *long long*, *naughty naughty*), adverbs (e.g. *much much*, *really really*, *very very*), and discourse items (e.g. *ok ok*, *please please*, *yes yes*). This shows that reduplication (especially for emphasis) occurs in most varieties of English present in SAOnE.

Furthermore, the column 'Non-words and interjections' shows the second highest frequency items (thirty-five types occurring 218 times), which are generally used in CMC irrespective of language, for example, laughter-related interjections (*ha ha*, *he he*, *lol lol*), and other interjections (*blah blah*, **cough cough**, **clap clap**, *eh eh*, *tch tch*, *tsk tsk*, *uff uff*). The examples with asterisks around are the CMC way to show the supposed sound or action the writer makes/does while typing the message. More familiar interjections were tagged separately, for example, *wow wow* is English specific (listed under the column 'English (unique to PK, South Asia)'), while its equivalent in meaning *wah wah* is an Urdu borrowed item that is categorised under interjections in the column 'Borrowed items (e.g. from Urdu)'. Overall,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

reduplication in SAOnE-PK occurs most frequently in tweets, followed by web forums, text messages, comments, and websites.

The columns 'English (unique to PK, South Asia)' and 'Borrowed items (e.g. from Urdu)' contain reduplicated items specific to Pakistan (in addition to the SA region). The former column shows ninety-six types (total frequency 141) that are English reduplicated words, but only found in SAOnE-PK or the PK, as well as SA data only. These include auxiliary verbs such as *must* in (12) or lexical verbs like *see see*, *send send*, and so forth. (13) has an adjective reduplicated, while *congrats* (a discourse item) is used two times in (14). In all of these cases, the purpose is to emphasise or intensify the meaning. In (15), the noun *condition* is reduplicated as the object of the verb *play*. This structure also occurs with other nouns: *playing Dr Dr*, *playing Twitter Twitter*, *playing Kashmir Kashmir*. In each case, the meaning is satirical, and the writer wants to emphasise that someone is making a mockery of a serious issue. This construction is a literal translation from Urdu. Similarly, numbers can be reduplicated in Urdu for emphasis, and the same can be seen in (16) with *two*.

The column 'Borrowed items (e.g. from Urdu)' shows fifty-one Urdu reduplicated lexical items that were found in English sentences. Discourse-related words are the most common category, for example, *Mashallah* 'Praise be to God' in (17) is on the same pattern as *Congrats* in (14), showing the tendency of reduplicating discourse items. The noun *parchi* ('receipt') is reduplicated in (18) to refer to the irregular selection of a cricketer in the national cricket team based on the reference of their father, who is an ex-cricketer, instead of a merit-based selection. Other examples include interjections like *chee chee* ('eww eww'), the discourse marker *jee jee* ('yes yes'), and negation like *nahe nahe* ('no no'). The examples and discussion show that reduplication is a productive process in the data of PakE analysed here.

(12) We all *must must* work together to bring peace in the whole region . . .. (SAOnE-PK, Comments)

(13) we have whatsapp group nd members full there so our *honarable honarable* admin make fb group [. . .] (SAOnE-PK, Text messages)

(14) *Congrats Congrats* Well played 👍 (SAOnE-PK, Tweets)

(15) When our reserves were going down during first six months and Asad Umar was playing *condition condition* with IMF. (SAOnE-PK, Web forums)

(16) [. . .] whether she was elected for one term or *two two*, does not matter. (SAOnE-PK, Tweets)

(17) *Mashallah Mashallah* blue shirt with black 😍❤️ (SAOnE-PK, Tweets)

(18) *Parchi Parchi* ('receipt receipt'), An Ex-Cricketer Father's *Parchi* ('receipt'). (SAOnE-PK, Tweets)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

3.4.3 Productive Combinations of Local Nouns and Adjectives in English CMC

The aim of the present subsection is to explore the use of cultural vocabulary in the form of noun + noun and adjective + noun combinations in PakE CMC. These combinations can be Urdu (i.e. both items are Urdu but used in English sentences) and English–Urdu mixed (i.e. Urdu adjectives and nouns in combination with English nouns). Using a frequency list from SAOnE-PK, Google Gemini was asked to classify the words that may be relevant to Pakistan's history, geography, politics, culture, religion, food, flora, fauna, clothing, and local languages. Then, we used the Multi-Feature Tagger of English (MFTE) (Le Foll & Shakir 2023; Le Foll & Shakir 2024) to extract productive two-word combinations tagged as Noun Compound (*NCOMP*) (i.e. compound nouns) where one item was from the selected vocabulary list. Finally, we selected those combinations that were used in an English sentence, resulting in 460 such bigrams. After manually checking the bigrams and their KWIC, we selected fifty-eight most interesting items, of which some occur only in one combination (e.g. *dheet* 'audacious' only occurs once in *dheet people*) and others are highly productive (e.g. *Eid* occurs in dozens of different combinations like *Eid offer*, *Eid prayer*, *Eid day*, *Eid vibes, Eid selfie*, etc.). At the same time, we classified each bigram according to different thematic categories and language (English, Urdu, mixed).

Figure 7 shows the thematic categories along with the language (Urdu or mixed) and their proportions in different registers of SAOnE-PK. The only locally relevant English word is an abbreviation Computerised National ID Card (*CNIC*), which occurs in bigrams like *CNIC number*, *CNIC copy*, *CNIC card*, and *CNIC condition* as in (19). The bigrams with mixed English and Urdu words are more frequent and diverse in terms of themes. One of the most frequent Urdu (also Hindi) words in the cultural category is *desi* ('native' or 'indigenous'), which occurs with many English words, for example, *desi parents*, *desi food*, *desi liberals*, *desi community*, *desi shampoo*. *Ghairat* ('honour') is comparatively more specific to Pakistani culture, occurring as *ghariat brigade* and *ghairat brigades* ('groups(s) of people who like to morally police others'). Another culturally relevant word is *rishta* ('marriage proposal'), which occurs in bigrams like *rishta process*, *rishta lady* ('the professional agent who mediates the process'): (20) shows the use of the latter.

Ethnolinguistic words are adjectives referring to different ethnolinguistic groups in Pakistan, for example, *Baloch/Baluch* (the main ethnolinguistic group from Balochistan), *Mohajir* (Urdu-speaking immigrants, see Section 2.2), *Pashtoon/Pukhtoon* (the Pashto-speaking ethnolinguistic community). These are used in bigrams like *Baloch people*, *Baloch tribes*, *Mohajir identity*, *Pashtoon belt*, and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

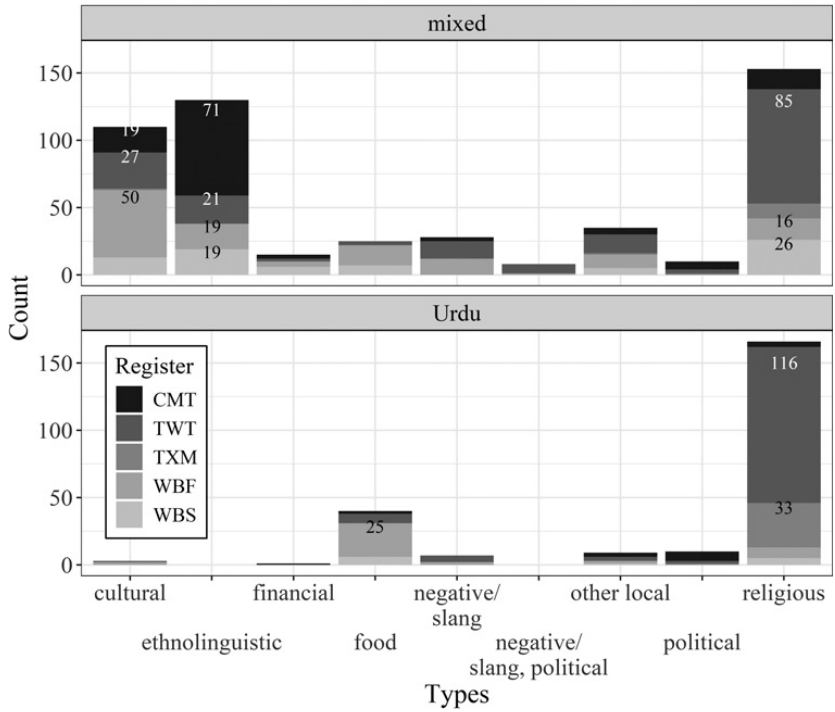


Figure 7 Thematic categories of local noun + noun and adjective + noun combinations in SAOnE-PK

so forth. In (21), the bigram *Pukhtoon revenge* refers to the tradition of revenge in Pashto speakers. The negative/slang category and political slang are thirty-six in number and are mainly used in Twitter. These include Urdu adjectives like *bakwas* ('crap') *player*, *ghunda* ('rascal') *police*, *jaahil* ('illiterate') *leader*, *youthia mentality* (mentality of Pakistan Tehreek-e-Insaaf (PTI) supporters), *patwari league* (Pakistan Muslim League Nawaz), and so forth. The word *lifafa* literally means envelope; in (22), it refers to journalists that supposedly accept bribe cash in envelopes. The category of 'other local' includes adjectives like *Lahori* ('someone from Lahore') as in *Lahori blogger*, *Lollywood* (Hollywood-inspired name of Pakistani film industry), as in *Lollywood movies*, *Lollywood way*, *Lollywood icon*, and *gora* ('white') as in (24). Lastly, religious bigrams include examples like *haram* ('forbidden') *money*, *Jihadi* ('religious militancy') *tendencies* as in (23), *Iftar* ('breaking fast') *time*, *Nikah* ('marriage vows') *dress*, and so forth.

(19) Mr.Zaidi introduced *CNIC condition* for business at just Rs.50,000/- . . .(SAOnE-PK, Comments)

(20) We got the *rishta* through a *rishta lady* and they've been talking for awhile. (SAOnE-PK, Web Forums)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

(21) APS *Pukhtoon children* massacre was your fatal mistake. The famous *Pukhtoon revenge* will come. (SAOnE-PK, Comments)

(22) They All are *lifafa Journalists* and Media Groups. (SAOnE-PK, Tweets)

(23) Such jingoism led us to *Jihadi tendencies* which became the tenth fault line. (SAOnE-PK, Web Forums)

(24) Grew up avoiding street food for questionable hygiene and a *gora vlogger* comes and calls that shit better than KFC. 😐. (SAOnE-PK, Tweets)

The bigrams that are completely in Urdu mainly consist of religious and food-related nouns. The most common bigram in the former category is *Eid Mubarak* ('happy Eid'), which mainly occurs in tweets and text messages. Food-related bigrams (e.g. *garam masala*, *chaat masala*, *tikka* 'BBQ' *masala*, *gosht* 'meat' *masala*) are mainly present in web forums where cooking recipes are posted by the forum members. In (25), *dabbul roti* ('box bread') refers to sandwich bread that is usually baked in a metal box. Other local Urdu nouns can be used as is in English sentences, for example, *Chai wala* ('tea maker') in (26). The word *jalsa* ('political gathering') is used a few times in comments, text messages, and tweets. In (27), such a meeting in Lahore is mentioned.

(25) It is *dabbul roti* . . . (SAOnE-PK, Web Forums)

(26) You have a monster *Chai wala* next door . . . (SAOnE-PK, Web Forums)

(27) The *Lahore jalsa* was a failure [. . .] (SAOnE-PK, Comments)

This subsection has shown that Urdu nouns and adjectives are frequently employed by Pakistani users in informal English communication, and that some of these lexical items are highly productive.

3.5 Grammar

In this subsection, we discuss four grammatical features of PakE. Three of these (i.e. relativisers *that* vs *which*, passive voice (*be* and *get*), (semi-)modal verbs of obligation) are related to ongoing change in English (Deuber et al. 2022). These can help gauge the extent of change happening in PakE and its orientation to AmE or BrE. The fourth one (i.e. noun phrase (NP) complexity) may be helpful to uncover L2 features of PakE (e.g. use of phrases vs clauses as NP modifiers, as we show in Section 3.5.4).

3.5.1 Relative Clauses with That and Which

Leech et al. (2009: 227) observe a decline in the frequency of *which* relatives – as in (28) – and an increase in *that* relative clauses – for example, (29) – in written AmE

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

and BrE using the Brown family corpora from 1960s to 1990s. They attribute the increase of *that* relatives to colloquialisation (Leech et al. 2009: 234). In their analysis, the tendency of replacing *which* with *that* relativiser is more pronounced in the AmE corpora (Brown & Frown), than the British ones (Lancaster-Oslo-Bergen (LOB), Freiburg-LOB (FLOB)). Deuber and Hänsel (2019: 54) note that the relatively higher frequency of *which* relative clauses in BrE (including in the press language) can be a good indicator of a variety's inclination towards AmE or BrE. Hence, the aim of the present subsection is to determine the frequency of *that* versus *which* relative clauses in PakE as compared to AmE and BrE using the written and CMC corpora as described in Section 3.2.

The data were tagged (POS, Named Entity Recognition or NER) and parsed (constituency, dependency) with the *stanza* (Qi et al. 2020) library. Then, we followed the criteria adopted by Hinrichs, Szmrecsanyi, and Bohmann (2015: 814–815) and by Deuber and Hänsel (2019: 54–55) to extract only restrictive relative clauses using a rule-based approach and a Python script. We excluded clauses that were marked by a comma before the relativiser *which*, pied piping relative clauses involving *which*, clauses where the antecedent was *that* and the relativizer was *which*, clauses with a human noun (PERSON tag in the NER tags) as the antecedent and the relativiser was *that*, and where the clause could function as adjunct (i.e. *that* was replaceable by *when*, *where*, *how*). For the last case, a list of adjunct nouns was used (*day*, *time*, *year*, *place*, *reason*, *way*, *date*) and the clause was excluded from final count where such a noun was the head of the antecedent NP and the relativiser was *that*. We also excluded the clauses marked as 'obl' or oblique (i.e. adjunct clauses) by the stanza dependency parser. For examples of these cases, please refer to Deuber and Hänsel (2019) and Deuber et al. (2022: 284).

As a further step, we checked for complement clauses incorrectly extracted as relatives using Google Gemini 2.5 Pro. This last step removed 1–10% clauses from the total (the highest number was from PWE): (30) shows such a clause identified by our script as a relative clause but flagged by Gemini. The overall accuracy of relative clause identification (based on a 2.5% sample of the data) remains around 96% in both datasets.

(28) He pointed out the United States had committed a blunder *which would further deteriorate law and order situation in Iraq*. (PWE, News)

(29) It was clearly a failure *that needs to be looked at and*, *if possible*, *addressed*. (Crown, Press reportage)

(30) One can infer from this brief discussion *that Saras ingenious literary expression consists in the blending of some deviant patterns into non-deviant patterns in her language*. (PWE, Research papers)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

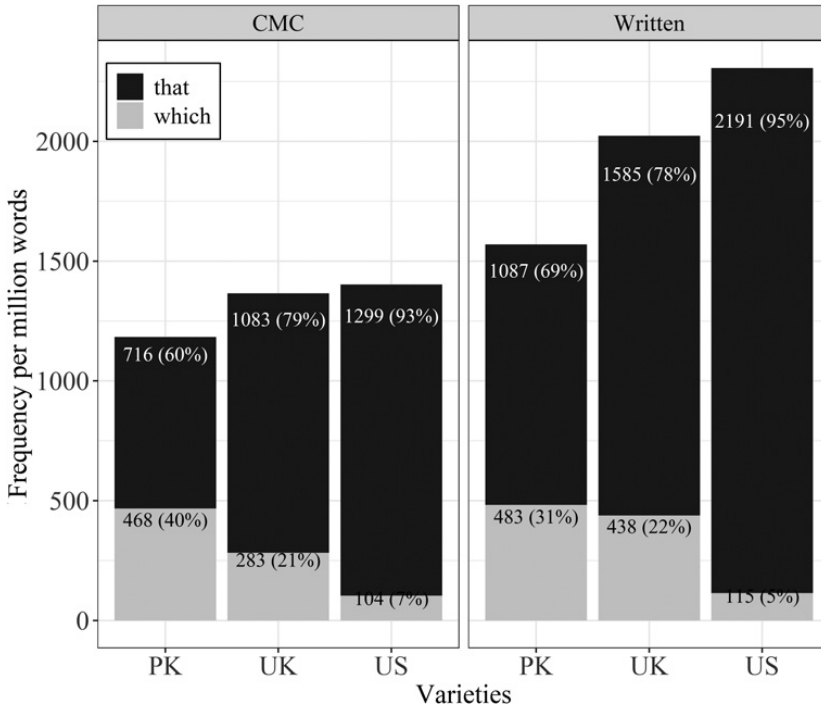


Figure 8 Relative clauses in written and CMC corpora of the three varieties of English

Figure 8 shows the per million word frequencies and proportions of the two types of relative clauses in the three varieties. The normalised frequencies of both types of relative clauses are much lower in both modalities of PakE, though the regional difference is less pronounced in CMC. The overall lower frequency of relative clauses might be attributable to the use of phrasal features (i.e. prepositional modifiers) instead of clausal modifiers in NPs, as also noted in Section 3.5.4 (last paragraph). The written English registers (news, editorials, and research papers) indicate that *which* relatives are the least prevalent in AmE, confirming previous studies on the Brown family corpora. The proportion of *which* relatives in PakE is higher than in both of the native varieties, and the differences are statistically significant based on overall chi-square comparisons using raw frequencies in both modalities (Written: χ^2 (2, N = 4496) = 332.04, $p < 0.001$; CMC: χ^2 (2, N = 3910) = 383.69, $p < 0.001$). Subsequently, PakE was also found significantly different when compared with both native varieties one by one (PK vs UK, PK vs US) using the function *pairwiseNominalIndependence* in the R package *rcompanion* (Mangiafico 2025; R Core Team 2025). Overall, the results show that these PakE registers (both in CMC and written) are different from their counterparts in AmE and BrE in the use of relative clauses.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

3.5.2 Be *and* Get *Passive*

In the past two decades, corpus-based research on passive voice in AmE and BrE has shown that the *be* passive, for example, (34), is declining in frequency, more in AmE and a little less in BrE, for example, Leech et al. (2009: 149, 156). At the same time, another variant of the passive voice, the *get* passive, as shown in (33), is very gradually increasing in frequency. Bohmann et al. (2025) have conducted a comprehensive analysis of *be* and *get* passive in varieties of English using very large corpora, for example, the Corpus of Historical American English (COHA) and the Global Web Based Englishes (GloWbE) corpus. Their results confirm previous observations that *get* passive has become more frequent with time (Bohmann et al. 2025: 129). This is especially true for central *get* passives, that is, the ones that are not formulaic as in example (31) or adjectival in nature, for example as in (32). The construction – though still mainly used in informal contexts – is also being increasingly used in formal contexts (Bohmann et al. 2025: 131). In terms of varieties, AmE leads with the highest proportion of *get* versus *be* (Bohmann et al. 2025: 163–164). PakE is among the varieties with the lowest proportion of *get* passives in their GloWbE data.

(31) [. . .] there's a chance you might actually enjoy it once you *get used to* it and gain some competency. (SAOnE, UK Web forums)

(32) But don't *get carried away.* (CLOB, Press editorials)

(33) The film's release also *got pushed back* from July 2014 to April 2015. (SAOnE, US Websites)

(34) However, he *was dispatched* without grace to Saudi Arabia immediately after landing. (PWE, Newspaper columns)

The aim of the present subsection is to explore the use of *be* and *get* passives in the three varieties using the corpora described in Section 3.2. These datasets include three written (editorials, news, research papers) and four CMC (comments, tweets, web forums, websites) registers. Hence, these are relatively more diverse than the web-based data used by Bohman et al. (2025). We have used MFTE (Le Foll & Shakir 2023; Le Foll & Shakir 2024) to extract the frequencies of the relevant tags *PASS* (*be* passive) and *PGET* (*get* passive) from the corpora. We then used Google Gemini 2.5 Pro to classify the *get* passive instances according to the six way classification of Collins (1996: 45–49): central as in (33), psychological (e.g. '*got frustrated*'), reciprocal ('*get married*'), adjectival as in (32), formulaic as in (31), and complex reflexive (e.g. '*Don't get yourself killed*'). A 10% sample of the *get* passives was used to check the accuracy of MFTE and subsequent classification by Gemini 2.5

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

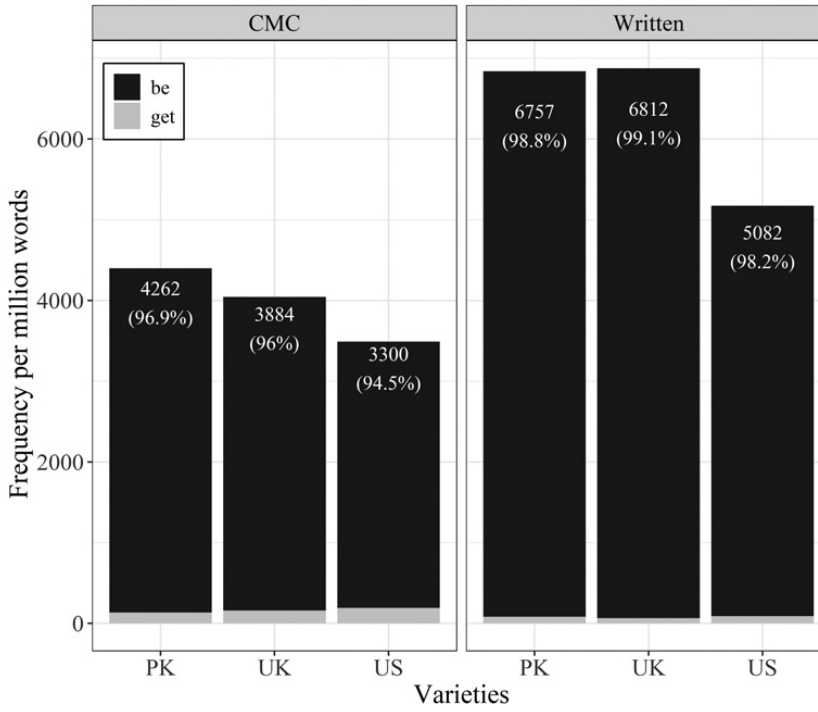


Figure 9 *Be* and *get* passives in written and CMC corpora of the three varieties of English

Pro. The accuracy of passive recognition by MFTE was 98.85%, while the accuracy of Gemini's classification of central versus non-central *get* passive was found to be 90%.

Figure 9 shows the words per million frequencies of *be* and *get* passives in the three varieties and two modalities. The *get* passive is slightly more frequent in CMC as compared to the three written English registers. AmE leads in the use of *get* passive in both modalities, which confirms previous studies. The PakE data trails the AmE and BrE data in CMC in the use of *get* passive. However, due to a slightly higher frequency of this variant in Pakistani newspaper columns, the frequency is overall higher than in BrE in the written medium. Otherwise, the main variant of the passive voice in PakE is *be* passives. It is either higher than in both of the native English varieties (in CMC) or near the BrE frequencies (in the three written registers). Deuber et al. (2022: 282) find a similar trend in Indian English. The differences in the use of *be* versus *get* passive in the written modality are not statistically significant among the three varieties (χ^2 (2, N = 15271) = 5.03, p = 0.081). However, these varieties do show significant differences in CMC (χ^2 (2, N = 11778) = 27.69, $p < 0.001$). Subsequent comparisons of PakE with

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

AmE and BrE are also statistically significant in CMC – performed using the function *pairwiseNominalIndependence* (Mangiafico 2025). Overall, passive voice and especially *be* passive appears to be comparatively more common in PakE.

3.5.3 Modal Verbs of Obligation

Shakir (2025) has conducted a study on the (semi-)modal verbs of obligation *must*, *have to*, *(have) got to*, and *need to* using the public registers (i.e. comments, tweets, web forums, and websites) in the SAOnE corpus. Like the other three SA varieties (Bangladeshi, Indian, and Sri Lankan Englishes), PakE has a higher preference for *must*, particularly in comments and tweets. The main use of *must* is in the root meaning to express personal opinions, referred to as 'discourse internal obligation' by the author. Example (35) shows the use of *must* from the SAOne corpus used by Shakir (2025). The two native varieties (i.e. AmE and BrE) mainly employ *need to* for this purpose. PakE (along with the other SA varieties), however, uses *need to* as well as *must*. The author attributes the higher usage of *must* (especially on Twitter) to the presence of political activists who express their opinion using this verb. The number of unique lexical verbs used with *must* is also significantly higher in PakE (143) as compared to AmE (91) and BrE (72) (Shakir 2025: 10–11). The semi-modal *need to* has the highest frequency in the Pakistani data (Shakir 2025: 9). The author also finds epistemic uses of the verbs (*have*) *got to* and *have to*, as in (36) and (37).

(35) Pakistan *must revive* PSM. It is serious and major matter. (SAOnE-PK, Comments)

(36) Excellent . . . This *has got to be* the first one . . . (SAOnE-PK, Tweets)

(37) And look at their achievements, Bangladesh (i know raw was created in '68 but there *has to be* some sort of agency before that). (SAOnE-PK, Web forums)

3.5.4 Noun Phrase Complexity

The past two decades saw the rise of research on NPs to explore syntactic complexity and its impact across registers. Early linguists like Aarts (1971) identified that 'light' (non-modified) NPs bear a specific relationship with spoken texts and subject positions, while 'heavy' (modified) NPs are associated with written texts and non-subject positions. NP complexity arises through the addition of modifiers – either pre-modifying or post-modifying the head noun (Quirk et al. 1985). Biber et al. (1999: 578) found that complex (i.e. pre- and post-modified) NPs occur more frequently in academic writing and news reports as opposed to spoken conversations.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Several studies have also been conducted on NP complexity in varieties of English. For example, Brunner (2014) comparatively analyses British, Singaporean, and Kenyan NP patterns using the respective International Corpus of English (ICE) components. He attributes the differences across the three varieties to the structural uniqueness of each substrate language. Hence, Singaporean English employs straightforward pre-modification reflecting Sinitic head-final typology, while Kenyan English exhibits considerable Bantu patterns favouring more post-modification (Brunner 2014: 37). Brato's (2020: 391) diachronic analysis of Ghanian English shows that NP complexity increases over time, which is attributable to increased education and proficiency in English, instead of language contact or substrate influence. Schilk and Schaub (2016: 81) analyse ICE components of Hong Kong, Singapore, Jamaica, India, and Canada. Their findings show that syntactic function of the NP (subject vs other) is the strongest predictor of complexity, while variety and register only become relevant in interaction effects. For example, they show that Indian English academic writing shows a strong tendency towards simple NPs (Schilk & Schaub 2016: 78).

No analysis of NP complexity is available for PakE, so our aim in this subsection is to conduct a preliminary study on NP complexity in PakE in comparison to AmE and BrE. Since the NP types are highly dependent on registers, we limit ourselves to the three comparable written registers from PWE (i.e. editorials, news, and research papers) and their counterparts from Crown and CLOB. We use Schilk and Schaub's (2016: 63) definitions to identify four types of NPs:

- Simple: (Det) + Noun, for example, *(The) book, UN*.
- Pre-modified: (Det) + Adjective(s) + Noun, as in *(The) original book*.
- Post-modified: (Det) + Noun + Prepositional phrase/relative clause, as in *(The) book that you read/of Mathematics*.
- pre- and post-modified: (Det) + Adjective (s) + Noun + Prepositional phrase/relative clause, as in (*The) old book that you sent me/of Mathematics*.

NPs with complex pre-modification (e.g. *good A and B*, *good A and bad B*) were also included in the last type, as proposed by Schilk and Schaub (2016: 63–64). The NPs were extracted from the tagged (POS and NER) and parsed (constituency and dependency) data using *stanza* (Qi et al. 2020) and the *SuPar* parser (Zhang, Li, & Min 2020; Zhang, Zhou, & Li 2020). The latter tool was selected because it performed slightly better than the built-in parser of *stanza* in constituency parsing. We used custom scripts written in Python to extract NPs from subject and other positions (e.g. object). NPs in stand-alone prepositional phrases were ignored. After extracting the NPs, we selected 100 random

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

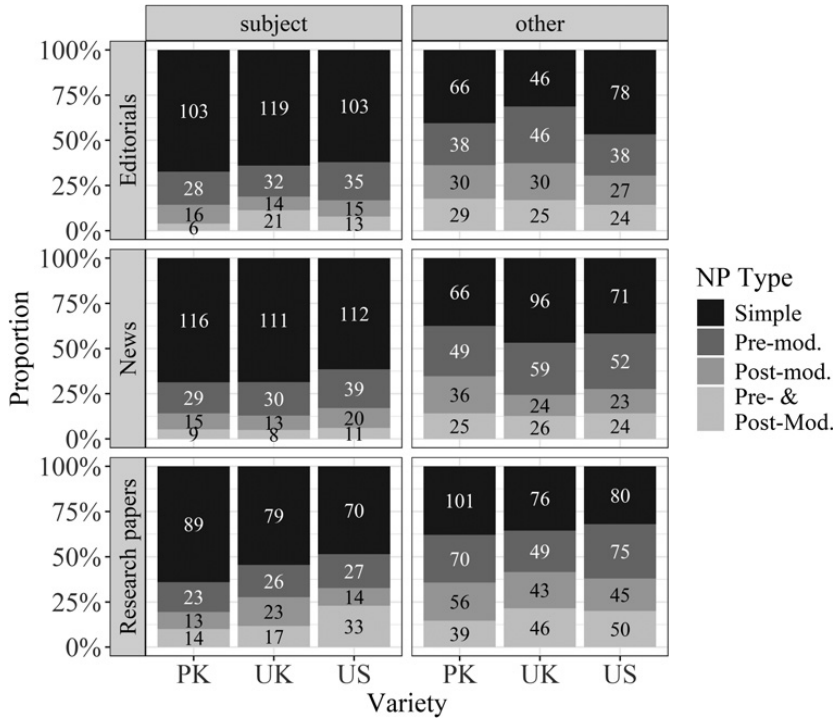


Figure 10 NP complexity according to syntactic function in the three varieties and selected registers

sentences (and the corresponding NPs) from each register category and each corpus. This generated about 420 NPs from each category and corpus, making a total of about 3700 NPs. These were manually checked for accuracy and false positives were discarded. The final dataset includes 3204 NPs.

The results are plotted with respect to the syntactic function (subject vs other) in the three varieties and three registers in Figure 10. The proportions show that there is hardly any difference in the use of NP types among the varieties under consideration. As shown in previous research, syntactic function shows the most visible differences in the use of NP types, where simple NP types are more common in subject position, while complex pre- and post-modified types are more frequent in all non-subject (i.e. other) positions. Nonetheless, two slight variations in PakE registers can be noted. Firstly, Pakistani research papers employ simple unmodified NPs more often when compared with both AmE and BrE, as depicted in the bottom panel 'Research Papers' under the panel titled 'subject'. An example of such an NP is presented in (38). This finding is in line with Schilk and Schaub's observation about Indian English, another SA L2 variety of English. Secondly, Pakistani news reports show a higher proportion

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

of post-modified NPs in non-subject (i.e. other) positions, which is visible in the panel 'News' under the vertical panel 'other'. As per our random sample, the post-modifiers in these NPs are mainly prepositional phrases, as shown in (39). In our mixed effects multinomial regression model (np_type ~ variety * register_category * np_position), the variables 'register category' and 'NP position' (subject vs other) are statistically significant predictors of NP type, but variety is not significant.

(38) *The teacher* acts as a facilitator of their learning. (PWE, Research papers, Syntactic function: *subject*)

(39) Speaking on the occasion, Tahir Iqbal said that [. . .] these are essential for [. . .] disseminating *the message of peace and fraternity.* (PWE, News, Syntactic function: *other*)

3.6 Section Summary

In this section, we have presented features of PakE on the phonological, orthographic, lexical, and grammatical levels. The phonological features (e.g. de-aspiration, retroflexion) are the result of substrate influence in the educated speech of Pakistanis and give them a distinct accent. The analysis of spelling variation has shown that Pakistani users of English follow BrE norms in formal contexts, but AmE is not far behind (especially in informal CMC registers and among young Pakistanis). The analysis of lexical features has demonstrated that Pakistanis use existing English resources productively (e.g. adjectives with *-ian*) as well as create new expressions by using borrowed vocabulary (e.g. local nouns and adjectives) and linguistic processes (i.e. reduplication). The four case studies of grammatical variation depict that PakE is relatively conservative compared to AmE and BrE (e.g. the use of *which* relatives, *be* passive, and the modal *must*). Additionally, it also shows features found in other new English varieties (e.g. reduplication patterns shared with other SA Englishes, the use of simple NPs in academic writing).

4 PakE on the Internet

The attitudes survey in Section 2 has indicated that the use of English is more likely in the situations associated with the internet, that is, social media (personal or professional) and email. Email communication is mostly relevant to professional settings (e.g. between different departments of a university). Moreover, it is a private register, so the data collection is not as straightforward as it is for public registers like blogs and social media. In the present section, we explore the use of English by Pakistani netizens in various public registers. We

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

start with a brief overview of the earliest use cases in various kinds of blogs. Afterwards, we present individual examples of English usage in various settings on Facebook. Lastly, we conduct two case studies based on larger datasets, showcasing the use of English on Instagram and Twitter.

4.1 PakE Blogs and Bloggers

This overview of PakE blogs is based on previous research (Shakir 2020; Shakir & Deuber 2019) and some more recent observations. The earliest English blogposts available in Shakir's (2020) dataset of PakE blogs are from 2004 (e.g. the post from Pakhana which can be found at https://pakhaana.blogspot.com/2004/01/flirting-with-disaster.html). As noted by Shakir and Deuber (2019: 13), many of these blogs are of the personal diary type (Grieve et al. 2010) where the bloggers talk about their day-to-day lives. By the mid-2010s most of these blogs were abandoned, while at the same time social media platforms like Facebook and Twitter were gaining popularity among these users. A look at the contemporary multi-writer blogs, for example, metro blogs (https://web.archive.org/web/20051224175750/http:/lahore.metblogs.com/), shows that these bloggers were most probably from urban centres of Pakistan (Karachi, Lahore, and Islamabad) and belonged to the millennial generation, who would have been in their teens or early twenties at that time.

In late 2000s and 2010s, individual and multi-writer blogs gave way to various thematic types (Grieve et al. 2010) of English blogs, including news blogs on English newspaper and news channel websites as well as stand-alone blog sites on topics like technology, internet gossip, celebrity news, and so forth. At the time of writing, the thematic blogs in the form of news and opinion blogs are still active on news websites (e.g. the Express Tribune found at https://tribune.com.pk/blogs) or other platforms (e.g. PakPositive found at https://www.pakpositive.com/opinion/). Technology blogs, which were written in the informal and personal style of blogs in the beginning, have now become full-fledged news media outlets, focusing not only on technology but many more topics – see Shakir (2020: 160–164) on the diachronic development of literate features in Pakistani technology blog ProPakistani.com. Along with these blog types, there are online magazine-type blogs – Shakir (2020: 100–101) labels these as 'new media blogs'– that feature content about Pakistani celebrities (including their Instagram posts and stories) in English.

4.2 English on Facebook

Facebook is one of the oldest and most widely used social media platforms in Pakistan. One of the main reasons for its popularity is that a text-based version of Facebook (i.e. no images or videos are downloaded) can be used without any cost

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

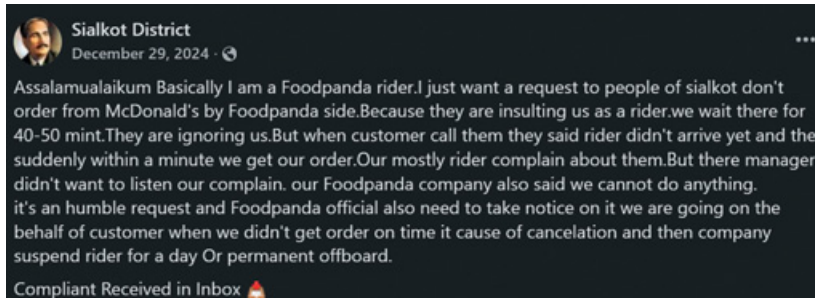


Figure 11 A complain-related post about a local branch of an international food chain

on many mobile phone networks (e.g. on Ufone found at https://web.archive.org/web/20260112171500/https:/www.ufone.com/free-facebook/). This makes the use of Facebook on smartphones very attractive in rural areas, where internet is only accessible through Edge, 3G, or in some cases 4G mobile networks. There are thousands of pages on Facebook made by Pakistani users that represent individuals, including writers or celebrities, particular cities, districts or subdistricts, businesses, including news outlets, (online) brands, local governments, government officials, public sector departments, universities, other educational institutes, and so forth. These pages post updates in the form of text and multimedia (i.e. images or videos mostly made by smartphones). Depending on the type of the page, the text accompanying these posts can be in Urdu (in majority of the cases), as well as in regional languages and English. The present subsection discusses some of these posts and their context to highlight various settings on Facebook where English is used by Pakistani Facebook users.

The text shown in Figure 11 is from a page named *Sialkot District*. This page is mostly dedicated to news updates from the district of Sialkot in the province of Punjab. The majority of the posts on this page are in Urdu, but English posts like the one shown in the figure are also possible. As the last line indicates, the post is not written by the page admin(s), but by another Facebook user and was sent as a private message to them. The writer is apparently a delivery boy and wants to complain about the behaviour of the local franchise of an international food chain. This is most probably one of the reasons why the post and most of the 800+ comments on the post are in English. The post language exhibits the usual social media conventions, for example, no space after period or no marking of sentence boundaries (third last line of the main paragraph). Additionally, it also shows some features of PakE and the local culture, for example the use of progressive *they are insulting us* and the show of respect *an humble request*. Features like absence of subject verb agreement in *when customer call* and plural marking in *Foodpanda official* indicate that it could be classified under mesolectal PakE

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

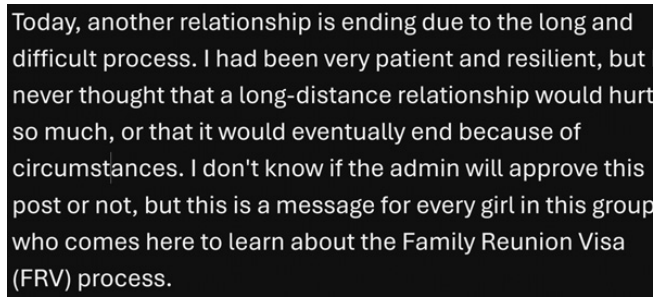


Figure 12 A paragraph from a post in a family reunion visa-related Facebook group

(variety C) used by middle/lower middle class Pakistanis educated through public sector schooling system (Rahman 2015: 22).

Apart from Facebook pages of different kinds, English is also employed in Facebook groups by Pakistani internet users. Though this phenomenon has been documented in previous research (e.g. Shakir 2020), we present some additional cases which have not been covered yet. The part of a post shown in Figure 12 is by a Pakistani woman married to a Pakistani man living and working in a European country. She is supposed to be waiting for a family reunion visa. The post tells about the unfortunate end of their marriage in divorce due to the very long waiting time required to get such a family visa. Though it is a very personal topic, the selection of English instead of Urdu shows the extent Pakistanis are able to employ this language in non-professional contexts. Another probable reason is the distancing and shielding provided by selecting a more formal language.

Overseas Pakistanis (living in English-speaking as well as other countries like Belgium, Germany, Spain, etc.) frequently use English among themselves and with their countrymen living in Pakistan when communicating through Facebook groups and other such platforms. This is another reason why the post in Figure 12 is in English. A similar non-professional context where English is employed very frequently by Pakistanis is the search for a life partner through Facebook groups created only for this purpose – possibly also applicable to WhatsApp and other such platforms. Pakistanis living within Pakistan as well as abroad are members of such *rishta* ('marriage proposal') groups. Candidate profiles are mainly posted in English, though Urdu posts also exist. The text contains information related to age, gender, education, caste, ethnicity, profession, current residency, hobbies, (dis)preferences in a potential life partner, and so forth. These texts, which are written once and only edited for minor details later, can be posted in multiple groups. Other group members can comment (mostly in English) on the posts to show their interest. Subsequent

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

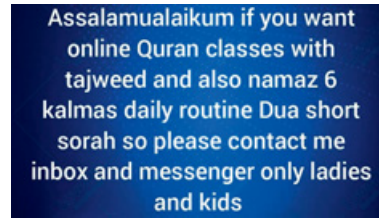


Figure 13 Business promotion post by a local Pakistani in a diaspora Pakistani group

communication through messaging services like Facebook Messenger usually occurs in Urdu or other local languages, though.

The last example of English use on Facebook is a post in a diaspora Pakistani group in which a graduate of a religious seminary in Pakistan is advertising her online teaching services (Figure 13). The post is written in plain English with no punctuation marks. It employs religious terminology: *namaz* ('prayer'), *tajweed* (rules of Quranic recitation), and *kalmas* (basic declarations of Islamic faith). It should be noted that such graduates are traditionally more proficient in Arabic and Urdu instead of English. However, the context (advertising and Pakistani diaspora in a non-English-speaking European country) has triggered the use of English instead of Urdu which would be equally understandable to the audience but may not be as effective. The use of English in social media advertisements is not limited to diaspora context, but is also common within Pakistan whether it is by bigger business establishments, individual online business owners, or Pakistani celebrities (see Section 4.3).

To summarise, in this subsection we have presented some situations on Facebook where the use of English by Pakistani netizens is more likely. These involve professional contexts (e.g. complaining about a business or advertising one's business) as well as non-professional ones like posting a marriage proposal. The interlocutors may reside within the country or live abroad.

4.3 The Use of English on Instagram by Pakistani Celebrities

The purpose of this subsection is to explore various uses of English and/or Urdu by Pakistani TV actors and celebrities in Instagram posts. The data have been extracted from post captions. We do not include the accompanying images or videos in this analysis. This is due to the time and resources that would be required to conduct multimodal analyses, and also due to the difficulties in obtaining such data, which can lead to several gigabytes of downloaded content and cause account blocking by Meta (Instagram's parent company). The original dataset is much larger with sometimes several hundred or a thousand plus

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

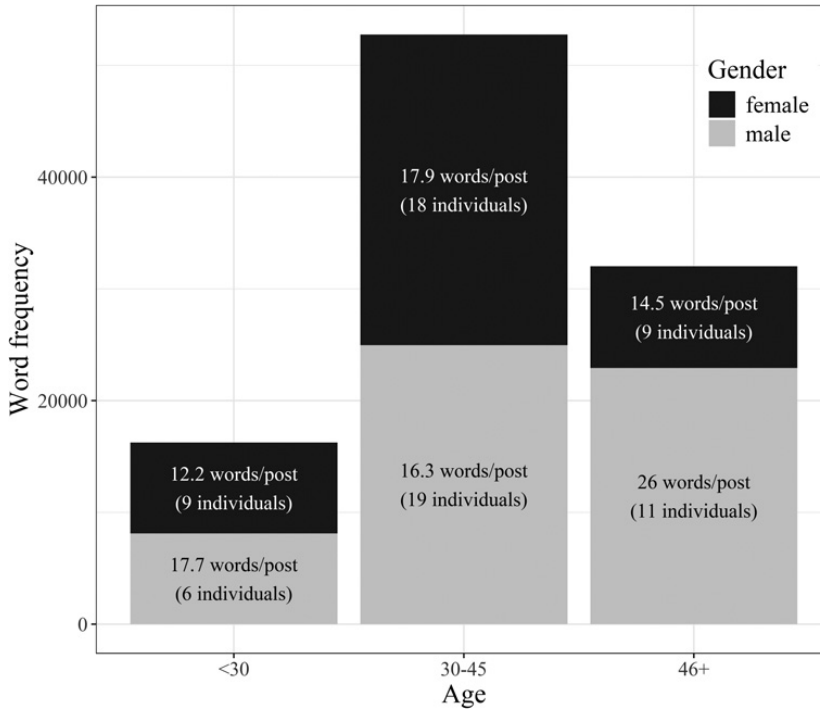


Figure 14 Demographic details, number of words, and post length in sub-categories

posts per account. The present data consists of randomly sampled posts (100 per celebrity or less if not enough posts are available) of 72 (36 each) male and female Pakistani celebrities. The accounts are mainly run by the celebrities themselves, though some are run by professional media management agencies/individuals. Figure 14 shows the age, gender, number of words, and post length of each demographic sub-category in the data.

The celebrities are mainly actors; very few of them have multiple professions listed on their Instagram profiles (e.g. director, producer, content creator, public figure). To keep the sample as consistent as possible, we have tried to exclude artists who are singers, songwriters, and so forth. Many of the celebrity accounts in these data write only 'Actor' in their bio. While there are also 'pure' Insta and Tiktok celebrities present in Pakistan, the 'traditional' celebrities are TV and film stars (many of the film stars are now TV drama actors due to almost a dead film industry). These may include people from families that have been in performing arts for more than one generation or newcomers that made their way to TV and media through talent, luck, or a mixture of both. Such celebrities used to be interviewed by film journalists for magazines in Urdu and English (or traditional

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

newspapers' Sunday magazines in both languages). That role has now been taken over by online news outlets in English (see Section 4.1) or Urdu, and social media. These news outlets also feature the Instagram posts of these celebrities as news or gossip. Hence, Instagram has become a very important source for these celebrities to connect with the traditional media as well as directly with their fan base. Many of them have Facebook and possibly also TikTok accounts as well, but we concentrate only on Instagram for this analysis. Instagram is a visual platform, in that pictures and images are first-class content here. Reels (vertical format videos) are a later innovation to compete with TikTok. Pakistani celebrities can use Instagram to post their latest snaps (especially model-cum-actors in both genders), updates about their on-air or in-production projects, post about themselves, family or work, advertise for brands, or comment on social/political issues. English is a very useful and prominent tool in their linguistic repertoire. They also use Urdu to connect with their audience and followers, but English is more formal and many times deemed appropriate for the occasion. The following subsections explore these uses of English.

4.3.1 Data Annotation

The posts have been annotated for theme and language through Google Gemini 2.5 Pro. Gemini was not provided with an explanation of the thematic categories. Instead, a 10% random stratified sample (approximately ten posts from each Instagram account) of the selected data was manually annotated for the nine categories listed as follows. Then Gemini was instructed to tag additional data (including the manually annotated posts) using the given scheme:

- Ads: Promotions of film/drama projects (of the celebrity) or of local/international brands and businesses (clothes, beauty products, housing societies, etc.).
- Follower engagement: Asking a question and/or for feedback about a TV show, or otherwise engaging with the followers.
- Hashtags: Just hashtags with no other theme deducible from the hashtags.
- Personal life: Posting about self, family, and friends, including travels or simply selfies with some text.
- Professional life: Posting about career, job, colleagues, professional relationships, drama/film projects (e.g. back the stage updates).
- Politics/social: Posting on local or international political or social issues, including the ones that may affect the actor community.
- Quotations: Motivational quotations or posts that address the audience in general, song lyrics or poetry verses in Urdu or English, or a funny quotation.
- Unclear: Posts with no text including emojis, or posts that are hard to fit in any of the other categories.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

- Wishes: On holidays, to colleagues or family members on birthdays, anniversaries, and simple prayers.

The data were also annotated for language using the same model. Code-switching was defined at phrase or clause level (two or more words) to make separation more realistic and hopefully get more accurate chunks of each language.

4.3.2 Validation of Annotated Data

The thematic annotation was completed in multiple runs to improve accuracy. Initially, the Large Language Model (LLM) was given a list of fourteen categories and brief explanations to annotate the data. A random sample of 100 posts indicated about 90% accuracy in theme and language annotation. Later on, a 10% sample (as described in Section 4.3.1) was manually tagged and compared with the LLM-based annotation. The comparison showed a 0.63 Kappa (unweighted for two raters: human vs machine). The accuracy was 69%. In the second run, the human-annotated data were used as examples for re-annotation of the complete dataset. The Kappa for the 10% sample (human vs machine) was 0.89, and the accuracy was 90%. The language annotation resulted in a Kappa of 0.46 in the first run, and 0.51 on the 10% sample in the second attempt. It was mostly mixed language posts that caused disagreement.

To understand the overall trends in the thematic categories, the 10% human-annotated sample and the overall dataset tagged by Gemini have been compared side by side in Figure 15. The bar charts show that Gemini tends to annotate social/political posts and wishes less frequently as compared to the hand-annotated data. The lack of cultural understanding required beyond the post text appears to be the main cause of this discrepancy. However, the three most frequent thematic classes (i.e. ads, personal life, and professional life) have almost similar proportions in both datasets.

4.3.3 Data Analysis

The data have been analysed for the effect of age, gender, and language on the post theme. The posts annotated by the LLM for English (74%) and Urdu (8%) were retained for the final analysis. Mixed posts (approx. 15%) and unclear cases were removed to reduce incorrectly annotated cases. Multinomial regression model from *nnet::multinom* was applied to the data, and the final model was selected using the *Mass::stepAIC* function (Venables & Ripley 2002) using R (R Core Team 2025). The model shows that all three explanatory variables interact with each other, with the interaction of language and gender being the weakest. The results have been visualised in the form of bar plots with average

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

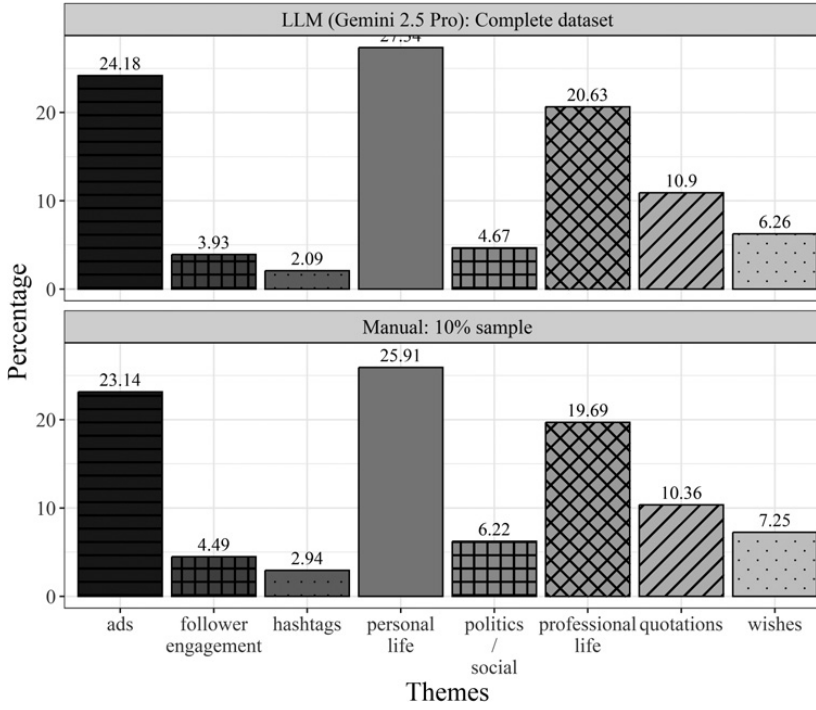


Figure 15 Percentage comparison of thematic annotation (machine vs human)

predicted probability of the eight thematic categories with respect to the selected explanatory variables. These marginal effects have been calculated using the *marginaleffects* package (Arel-Bundock, Greifer, & Heiss 2024) with the option *newdata*='*balanced*' in the function *avg_predictions* to account for unequal subgroups in age, gender, and language.

Figure 16 shows the average predicted probabilities of post theme as per age groups and gender. The likelihood of professional life-related posts increases with age in both genders, while personal life-related posts remain almost steady in females, with a decrease in older males. Example (40) shows a professional life-related post where a female celebrity is talking about an old drama she performed in. A post captioned with *my family* has been given the category of personal life, as in (41). Posting ads (own drama/file project or a brand's advertisement) decreases with age. The highest probability appears to be in younger female celebrities. Example (42) shows the advertisement of a clothing brand without clarifying that it is an ad. Example (43) is a much more clearly phrased advertisement of a brand. The promotion of own drama project in (44) is an ad, but it also shows why it is difficult to annotate for language even for a capable model like Gemini 2.5 Pro. This is especially the case when post text

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

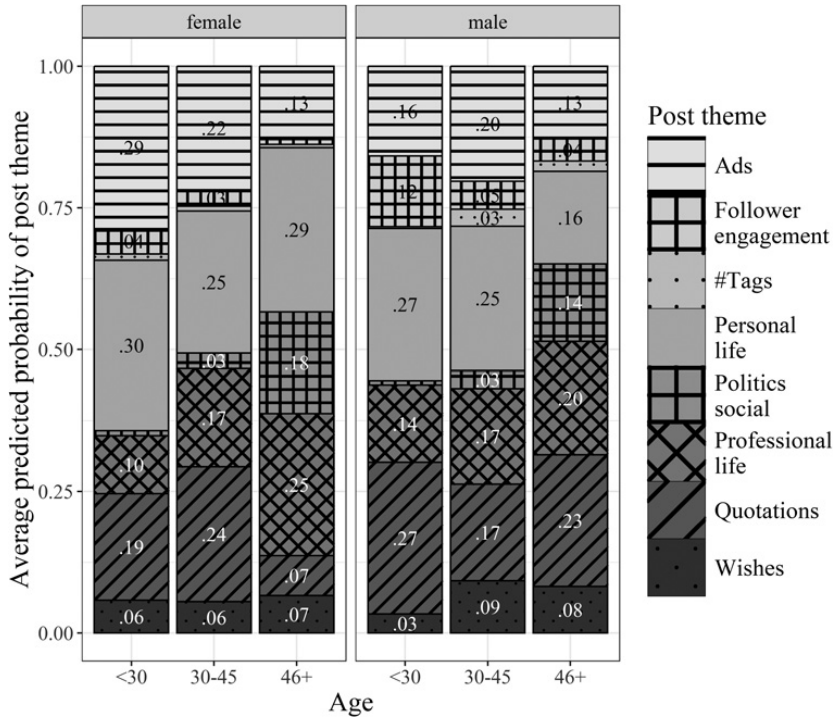


Figure 16 Average probability of post theme according to age and gender

is shorter. Lastly, posts about social/political issues are more likely in older age groups in both genders, as shown in (47) and (48).

(40) Blast from the past ❤ throwback. AROOSA days with the most amazing Fatima Surraya Bajia (May her soul rest in peace Ameen). What a personality. Am so lucky to have worked with such a stalwart. #xxxxxxxxx #xxxxxxxxxxxxxxxxx #ptv #ptvhome #pakistanicelebrities (female, 46+, professional life)

(41) My Family ❤ MashaAllah [Family, Support System, Love, Farooqui Brothers, twins, Xxxxxx Farooqui, Xxxxx Farooqui, Bleesed] (male, 30–45, personal life)

(42) Dark feminine Wearing @xxxxxxxxxxxxx❤ (female, <30, ads)

(43) Check out the amazing LUXURY LAWN COLLECTION 2022 by @xxxxxxxxxxx well known for introducing remarkable luxurious and demanding articles with high quality fabric. Shop Now: <URL> (female, <30, ads)

(44) SHAMI KI SHADI ('Shami's wedding') A tale of love, laughter, and unexpected chaos! Starting from 29th August Every Thursday at 9:00PM Only on @aurlifeofficial Staring @xxxxxxxxxxxxxxxxx @xxxxxxxx @xxxxxxxxxxxxxxxxxxx Writer: @xxxxxxxxxxxxx & @xxxxxxxxxx Director: @xxxxxxxxxxxxxxxxxxx Makeup Artist: @xxxxxxxxxxx DOP: @xxxxxxxxxx #xxxxfilmsproduction (male, 30–45, ads)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

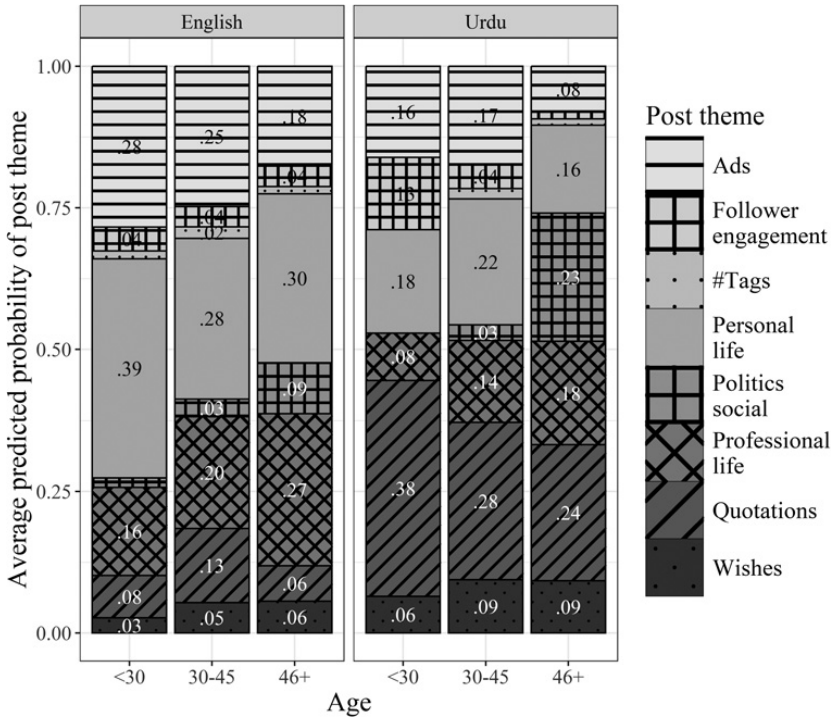


Figure 17 Average probability of post theme according to age and language

The interaction of age and language to explain the post theme has been depicted in Figure 17. Urdu posts are only a small fraction of the total dataset. Nonetheless, the data show some interesting patterns. Urdu posts have been mostly tagged as quotations, especially in younger people (less than 30), and as political/social commentary mostly in older people of both genders. Example (45) shows Hindi/Urdu song lyrics posted by a male celebrity. Follower engagement-related posts in under thirty males in Urdu are more likely. An example of such a post with a simple question is shown in (46). Two posts in the category of social/political are shown in (47) and (48) in Urdu and English, respectively. Posts related to personal wishes (e.g. a birthday) or holiday wishes (e.g. Eid) can occur both in English and Urdu, but the latter is slightly more favoured. Two such examples are shown in (49) and (50). Furthermore, it can be seen in the figure that professional life and advertisement-related posts are less likely to occur in Urdu; especially the under thirty and thirty to forty-five age groups are more likely to use English for such posts. This shows that even if the content they create is in Urdu (e.g. dramas), they prefer to talk about their professional life and promote the content in English. Lastly, post captions where personal life and family are discussed, as in (41), are also preferably written in English.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

(45) Andhere ko baahon main leke (‘Taking darkness in its arms’)
Ujaale ne ghar basaya hay (‘The light has built a home’)
Churaya tha jo chukaya hay!!! (‘It has paid back what it had stolen’) (male, <30, quotation, Hindi/Urdu Bollywood song lyrics)

(46) Kiya bolti hai public? (‘what does the public say?’) (male, <30, follower engagement, Urdu)

(47) Kash social media logoon ko Jordta . . .. (‘Wish social media would unite people’)
Kash social media kay chaudry pages chalanay kay liay mohabat pahlatay na kay nafrateen. [. . .] (‘Wish the elites of social media promoted love instead of hate.’) (male, 46+, social/political, Urdu)

(48) 2020 just keeps getting worse and worse but I cannot comprehend such a great loss. Not just for me but for all of us. Sister Ruth Lewis, your service to mankind has been and always will be one that no words can do justice to. [. . .] (female, 46+, social/political, English)

(49) Happy birthday to my very old friend @xxxxxxxxxxx 🎂
You hold a very special place in my heart. God bless you and always stay happy and healthy. [. . .] (female, 30–45, wishes, English)

(50) Chaand raat Mubarak 🤍🌙 [. . .] (‘Happy moon night’: refers to the night before Eid) (female, <30, wishes, Urdu)

To summarise this subsection, we have presented background and analysis of Instagram posts of Pakistani celebrities (mainly actors) from two genders and different age groups. The post themes and language have been automatically tagged through an LLM, which may have inaccuracies as shown by the comparisons in Section 4.3.2. However, the overall trends for thematic categories are approximately the same in the human-annotated sample and the machine-annotated complete dataset. The results show that Pakistani celebrities use English for posts belonging to such themes like ads, personal life, and professional life. Urdu and mixed language posts also occur, but they are only a small portion of their online content. This shows that Pakistani celebrities embrace English as the dominant language on the internet while communicating with their fans within Pakistan as well as with those living abroad, and possibly also with the international audience of Pakistani dramas that may reside in other SA countries (e.g. Bangladesh) or elsewhere.

4.4 The Use of English on Twitter

Twitter is also one of the most popular social media platforms among Pakistani netizens. Its wider adoption started in the late 2000s and early 2010s when blogs were being replaced with social media platforms. Around the same time, a new

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

political party (Pakistan Tehreek-e-Insaaf, PTI) started gaining popularity among the urban middle class. The party's popularity started spreading with a powerful political show in Lahore on 30 October 2011 (Web Desk 2011). Over the next few years, young Pakistanis (especially those under thirty years) increasingly started using Twitter to show their support for PTI and voice their political views. The effectiveness of social media (especially Twitter) in politics was already proven by Barack Obama in the US presidential elections of 2008 and 2012 (Pew Research Center 2012). PTI and its leader Imran Khan also employed Twitter along with other platforms like Facebook to campaign against the political elites of Pakistan. They were helped by scores of political activists posting on their behalf or voluntarily on Twitter and other social media in Urdu as well as in English. Even after Imran Khan's four-year term as a prime minister (2018–2022), Twitter has remained a symbol of resistance and political organisation among his supporters. The platform was blocked for some time in Pakistan in 2024, but in late 2025 it is again accessible. Against this backdrop, our aim for this subsection is to explore the main themes and discourses that occur in English communication of Pakistani users of Twitter.

4.4.1 Data and Methodology

The data for the present analysis are based on the SAOnE corpus (see Section 3.2). The tweets category includes one million words in 500 text files (one Twitter user/file; 100+ tweets per user) from each of the six countries. The users' profiles were manually checked to confirm their country of origin. The data were collected in 2021, so most of the tweets were posted in this year or a couple of years earlier. During the compilation phase of SAOnE and while conducting studies on parts of these data (Shakir 2024; Shakir 2025), it was observed that Pakistani tweets include political discussions. Hence, this dataset is well-suited to explore main themes and discourses in English tweets of Pakistani users.

Lexical Multidimension Analysis (LMDA) proposed by Sardinha and Fitzsimmons-Doolan (2025: 26–27) is employed for this study. LMDA is inspired by Multidimensional Analysis (MDA) propagated by Biber (1988) and many other studies who followed him. An MDA starts with the selection of functionally relevant lexico-grammatical features (e.g. adverbs and their kinds, verbs and their semantic classes, types of *that* complement clauses). Normalised per-1,000-word frequencies of these features are obtained from a representative corpus. Then a dimensionality reduction statistical technique like Exploratory Factor Analysis (EFA) or Principal Component Analysis (PCA) is applied to obtain clusters of co-occurring linguistic features. These groups of features are

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

then functionally interpreted by taking into account the data included and the situational characteristics of the register categories in the data. LMDA uses content words (i.e. nouns, verbs, adjectives, and adverbs) and excludes function words like prepositions as well as structural features like clause types. LMDA builds on the concept of keyword analysis – which produces a list of statistically significant words by comparing one corpus to another (Baker 2023: 165) – to find clusters of keywords that co-occur and may uncover underlying discourses and themes in a clearer way (Clarke, McEnery, & Brookes 2021; Sardinha & Fitzsimmons-Doolan 2025: 13–26).

For the present analysis, we identify keywords by comparing the Pakistan tweets sub-corpus from SAOnE to the tweets of the five other countries included in the corpus. The corpus analysis tool AntConc (Anthony 2024) has been used for this purpose. The data are tagged and lemmatised with the *stanza* (Qi et al. 2020) package in Python (Python Core Team 2025). As a pre-processing step, we remove # and @ from Twitter handles and hashtags, respectively. This way proper nouns can be easily identified by the tagger and later included in the keyword analysis. We use the dispersion-based keyness computation method proposed by Egbert and Biber (2019), which is available as an option in AntConc. This way, we are able to identify a list of 1735 lemmas that are used in many text files, and by extension by the Pakistani Twitter users. The normalised per-1,000-word frequencies are calculated for these features and compiled in a table. As a further step before LMDA, we remove 170+ function words based on the list from Dang and Webb (2016: 188–189) to keep only the lemmas of content words for further analysis. Additionally, we remove any lexical items with zero frequency in 80% or more documents. We also remove features which have very low correlation (<|0.1|) to any other feature, communalities of less than 0.2 as recommended by Egbert and Staples (2019), or a Kaiser–Meyer–Olkin (KMO) value of less than 0.5. Lastly, we apply signed-log transformation to the frequency table to reduce the impact of skewed frequencies (Neumann & Evert 2021: 155). For dimensionality reduction, we use PCA with no rotation – see Shakir (2024: 377–379) for a discussion on this. Figure 18 shows that about six dimensions can be extracted through both EFA and PCA.

We include four dimensions in our analysis here, as the fourth and sixth dimensions are not interpretable. The final analysis has seventy-five keywords with an overall KMO value of 0.80 ('Meritorious'), which shows the suitability of the data for EFA or PCA (Egbert & Staples 2019). The variance explained by the PCA solution is 35%. The cumulative variance explained by the first dimension is the highest, and it drops with subsequent dimensions. Only those lemma keywords are retained which have a communality of |0.295| on either dimension, as shown in Table 5. The table also shows the interpreted thematic labels for each of the four principal components.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Table 5 Dimensions and lemma keywords of lexical multidimensional analysis

Dim	+/-Label	Keywords
Dim1 +	Issues related to Pakistani politics	election (0.55), PTI (0.54), leader (0.52), Khan (0.51), Nawaz (0.51), Imran (0.5), govt (0.49), Minister (0.49), ImranKhanPTI (0.49), Pakistan (0.49), ppp (0.48), corruption (0.47), corrupt (0.47), court (0.46), political (0.45), media (0.44), minister (0.44), nation (0.44), Kashmir (0.41), law (0.4), force (0.39), country (0.39), Muslim (0.38), Sindh (0.37), Asad (0.37), politics (0.36), HamidMirPAK (0.36), Army (0.36), fawadchaudhry (0.36), Pakistani (0.36), MaryamNSharif (0.36), BBhuttoZardari (0.35), PPP (0.35), stand (0.35), Umar (0.34), PTIofficial (0.33), terrorist (0.33), innocent (0.33), citizen (0.33), foreign (0.32), imrankhanpti (0.32), US (0.32), Shah (0.31), ik (0.3), shame (0.3), action (0.3), issue (0.3), state (0.3)
Dim2 +	PPP Sindh vs	BBhuttoZardari (0.57), PPP (0.54), Sindh (0.54), Karachi (0.45), visit (0.36), ppp (0.34), regard (0.3)
Dim2-	regional issues	country (-0.3), Pakistani (-0.31), Army (-0.32), foreign (-0.34), terrorist (-0.37), force (-0.39), Kabul (-0.49), US (-0.54), Afghan (-0.62), Taliban (-0.63), Afghanistan (-0.67)
Dim3 +	Prayers and blessings	bless (0.63), Allah (0.62), prayer (0.49), stay (0.46), Ameen (0.44), peace (0.42), PK (0.41), beautiful (0.35), soul (0.33), recovery (0.32), Muslim (0.3)
Dim3 -	–	Umar (-0.38), Asad (-0.4)
Dim4 +	Citizen issues and pleadings to govt. officials	action (0.38), Sir (0.37), official (0.36), citizen (0.34), issue (0.33), city (0.31), road (0.31)
Dim4 -	–	Khan (-0.31), Nawaz (-0.32), Imran (-0.34)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

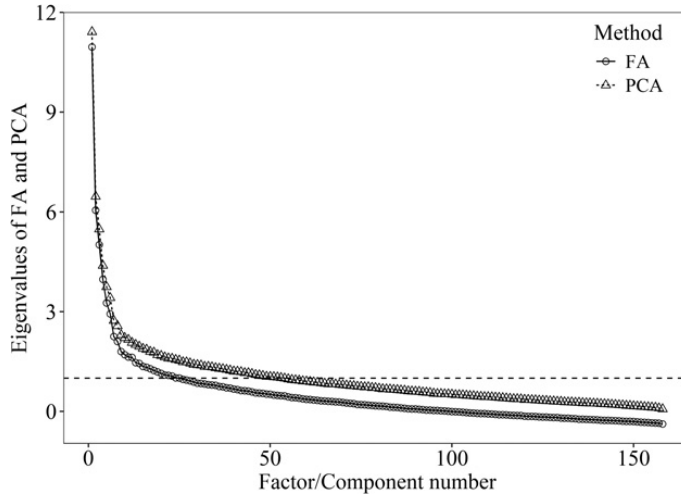


Figure 18 Scree plot of data (EFA and PCA)

4.4.2 Themes and Discourses in PakE Tweets

The first principal component consists of forty-eight keywords. It is the biggest dimension in terms of lemma keywords loaded and the variance explained in the PCA model (8% of the cumulative variance explained). The keywords can be grouped in various categories. There are proper nouns that are either names of Pakistani politicians (e.g. *Nawaz*, *Umar*, *Khan*) or their Twitter account handles (e.g. *ImranKhanPTI*, *MaryamNSharif*, *BBhuttoZardari*). Other proper nouns are related to party names (e.g. *ppp*, *PTI*), famous journalist names or Twitter handles (e.g. *HamidMirPAK*) or place names (e.g. *Kashmir*, a region disputed with neighbouring India). Some of these proper nouns may also occur as hashtags. Common nouns such as *election*, *leader, govt.*, *corruption*, *court*, *nation*, *minister*, *law, force*, *citizen*, *shame*, *action*, *issue*, *state* show that a common theme of politics with various aspects (e.g. corruption) is present in these tweets. The adjectives *political*, *terrorist*, *innocent*, *foreign* also strengthen this interpretation as they cover various internal and external political and social challenges that may be under discussion in English tweets of Pakistanis. This dimension appears to represent 'Issues related to Pakistani politics' on Pakistani Twitter.

As mentioned in Section 4.4.1, these tweets are from 2021 when Mr Imran Khan was still the prime minister of Pakistan. The Twitter accounts active at that time and included in these data are not only PTI-affiliated activists but also from other parties. In (51) and (52) most probably a sympathiser of or a social media account of another political party, Pakistani Muslim League Nawaz (PMLN), is

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

alleging the same *corruption* charges that would be levelled against Mr Nawaz Sharif (the PMLN leader) by PTI-affiliated and sympathiser accounts. In (53), the governance abilities of Pakistan People's Party (PPP), which is in power in the province Sindh, are being called into question. Examples (54) and (55) are from a PTI-affiliated account and the same discourse of *corruption* is now being wielded against PMLN and PPP leaders (Zardari) in Sindh. Along with politicians the tweets may also mention army – which can refer to Pakistan's army or another foreign army – and other security challenges like *terrorism*. On the whole, this dimension confirms previous observations that the most dominant discourse or theme on Pakistani Twitter revolves around various political issues, and a sizeable part of it is communicated in English as well.

(51) *ImranKhanPTI Imran Khan* never take any *action* Janhangir *corruption*. (PK-TWT135.txt)

(52) I condemn *Imran Khan PTI* chairman to open first offshore company with out name for the purpose to evade Tax and wealth. (PK-TWT135.txt)

(53) BBZ Why Afghan influx in *Sindh* is not checked by *Sindh govt*. (PK-TWT300.txt)

(54) ArifAlvi *ImranKhanPTI* ImranIsmailPTI *Asad Umar*: Zardaris *Sindh* Police criminally assaulting Karachites & enforcing unannounced Curfew. (PK-TWT-025.txt)

(55) You bulldozed the artificial majority of *corrupt* N- League Mafia- that's like a real roaring tiger of *PTI*. (PK-TWT025.txt)

The second dimension has eighteen keywords with seven on the positive side and eleven on the negative side. The keywords on the positive side are again from one political party, that is, PPP, and its affiliated accounts, where they tweet about various activities of the Sindh provincial government, for example, (56). On the negative side, the keywords are related to regional and international security, and political issues. Hence, there are country names (*US*, *Afghanistan*), foreign groups and political forces (*Afghan*, *Taliban*), and adjectives like *terrorist*, showing that these accounts are more interested in tweeting about international and regional affairs. This is also relevant because the US was preparing to leave Afghanistan in this time period, so this theme emerges as a concern for Pakistan. Examples (57) and (58) present example sentences from the tweets of two different Pakistani accounts. Both of them are either concerned about a Taliban-led government or the future of foreigners in Afghanistan (here most probably referring to the Americans who would soon leave the country). The themes related to terrorism are in line with other such studies on different datasets originating from Pakistan; for example, Gries (2025: 16) finds *terror* and relevant keywords based on English newspaper data collected around 2020. Combining

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

both the positive and negative sides, the second dimension can be labelled as 'PPP Sindh vs regional issues'.

(56) Spokesperson *Sindh* Govt murtazawahab1 and Minister Information Syed NasirHShah condolence with PLF G S *Sindh* Arshad Naqvi on sad demise of his father in clifton karachi. (PK-TWT392.txt)

(57) This is interesting: Though *Taliban* reportedly cancelled inauguration ceremony but Russia it will not attend any inauguration of *Taliban* govt. (PK-TWT247.txt)

(58) These *foreign* stooge have no future in *Afghanistan*. (PK-TWT208.txt)

The third dimension has eleven keywords on the positive side and two on the negative side. The positive keywords appear to be non-political (e.g. *bless*, *prayer*, *Allah*, *recovery*, *soul*) and rather religious. However, as (59) and (60) show, many such tweets are from the same politically affiliated accounts that may praise their respective leaders or wish them good health. The two negative keywords (*Asad*, *Umar*) make the name of a prominent PTI politician *Asad Umar*. The negative loadings of these keywords suggests that they are absent from the text files where the positive – seemingly less political – keywords occur. Thus, the majority of dimension 3 is related to themes where political issues are not under discussion but prayers and blessings for others (including politicians). Hence, it can be named as 'Prayers and blessings'.

(59) GetWellSoonMaryamNawaz May *Allah bless* you with a long and healthy life and [. . .] (PK-TWT265.txt)

(60) Our Shah Jee always relish our minds & hearts by posting *beautiful* roses. (PK-TWT225.txt)

The last dimension consists of seven keywords on the positive side (common nouns *action*, *official*, *citizen*, *issue*, *city*, *road* and the address form *Sir*) and three on the negative side (politician names). The keywords on the positive side along with examples (61) and (62) show that certain Pakistani Twitter users try to engage with public officials to lodge their complaints or plead for help on issues like road repair. Such accounts probably do not or less often tweet about political leaders, which is shown by the negatively loaded politician names (*Khan*, *Nawaz*, *Imran*) that are the hallmark of politically affiliated accounts. This dimension can be interpreted as 'Citizen issues and pleadings to govt. officials'. It is noteworthy that this tendency to use social media and viral posts is not limited to Twitter. In recent years, smartphone captured videos of incidents (e.g. of police brutality) have been used to get the attention of relevant authorities to seek justice. The language of such videos is mostly not English,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

but the examples here show that English in also used in this context along with other local languages in Pakistan.

(61) Thanks for sharing the information' many people will be saved from fraudulent messages' hope *action* will be initiated against such fake and fraudulent peoples. (PK-TWT459.txt)

(62) HyderabadDeputy Dear sir, great initiative by ur kind, one request to ur gdself, plz repair *road* or make paver block from *citizen* colny to bhittai nagar *road* whch z in bad condtion since three years no repair work done. (PK-TWT384.txt)

To summarise, the LMDA has shown that politics and political issues is the main theme of PakE tweets. Pakistani netizens on Twitter appear to be either active members of a political party, paid by them, or are their sympathisers. It is also noteworthy that already in 2021 other political parties had tried to catch up to PTI, the main party who successfully utilised social media to advance their narrative and gained power in 2018. Thus, we have seen the discourse of *corruption* being used both for and against PTI. The other themes discussed on Twitter are regional and security issues, and lastly some users try to connect with public sector officials and departments to get their local issues solved. This analysis reveals how, on social media, English – along with other languages like Urdu – is employed in communication about a highly local issue, that is, politics, where the interlocutors are also locals. This is despite the fact that Urdu (the national language and lingua franca) and the regional languages are the main vehicles of communication for politics in most other mediums like face-to-face conversations or electronic media.

4.5 Section summary

The section has highlighted four different domains related to the internet where English has become an important tool for Pakistanis in the twenty-first century. The first subsection briefly reviewed the use of English in different blogs, the earliest use case of English by Pakistanis on the internet. Then, in part two, the section explored the use of English in various settings on Facebook, including public pages and public/private Facebook groups where individuals may post on topics like divorce, marriage proposals, and for advertisement. It also showed how overseas Pakistanis may also motivate local Pakistanis to use English in these settings. Thirdly, we explored the use of English by Pakistani celebrities on Instagram who may use English to post about their personal and professional life, and promote their own work as well as other brands, among other uses. Lastly, we have shown that one of the unique aspects of PakE tweets is to propagate various political discourses and talk about internal/local issues, which would be discussed in Urdu and the regional languages otherwise.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

5 Conclusion

In the previous sections, we have provided a detailed descriptive account of PakE at the end of the first quarter of the twenty-first century. In the first section, we briefly described the history of the country with special attention to English policy and English teaching and learning in the country. Then, we looked at the attitudes of Pakistanis towards English in relation to Urdu and other local languages of Pakistan. In Section 3, we described various features of PakE ranging from phonology to spelling, lexis, and grammar. In the second last section, we presented an overview of the use of English in various domains on the internet. In the present section, we synthesise our findings to discuss the place of English in Pakistani society, the perceptions regarding the variety, its distinct linguistic features, and its evolutionary status within the framework of World Englishes. Finally, we put forward directions for future research.

5.1 Place of English in Pakistani Society

Previous research (e.g. Khalid 2016; Abbas & Iqbal 2018) as well as our own findings in Section 2 show that Pakistanis have a very positive attitude towards learning and using English. It is a co-official language along with Urdu and a compulsory subject from day one at schools. It is also increasingly the language of instruction in the private sector and an optional medium of instruction in public sector schools, while higher education is almost always in English. The motivation to learn English remains largely instrumental; English is viewed as a prerequisite for educational success and professional upward mobility. These positive perceptions are also fuelled by the trend (mainly increased in 2010s and 2020s) among the youth of moving abroad for study or work where English proficiency for scholarships and employment in Europe, North America, and beyond is very important.

The influence of and domains created by the internet are another catalytic factor. As per the survey results in Section 2, Pakistanis are more likely to use English in email communication. We have also shown how English is used in web-based registers (i.e. blogs) and social media for purposes like looking for marriage proposal on Facebook, posting about personal and professional life on Instagram, and talking about local politics on Twitter. The attitudes study has furthermore indicated that the younger generation and women (in comparison to men) may be more attracted to English than other demographic groups. Pakistanis, hence, appear to have integrated English in their linguistic repertoire more than at any point in the past and employ this language in an increasingly expanding set of domains along with Urdu and the regional languages.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

5.2 Perceptions about PakE

Despite the expanding domains of the language use, the perception of the local variety (i.e. 'Pakistani English') as a distinct and legitimate form of English remains elusive and contradictory to the usage patterns and localised features (see Section 5.3). Previous research from the 1990s (Baumgardner 1995) and early 2010s (Jabeen, Mahmood, & Rasheed 2011) have indicated that a majority of Pakistanis consider 'standard English' important – the 'standard' being BrE or AmE and/or both. Only limited recognition of the local variety was found. This has not changed as per our findings in Section 2. The majority of our respondents show no clear indication of variety recognition. However, they do acknowledge the presence of local speech patterns, which may be accent differences as well as lexis and grammar. Other recent research has shown that localised accents may attract negative attitudes among Pakistani students (Gul & Channa 2023). Hence, while there is a strong motivation to learn and use English, the recognition of the local variety of English is not clear at the attitudinal level.

5.3 Features of the Variety

The features of PakE described in the previous sections broadly stem from three sources: language contact with Urdu and the regional languages, AmE and BrE influences, and a combination of conservativism and possibly variety-internal innovations. We discuss these one by one.

PakE, like other postcolonial Englishes in the region and around the world, exists in a language contact situation with the local languages – Urdu to the greatest extent but also with the regional Pakistani languages. Language contact-related features are highly visible at the levels of phonology and lexis. The distinct patterns of pronunciation at the consonantal, vocalic, and suprasegmental levels are influenced by the sound systems of Urdu and the regional languages. Furthermore, sometimes speakers of different regional languages (e.g. Punjabi vs Pashto) may also show differences in their pronunciation of English words due to the influence of the respective mother tongue. The contact with Urdu also introduces a host of borrowed items which also occur in productive patterns with English nouns in adjective + noun and noun + noun combinations (e.g. *ghairat brigade*, *rishta lady)*. Moreover, we have also shown the application of reduplication – which is a very productive process in the local languages of Pakistan – in informal CMC. Thus, there are patterns literally translated from Urdu like *play X X* (e.g. *playing condition condition*) as well as reduplication of English lexical resources for emphasis (e.g. *must must*). Some of these items are shared with AmE and BrE CMC, but others are specific to South Asia or even just PakE.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Traces of competing norms and influences are also visible in the features discussed in Section 3. This competition is apparent in the selection of spelling variants. The spelling standard of the historical input variety BrE (dubbed as a 'super-central variety' by Mair 2013) are taught at schools and preferred in formal contexts. Hence, we have seen the dominance of BrE spellings in PWE (the written English corpus). However, the hyper-central variety (Mair 2013), that is, AmE, is not far behind in this regard. A substantial amount of spellings in PWE (especially the <-z-> variants) are those of AmE, and these variants become dominant if the CMC registers from SAOnE-PK are considered. The preference of AmE spellings is catalysed by the default spellchecker settings in American word processors, browsers, social media platforms, and smartphone operating systems used by Pakistanis. Also at the attitudinal level, the younger generation prefers AmE variants, considering them easier, as shown by the survey studies cited in Section 3.3. On a whole, BrE spellings are preferred for certain patterns (e.g. *flavour* vs *flavor*), while AmE variants are preferred in cases like *analyze* versus *analyse*. The influence of both native varieties is also visible at the grammatical level, for example, the use of semi-modals *have to*, (*have*) *got to*, and *need to*, as noted in Shakir (2025).

Along with the contact-induced features and competing norms of AmE and BrE, the variety also exhibits locally preferred norms that either originate from conservativism (and/or resistance to change) or possibly through local innovations. At the grammatical level, a higher percentage of *which* relatives, a higher frequency of *be* passives (at least in CMC in these data), and the frequent use of the modal verb *must* (Shakir 2025) can be listed as candidate features that have declined in the two native Englishes rapidly over the years, but are still more frequently employed in PakE. The infrequent use of *get* passives (also noted by Bohmann et al. 2025) can also be considered in this connection. At the lexical level, the use of English suffixes to create new nouns (e.g. *patwarism*, *insafian*) exemplifies local innovations. The use of the progressive ('*they are insulting us*') where AmE and BrE would prefer simple present tense could be a local innovation and/or a contact-induced feature as well. Other grammatical features can be considered local innovations or tied to the domains of use of PakE as a formal and L2 variety. The use of simple NPs in academic writing – also shared with regional varieties like Indian English – is an example of L2 features linked to the typical functions of English in Pakistan. The presence of post-modified NPs (especially prepositional phrases as modifiers) in news reports may be attributable to phrasal discourse (as opposed to clausal discourse) and a preference for more informational features, as has been observed by numerous MDA studies of the variety (e.g. Shakir & Deuber 2019; Shakir 2020; Shakir 2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

From a regional perspective, these features show that PakE has many commonalities with other SA varieties (i.e. Bangladeshi, Indian, and Sri Lankan Englishes) on different levels, but also distinct features of its own. The influence and borrowing of Hindi/Urdu shared features is one of the main reasons of the similarity with Indian English, for example, at the phonological level Hindi/Urdu sounds [ʈ] and [ɖ] and at the lexical level words like *desi* as noted in Section 3.4.3, and *yaar* ('friend') and *na* ('no') as noted by Shakir & Deuber (2023: 130–131). At the grammatical level, we have mentioned the comparatively frequent use of simple NPs in the previous paragraph. Similarly, common grammatical features have been observed in other studies, for example, the frequent use of the modal *must* in PakE and the three SA varieties (Shakir 2025). Also, at the functional level a more informational discourse (as noted by Shakir 2024) is a common feature with the other SA varieties. Examples of features specific to PakE are borrowed lexical items like *ghairat* and borrowed patterns like *play X X*. Furthermore, a relatively higher frequency of the semi-modal *need to* distinguishes PakE from the other three varieties (Shakir 2025). Thus, PakE shows similarities as well as differences from its regional peers at various levels.

5.4 Evolutionary Status of the Variety

Based on the discussion in Sections 5.1, 5.2, and 5.3, PakE shows signs of evolution related to more than one stages of Schneider's (2007) Dynamic Model. At the attitudinal level, there is little evidence to support the recognition of inward or local standardisation. The users generally look outwards to BrE and/or AmE for standards, but a considerable number of them also recognises the local speech patterns. At the same time, intra- and extra-territorial forces (Buschfeld 2020) – in this case the force of globalisation through the internet and sociodemographic developments due to moving abroad and the diaspora's contact back home through English – appear to be playing an important role too. Structural features and the expanding use cases show that the variety has an increasingly important and most probably stabilised status in Pakistani society. Language contact-related innovations at the lexical level have been documented for decades now, also using corpus-based studies like Rashid Mahmood (2009). Our case studies in Section 3.4 have shown that these features are used as functional adaptations for local expression by Pakistanis. Grammatically, the variety displays autonomy in certain aspects through its specific mixture of conservative traits (e.g. the modal verb *must* and the relative pronoun *which*) and follows the path taken by the two native varieties in other aspects (e.g. the use of semi-modal verbs like *need to*). Functionally, it is used in formal

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

domains, and MDA studies have indicated the use of informational discourse that matches, with its function as an L2 variety. Taking all of this together, at the attitudinal level, the variety could be placed in the phase of late 'exonormative stabilisation' (phase 2) or early 'nativisation' (phase 3). Some norm dependence at the feature level (e.g. the use of semi-modals) can be taken as supportive evidence in this regard. Simultaneously, the variety exhibits (strong) characteristics of 'nativisation' if conservative traits and communicative functions are considered. This view is further supported by the widespread local use of the language in traditional formal contexts (government, judiciary, higher education, etc.) and in new domains on the internet, and various features shared with other regional varieties (e.g. see Section 5.3, last paragraph). Overall, we propose that PakE is in the 'nativisation' phase.

5.5 Future Research Directions

While this Element covers various aspects of PakE, it was constrained in many aspects due to its short and introductory nature. Further research is required to complete the picture of PakE. The impressionistic description of the sound system of PakE needs to be followed up by a detailed phonetic analysis of vocalic and consonantal features using a representative speech dataset. A wider and closer look at NP complexity in additional registers also warrants further investigation. Moreover, there is a need to analyse spoken registers. Compiling an ICE-like corpus of PakE could provide a useful basis for this purpose. This study looked at contact-induced lexical features; future studies could expand this by focusing on syntactic features that may be contact-induced.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Appendices

Appendix A

Survey Questions

Use of English, Urdu, and regional languages in Pakistan

Background information

1 Age

☐ 18–25	☐ 26–35	☐ 36–45	☐ 46–55	☐ 56+

2 Gender

☐ Man	☐ Woman

3 Education

☐ Matric/O Levels/High school (10 years)	☐ Inter FA/FSc/ICs/Associate Engineer Diploma/A Levels (12 years or equivalent)	☐ BA/BSc (14 years)
☐ MA/BS Hons (16 years)	☐ M Phil/MS (18 years)	☐ PhD
☐ Other: _____		

4 Occupation

☐ Teacher (School)	☐ Teacher (College/ University)	☐ Student	☐ Other: _____

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

5 In which part of the country are you residing?

☐ Azad Jammu & Kashmir	☐ Balochistan	☐ Gilgit Baltistan
☐ Islamabad (Capital Territory)	☐ Khyber Pakhtunkhwa	☐ Punjab
☐ Sindh		

6 Did you move to your current place of residence from somewhere else in Pakistan?

☐ No	☐ Yes

7 Which language is your mother tongue?

☐ Balochi	☐ Balti	☐ Brahwi	☐ Burushaski	☐ Domaaki	☐ English	☐ Hindko
☐ Khowar	☐ Ladakhi	☐ Pashto	☐ Punjabi	☐ Purgi	☐ Saraiki	☐ Shina
☐ Sindhi	☐ Urdu	☐ Wakhi	☐ Other_____			

8 Which additional languages have you learned apart from your mother tongue?

☐ Balochi	☐ Balti	☐ Brahwi	☐ Burushaski	☐ Domaaki	☐ English	☐ Hindko
☐ Khowar	☐ Ladakhi	☐ Pashto	☐ Punjabi	☐ Purgi	☐ Saraiki	☐ Shina
☐ Sindhi	☐ Urdu	☐ Wakhi	☐ Other_____			

9 What is your proficiency in these languages?

	None	Beginner	Intermediate	Fluent	Native
English	☐	☐	☐	☐	☐
Urdu	☐	☐	☐	☐	☐
Regional language (mother tongue)	☐	☐	☐	☐	☐
Regional language (other)	☐	☐	☐	☐	☐

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

10 Please indicate how frequently you use your **(main) regional language** in the following contexts.

The regional language that is your **mother tongue** (as indicated in question 7). If no regional language is your mother tongue the regional language of Pakistan that you know best (as indicated in question 8). Use *never* if you don't know any regional language.

	Never	Sometimes	Often	Usually	Always
Talking with family	☐	☐	☐	☐	☐
Talking with friends	☐	☐	☐	☐	☐
Talking with colleagues/ class fellows	☐	☐	☐	☐	☐
Talking with strangers (e.g. shopkeeper, someone in bazaar)	☐	☐	☐	☐	☐
Writing an email (personal or professional)	☐	☐	☐	☐	☐

11 Please indicate how frequently you use **Urdu** (or **Roman Urdu** on computers and smartphones) in the following contexts.

	Never	Sometimes	Often	Usually	Always
Talking with family	☐	☐	☐	☐	☐
Talking with friends	☐	☐	☐	☐	☐
Talking with colleagues/ class fellows	☐	☐	☐	☐	☐
Talking with strangers (e.g.	☐	☐	☐	☐	☐

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

(cont.)

shopkeeper, someone in bazaar)					
Writing a personal text message/ social media post	☐	☐	☐	☐	☐
Writing a professional text message/ social media post	☐	☐	☐	☐	☐
Writing an email (personal or professional)	☐	☐	☐	☐	☐

12 Please indicate how frequently you use **English** in the following contexts.

	Never	Sometimes	Often	Usually	Always
Talking with family	☐	☐	☐	☐	☐
Talking with friends	☐	☐	☐	☐	☐
Talking with colleagues/ class fellows	☐	☐	☐	☐	☐
Talking with strangers (e.g. shopkeeper, someone in bazaar)	☐	☐	☐	☐	☐
Writing a personal text message/ social media post	☐	☐	☐	☐	☐
Writing a professional text message/ social media post	☐	☐	☐	☐	☐
Writing an email (personal or professional)	☐	☐	☐	☐	☐

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

13 Please read through the following sentences. Do you agree with the following statements?

	I strongly agree	I agree	Indifferent	I disagree	I strongly disagree
English offers advantages in seeking good job opportunities.	☐	☐	☐	☐	☐
All children should be required to learn English at school.	☐	☐	☐	☐	☐
There is no 'Pakistani English' in Pakistan.	☐	☐	☐	☐	☐
Pakistanis speak English differently than other countries.	☐	☐	☐	☐	☐
Urdu should replace English in government, judiciary, and higher education in Pakistan.	☐	☐	☐	☐	☐
I prefer using Urdu in most situations whenever possible.	☐	☐	☐	☐	☐
I prefer to talk to my kids in Urdu instead of my regional language (mother tongue).	☐	☐	☐	☐	☐
I prefer using my regional language (mother tongue) in most situations whenever possible.	☐	☐	☐	☐	☐
I strongly identify myself with my regional language (mother tongue) and the group that speaks it.	☐	☐	☐	☐	☐

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Appendix B

Table 6 ANOVA Results for Language Proficiency[1]

[1] The second column 'Anova p value' indicates whether the predictor has a statistically significant overall effect on the dependent variable (here 'language proficiency'), calculated using Type III tests via the *car* package in R. If yes, the *emmeans* package is used to calculate which subgroups within that predictor are significantly different from each other (shown in the column 'Significant posthoc pairs'). Lastly, the 'Nagelkerke Pseudo R-Squared' column represents the effect size, that is, it indicates the unique explanatory power of the current term. This is calculated by removing the current predictor from the model and calculating the difference in the pseudo R-squared values between the full and reduced models. This way, it isolates the proportion of variance uniquely explained by the current predictor variable.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Predictor	Anova p value	Significant post hoc pairs	Nagelkerke Pseudo R Squared Contribution
Gender	0.1651	N/A ($p >= 0.05$)	0.0000
Residence	0.0120	Balochistan – Punjab ($p = 0.019$)	0.0022
Education	0.0000	10y – 16y ($p < 0.01$); 10y – (18y+) ($p < 0.001$); 12 year–14 year ($p < 0.01$); 12 year–16 year ($p < 0.001$); 12 year–(18 year+) ($p < 0.001$); 14 year–(18 year+) ($p < 0.001$); 16 year–(18 year+) ($p < 0.001$)	0.0095
Occupation	0.0017	No significant pairs	0.0000
Mother_tongue	0.3259	N/A ($p >= 0.05$)	0.0000
Language	0.0000	English – Reg. (mother tongue) ($p < 0.001$); English – Regional (other) ($p < 0.001$); English – Urdu ($p < 0.001$); Reg. (mother tongue) – Regional (other) ($p < 0.001$); Reg. (mother tongue) – Urdu ($p < 0.001$); Regional (other) – Urdu ($p < 0.001$)	0.0000
Gender: Language	0.0385	For Language = Regional (other): Man – Woman ($p < 0.001$)	0.0012
Occupation: Language	0.0002	For Language = English: Other – Student ($p < 0.01$), Student – Teacher ($p = 0.022$); For Language = Reg. (mother tongue): Other – Student ($p = 0.021$); For Language = Regional (other): Other – Student ($p < 0.01$), Other – Teacher ($p < 0.01$)	0.0039

(cont.)

Predictor	Anova p value	Significant post hoc pairs	Nagelkerke Pseudo R Squared Contribution
Mother_tongue: Language	0.0000	For Language = Reg. (mother tongue): Balti – Saraiki ($p = 0.044$), Balti – Urdu ($p = 0.040$), Hindko – Saraiki ($p = 0.024$), Hindko – Urdu ($p = 0.012$), Other – Pashto ($p = 0.045$), Pashto – Saraiki ($p = 0.020$), Pashto – Urdu ($p < 0.01$), Saraiki – Sindhi ($p = 0.044$), Sindhi – Urdu ($p = 0.019$); For Language = Regional (other): Balochi – Hindko ($p < 0.01$), Balochi – Pahari/Poth ($p < 0.001$), Balochi – Saraiki ($p = 0.018$), Balochi – Shina ($p < 0.01$), Balti – Pahari/Poth ($p < 0.01$), Balti – Shina ($p < 0.01$), Hindko – Other ($p = 0.023$), Hindko – Pashto ($p < 0.01$), Other – Pahari/Poth ($p < 0.001$), Other – Shina ($p = 0.010$), Pahari/Poth – Pashto ($p < 0.001$), Pahari/Poth – Punjabi ($p = 0.042$), Pahari/Poth – Sindhi ($p = 0.019$), Pahari/Poth – Urdu ($p = 0.012$), Pashto – Saraiki ($p = 0.015$), Pashto – Shina ($p < 0.01$); For Language = Urdu: Balochi – Hindko ($p = 0.024$), Balochi – Pahari/Poth ($p < 0.001$), Balochi – Punjabi ($p < 0.001$), Balochi – Urdu ($p < 0.001$), Other – Pahari/Poth ($p = 0.019$), Other – Urdu ($p = 0.014$), Pahari/Poth – Pashto ($p < 0.01$), Pahari/Poth– Sindhi ($p < 0.001$), Pashto – Punjabi ($p = 0.021$), Pashto – Urdu ($p < 0.001$), Punjabi – Sindhi ($p < 0.001$), Sindhi – Urdu ($p < 0.001$)	0.0267

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Appendix C

Table 7 ANOVA Results for Situational Use (Mother Tongue)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Predictor	ANOVA p value	Significant post hoc pairs	Nagelkerke Pseudo R-Squared Contribution
Gender	0.4681	N/A ($p >= 0.05$)	0.0000
Residence	0.0000	AJK – KPK ($p < 0.001$); Balochistan – KPK ($p < 0.001$); KPK – Punjab ($p < 0.01$); KPK – Sindh ($p < 0.001$)	0.0000
Education	0.0009	12 year–(18 year+) ($p < 0.001$); 14 year–(18 year+) ($p < 0.001$); 16 year–(18 year +) ($p < 0.01$)	0.0021
Occupation	0.0060	Other – Student ($p < 0.01$)	0.0011
Mother_tongue	0.0050	Balochi – Sindhi ($p < 0.01$); Balochi – Urdu ($p < 0.001$); Hindko – Sindhi ($p = 0.045$); Hindko – Urdu ($p < 0.001$); Other – Sindhi ($p < 0.01$); Other – Urdu ($p < 0.001$); Pahari/Poth – Sindhi ($p < 0.001$); Pahari/Poth – Urdu ($p < 0.001$); Pashto – Punjabi ($p < 0.01$); Pashto – Saraiki ($p = 0.035$); Punjabi – Sindhi ($p < 0.001$); Punjabi – Urdu ($p < 0.001$); Saraiki – Sindhi ($p < 0.001$); Saraiki – Urdu ($p < 0.001$)	0.0000
Situation	0.0000	Family – Friends ($p < 0.001$); Family – Collegeues_Cl.Fellows ($p < 0.001$); Family – Strangers ($p < 0.001$); Family – Email ($p < 0.001$); Friends – Collegeues_Cl.Fellows ($p < 0.001$); Friends – Strangers ($p < 0.001$); Friends – Email ($p < 0.001$); Collegeues_Cl.Fellows – Strangers ($p < 0.001$); Collegeues_Cl.Fellows – Email ($p < 0.001$); Strangers – Email ($p < 0.001$)	0.0000
Gender: Situation	0.0004	For Situation = Friends: Man – Woman ($p < 0.01$); For Situation = Collegeues_Cl.Fellows: Man – Woman ($p < 0.001$); For Situation = Strangers: Man – Woman ($p < 0.001$); For Situation = Email: Man – Woman ($p < 0.001$)	0.0023

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Residence: Situation	0.0002	For Situation = Family: AJK – Balochistan ($p = 0.028$), AJK – KPK ($p < 0.01$), Balochistan – Sindh ($p < 0.01$), KPK – Sindh ($p < 0.001$); For Situation = Collegeues_Cl.Fellows: AJK – KPK ($p < 0.01$), Balochistan – KPK ($p < 0.001$), KPK – Punjab ($p = 0.011$), KPK – Sindh ($p < 0.001$); For Situation = Strangers: Balochistan – KPK ($p = 0.033$), KPK – Punjab ($p = 0.035$); For Situation = Email: Punjab – Sindh ($p < 0.01$)	0.005
Mother_tongue: Situation	0.0000	For Situation = Family: Punjabi – Sindhi ($p = 0.036$), Saraiki – Sindhi ($p = 0.041$); For Situation = Friends: Balochi – Punjabi ($p < 0.01$), Balochi – Saraiki ($p < 0.001$), Punjabi – Urdu ($p = 0.034$), Saraiki – Sindhi ($p = 0.040$), Saraiki – Urdu ($p < 0.01$); For Situation = Collegeues_Cl.Fellows: Balochi – Sindhi ($p < 0.01$), Balochi – Urdu ($p < 0.001$), Hindko – Urdu ($p < 0.01$), Other – Urdu ($p = 0.014$), Pahari/Poth – Urdu ($p = 0.011$), Punjabi – Sindhi ($p < 0.01$), Punjabi – Urdu ($p < 0.001$), Saraiki – Sindhi ($p < 0.01$), Saraiki – Urdu ($p < 0.001$); For Situation = Strangers: Punjabi – Urdu ($p < 0.001$), Saraiki – Urdu ($p = 0.012$); For Situation = Email: Balochi – Pashto ($p < 0.01$), Balochi – Saraiki ($p = 0.017$), Balochi – Sindhi ($p < 0.01$), Balochi – Urdu ($p < 0.01$), Pahari/Poth – Pashto ($p = 0.018$), Pahari/Poth – Saraiki ($p = 0.025$), Pahari/Poth – Sindhi ($p < 0.01$), Pahari/Poth – Urdu ($p < 0.01$), Punjabi – Urdu ($p < 0.01$)	0.0106

Appendix D

Table 8 ANOVA Results for Situational Use (Urdu, English)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Predictor	ANOVA p value	Significant post hoc pairs	Nagelkerke Pseudo R-Squared Contribution
Age	0.0016	(18–25) – (36–45) ($p = 0.021$)	0.0000
Gender	0.7705	N/A ($p >= 0.05$)	0.0000
Residence	0.0122	No significant pairs	0.0000
Education	0.0000	12 year–14 year ($p < 0.01$); 12 year–16 year ($p < 0.001$); 12 year–(18 year+) ($p < 0.001$)	0.0000
Occupation	0.0014	No significant pairs	0.0000
Mother_tongue	0.0000	Balochi – Balti ($p < 0.01$); Balochi – Hindko ($p < 0.001$); Balochi – Other ($p < 0.001$); Balochi – Pahari/Poth ($p < 0.001$); Balochi – Pashto ($p < 0.001$); Balochi – Punjabi ($p < 0.001$); Balochi – Saraiki ($p < 0.001$); Balochi – Shina ($p < 0.001$); Balochi – Sindhi ($p < 0.001$); Balochi – Urdu ($p < 0.001$); Hindko – Pahari/Poth ($p < 0.01$); Hindko – Urdu ($p < 0.001$); Other – Pashto ($p < 0.01$); Other – Sindhi ($p < 0.001$); Pahari/Poth – Pashto ($p < 0.01$); Pahari/Poth – Sindhi ($p < 0.001$); Pashto – Urdu ($p < 0.001$); Punjabi – Sindhi ($p < 0.001$); Punjabi – Urdu ($p = 0.021$); Saraiki – Sindhi ($p < 0.001$); Sindhi – Urdu ($p < 0.001$)	0.0000
Language	0.1166	N/A ($p >= 0.05$)	0.0000
Situation	0.0000	Family – Friends ($p = 0.029$); Family – Collegeues_Cl.Fellows ($p < 0.001$); Family – Social_media_personal ($p < 0.001$); Family – Social_media_prof ($p < 0.001$); Family – Email ($p < 0.01$); Friends – Strangers ($p < 0.001$); Friends – Social_media_personal ($p < 0.001$); Friends – Social_media_prof ($p = 0.014$); Collegeues_Cl.Fellows –	0.0000

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

(cont.)

Predictor	ANOVA p value	Significant post hoc pairs	Nagelkerke Pseudo R-Squared Contribution
		Strangers ($p < 0.001$); Strangers – Social_media_personal ($p < 0.001$); Strangers – Social_media_prof ($p < 0.001$); Strangers – Email ($p < 0.001$)	
Age:Language	0.0796	N/A ($p >= 0.05$)	0.0000
Gender:Language	0.4870	N/A ($p >= 0.05$)	0.0000
Residence: Language	0.5751	N/A ($p >= 0.05$)	0.0000
Education: Language	0.0023	For Language = English: 12 year–14 year ($p < 0.001$), 12 year–16 year ($p < 0.01$), 12 year–(18 year+) ($p < 0.001$), 16 year–(18 year+) ($p < 0.01$); For Language = Urdu: 12 year–16 year ($p = 0.011$)	0.0008
Occupation: Language	0.0085	For Language = English: Other – Student ($p = 0.040$), Other – Teacher ($p < 0.001$)	0.0004
Mother_tongue: Language	0.0793	N/A ($p >= 0.05$)	0.0000
Age:Situation	0.0000	For Situation = Family: (18–25) – (36–45) ($p = 0.033$); For Situation = Friends: (18–25) – (36–45) ($p = 0.026$)	0.0000
Gender:Situation	0.0261	For Situation = Friends: Man – Woman ($p < 0.01$); For Situation = Collegeues_Cl.Fellows: Man – Woman ($p = 0.014$); For Situation = Email: Man – Woman ($p < 0.01$)	0.0000

Residence: Situation	0.0000	For Situation = Family: AJK – Balochistan ($p = 0.027$), AJK – GB ($p < 0.01$), Balochistan – Punjab ($p < 0.01$), GB – Punjab ($p < 0.01$), KPK – Punjab ($p = 0.046$); For Situation = Collegeues_Cl.Fellows: Balochistan – KPK ($p = 0.043$); For Situation = Social_media_prof: Balochistan – Punjab ($p = 0.040$), GB – Punjab ($p = 0.025$)	0.0000
Mother_tongue: Situation	0.0000	For Situation = Family: Balochi – Balti ($p < 0.01$), Balochi – Hindko ($p = 0.026$), Balochi – Other ($p < 0.001$), Balochi – Pahari/Poth ($p < 0.001$), Balochi – Pashto ($p < 0.001$), Balochi – Punjabi ($p < 0.001$), Balochi – Saraiki ($p < 0.001$), Balochi – Shina ($p < 0.001$), Balochi – Sindhi ($p < 0.001$), Balochi – Urdu ($p < 0.001$), Balti – Shina ($p = 0.037$), Hindko – Shina ($p = 0.031$), Hindko – Urdu ($p < 0.001$), Pashto – Urdu ($p < 0.001$), Punjabi – Urdu ($p < 0.001$), Sindhi – Urdu ($p < 0.001$); For Situation = Friends: Balochi – Balti ($p < 0.01$), Balochi – Hindko ($p < 0.001$), Balochi – Other ($p < 0.001$), Balochi – Pahari/Poth ($p < 0.001$), Balochi – Pashto ($p < 0.001$), Balochi – Punjabi ($p < 0.001$), Balochi – Saraiki ($p < 0.001$), Balochi – Shina ($p < 0.01$), Balochi – Sindhi ($p < 0.001$), Balochi – Urdu ($p < 0.001$), Hindko – Urdu ($p < 0.01$), Sindhi – Urdu ($p < 0.01$); For Situation = Collegeues_Cl. Fellows: Balochi – Urdu ($p < 0.01$), Sindhi – Urdu ($p < 0.01$); For Situation = Strangers: Balochi – Hindko ($p < 0.01$), Balochi – Other ($p < 0.001$), Balochi – Pahari/Poth ($p < 0.01$), Balochi – Pashto ($p < 0.001$), Balochi – Punjabi ($p < 0.001$), Balochi – Saraiki ($p < 0.001$), Balochi – Shina ($p = 0.048$), Balochi – Sindhi ($p = 0.042$), Balochi –	0.0000

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

(cont.)

Predictor	ANOVA p value	Significant post hoc pairs	Nagelkerke Pseudo R-Squared Contribution
		Urdu ($p < 0.001$), Hindko – Saraiki ($p = 0.044$), Other – Sindhi ($p = 0.049$), Punjabi – Sindhi ($p < 0.01$), Saraiki – Sindhi ($p < 0.001$), Sindhi – Urdu ($p < 0.001$); For Situation = Social_media_prof: Pahari/ Poth – Shina ($p = 0.018$)	
Language: Situation	0.0000	For Situation = Family: English – Urdu ($p < 0.001$); For Situation = Friends: English – Urdu ($p < 0.001$); For Situation = Collegeues_Cl.Fellows: English – Urdu ($p < 0.001$); For Situation = Strangers: English – Urdu ($p < 0.001$); For Situation = Social_media_personal: English – Urdu ($p < 0.001$); For Situation = Social_media_prof: English – Urdu ($p < 0.001$); For Situation = Email: English – Urdu ($p < 0.001$)	0.0000
Age:Language: Situation	0.0000	For Language = English, Situation = Family: (18–25) – (26–35) ($p < 0.01$), (18–25) – (36–45) ($p < 0.01$); For Language = English, Situation = Email: (18–25) – (26–35) ($p < 0.01$), (18–25) – (36–45) ($p < 0.001$), (18–25) – (46–55) ($p = 0.016$); For Language = Urdu, Situation = Email: (18–25) – (26–35) ($p < 0.01$), (18–25) – (36–45) ($p < 0.01$), (18–25) – (46–55) ($p = 0.019$)	0.0042
Gender:Language: Situation	0.0000	For Language = English, Situation = Strangers: Man – Woman ($p < 0.01$); For Language = Urdu, Situation = Friends: Man – Woman ($p < 0.01$); For Language = Urdu, Situation = Collegeues_Cl.Fellows: Man – Woman ($p < 0.01$); For Language = Urdu, Situation = Social_media_prof: Man – Woman ($p < 0.01$); For Language = Urdu, Situation = Email: Man – Woman ($p < 0.001$)	0.0021

Residence: Language: Situation	0.0000	For Language = English, Situation = Social_media_prof: AJK – Sindh ($p = 0.044$), Balochistan – Punjab ($p = 0.040$), GB – Punjab ($p = 0.047$), Punjab – Sindh ($p < 0.001$); For Language = English, Situation = Email: AJK – Sindh ($p < 0.001$), Balochistan – Sindh ($p = 0.016$), Punjab – Sindh ($p < 0.001$); For Language = Urdu, Situation = Family: Balochistan – Punjab ($p = 0.024$); For Language = Urdu, Situation = Collegeues_Cl.Fellows: AJK – KPK ($p = 0.023$), Balochistan – KPK ($p < 0.001$), KPK – Punjab ($p = 0.028$); For Language = Urdu, Situation = Email: AJK – Sindh ($p < 0.001$), Balochistan – Sindh ($p < 0.001$), KPK – Sindh ($p < 0.01$), Punjab – Sindh ($p < 0.001$)	0.0043
Mother_tongue: Language: Situation	0.0000	For Language = English, Situation = Family: Balochi – Other ($p < 0.01$), Balochi – Pashto ($p = 0.020$), Balochi – Saraiki ($p < 0.01$), Balochi – Sindhi ($p < 0.01$), Balochi – Urdu ($p < 0.001$), Hindko – Urdu ($p < 0.001$); For Language = English, Situation = Friends: Balochi – Other ($p < 0.001$), Balochi – Pahari/Poth ($p < 0.01$), Balochi – Pashto ($p < 0.001$), Balochi – Punjabi ($p < 0.01$), Balochi – Saraiki ($p < 0.01$), Balochi – Sindhi ($p < 0.001$), Balochi – Urdu ($p < 0.001$); For Language = English, Situation = Collegeues_Cl.Fellows: Balochi – Other ($p = 0.017$), Balochi – Urdu ($p = 0.022$); For Language = English, Situation = Strangers: Balochi – Other ($p = 0.020$), Balochi – Saraiki ($p < 0.001$), Balochi – Sindhi	0.0071

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

(cont.)

Predictor	ANOVA p value	Significant post hoc pairs	Nagelkerke Pseudo R-Squared Contribution
		($p = 0.028$), Balochi – Urdu ($p < 0.001$), Hindko – Saraiki ($p = 0.033$), Hindko – Urdu ($p < 0.01$); For Language = English, Situation = Email: Balochi – Sindhi ($p < 0.01$), Hindko – Sindhi ($p < 0.01$), Other – Sindhi ($p < 0.01$), Pahari/Poth – Sindhi ($p = 0.015$); For Language = Urdu, Situation = Family: Balochi – Balti ($p = 0.049$), Balochi – Hindko ($p = 0.027$), Balochi – Other ($p < 0.001$), Balochi – Pahari/Poth ($p < 0.01$), Balochi – Punjabi ($p < 0.001$), Balochi – Saraiki ($p < 0.001$), Balochi – Shina ($p < 0.01$), Balochi – Sindhi ($p = 0.020$), Balochi – Urdu ($p < 0.001$), Hindko – Urdu ($p < 0.001$), Other – Pashto ($p = 0.037$), Pashto – Urdu ($p < 0.001$), Punjabi – Urdu ($p = 0.033$), Sindhi – Urdu ($p < 0.001$); For Language = Urdu, Situation = Friends: Balochi – Hindko ($p = 0.014$), Balochi – Other ($p < 0.001$), Balochi – Pahari/Poth ($p < 0.001$), Balochi – Pashto ($p < 0.01$), Balochi – Punjabi ($p < 0.001$), Balochi – Saraiki ($p < 0.001$), Balochi – Urdu ($p < 0.001$), Other – Sindhi ($p = 0.026$), Punjabi – Sindhi ($p < 0.01$), Sindhi – Urdu ($p < 0.001$);	

For Language = Urdu, Situation = Collegeues_Cl.Fellows: Punjabi – Sindhi ($p = 0.011$), Sindhi – Urdu ($p < 0.01$);
For Language = Urdu, Situation = Strangers: Balochi – Hindko ($p < 0.01$), Balochi – Other ($p < 0.001$), Balochi – Pahari/Poth ($p < 0.001$), Balochi – Pashto ($p = 0.020$), Balochi – Punjabi ($p < 0.001$), Balochi – Saraiki ($p < 0.001$), Balochi – Urdu ($p < 0.001$), Hindko – Sindhi ($p = 0.019$), Other – Sindhi ($p < 0.001$), Pahari/Poth – Sindhi ($p < 0.01$), Punjabi – Sindhi ($p < 0.001$), Saraiki – Sindhi ($p < 0.001$), Sindhi – Urdu ($p < 0.001$);
For Language = Urdu, Situation = Social_media_personal: Balochi – Hindko ($p = 0.029$), Balochi – Pahari/Poth ($p = 0.025$);
For Language = Urdu, Situation = Email: Balochi – Saraiki ($p < 0.01$), Balochi – Sindhi ($p = 0.015$), Balochi – Urdu ($p < 0.001$)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Appendix E

Table 9 ANOVA Results for Attitudes towards Various Statements

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Predictor	ANOVA p value	Significant post hoc pairs	Nagelkerke Pseudo R-Squared Contribution
Age	0.1206	N/A ($p >= 0.05$)	0.0000
Gender	0.0148	No significant pairs	0.0000
Residence	0.0000	AJK – Punjab ($p = 0.032$); AJK – Sindh ($p < 0.001$); Balochistan – KPK ($p < 0.001$); Balochistan – Sindh ($p < 0.01$); GB – Sindh ($p < 0.01$); KPK – Punjab ($p < 0.001$); KPK – Sindh ($p < 0.001$)	0.0047
Education	0.0722	N/A ($p >= 0.05$)	0.0007
Occupation	0.4722	N/A ($p >= 0.05$)	0.0000
Mother_tongue	0.0006	Balochi – Hindko ($p < 0.01$); Balochi – Other ($p < 0.01$); Balochi – Punjabi ($p < 0.001$); Balochi – Saraiki ($p < 0.001$); Balochi – Sindhi ($p < 0.001$); Balochi – Urdu ($p < 0.001$); Balti – Punjabi ($p = 0.012$); Balti – Urdu ($p = 0.013$); Hindko – Punjabi ($p < 0.01$); Hindko – Saraiki ($p = 0.048$); Hindko – Urdu ($p < 0.001$); Other – Punjabi ($p < 0.001$); Other – Saraiki ($p = 0.013$); Other – Urdu ($p < 0.001$); Pahari/Poth – Punjabi ($p < 0.001$); Pahari/Poth – Saraiki ($p < 0.01$); Pahari/Poth – Urdu ($p < 0.001$); Pashto – Punjabi ($p < 0.001$); Pashto – Saraiki ($p < 0.001$); Pashto – Urdu ($p < 0.001$); Punjabi – Sindhi ($p = 0.046$); Sindhi – Urdu ($p = 0.034$)	0.0000
Question	0.0000	English advantages – There's no PakE ($p < 0.001$); English advantages – Pakistanis speak EN differently ($p < 0.001$); English advantages – UR should replace EN ($p < 0.001$); English	0.0000

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

(cont.)

Predictor	ANOVA p value	Significant post hoc pairs	Nagelkerke Pseudo R-Squared Contribution
		advantages – Prefer UR mostly ($p < 0.001$); English advantages – (Prefer UR on MT w/ kids) ($p < 0.001$); Children should learn English – There's no PakE ($p < 0.001$); Children should learn English – Pakistanis speak EN differently ($p < 0.001$); Children should learn English – UR should replace EN ($p < 0.001$); Children should learn English – Prefer UR mostly ($p < 0.001$); Children should learn English – (Prefer UR on MT w/ kids) ($p < 0.001$); There's no PakE – Prefer MT mostly ($p < 0.001$); There's no PakE – (Identify w/ MT) ($p < 0.001$); Pakistanis speak EN differently – Prefer UR mostly ($p < 0.001$); Pakistanis speak EN differently – Prefer MT mostly ($p < 0.001$); Pakistanis speak EN differently – (Identify w/ MT) ($p < 0.001$); UR should replace EN – Prefer UR mostly ($p < 0.001$); UR should replace EN – Prefer MT mostly ($p < 0.001$); UR should replace EN – (Identify w/ MT) ($p < 0.001$); Prefer UR mostly – (Identify w/ MT) ($p = 0.031$); (Prefer UR on MT w/ kids) – Prefer MT mostly ($p < 0.001$); (Prefer UR on MT w/ kids) – (Identify w/ MT) ($p < 0.001$)	
Age:Question	0.0001	For Question = Pakistanis speak EN differently: (18–25) – (36–45) ($p < 0.001$), (18–25) – (46–55) ($p < 0.001$), (18–25) – (56+) ($p = 0.041$), (26–35) – (36–45) ($p < 0.001$), (26–35) – (46–55) ($p = 0.023$); For Question = UR should replace EN: (18–25) – (36–45) ($p = 0.044$);	0.0057

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

		For Question = Prefer UR mostly: (18–25) – (46–55) ($p = 0.013$), (18–25) – (56+) ($p = 0.027$)	
Gender:Question	0.0000	For Question = English advantages: Man – Woman ($p = 0.015$); For Question = Children should learn English: Man – Woman ($p = 0.013$); For Question = UR should replace EN: Man – Woman ($p < 0.001$); For Question = Prefer MT mostly: Man – Woman ($p < 0.01$)	0.0031
Occupation: Question	0.0042	For Question = Pakistanis speak EN differently: Other – Teacher ($p < 0.01$); For Question = UR should replace EN: Other – Teacher ($p < 0.01$); For Question = Identify w/ MT: Other – Teacher ($p = 0.012$)	0.0028
Mother_tongue: Question	0.0000	For Question = English advantages: Balochi – Other ($p = 0.015$), Balochi – Urdu ($p = 0.030$), Other – Pashto ($p = 0.015$), Pashto – Urdu ($p = 0.020$); For Question = Children should learn English: Balochi – Saraiki ($p = 0.027$), Balochi – Urdu ($p < 0.001$); For Question = Pakistanis speak EN differently: Balochi – Balti ($p < 0.001$), Balochi – Hindko ($p < 0.001$), Balochi – Other ($p < 0.001$), Balochi – Pahari/Poth ($p < 0.001$), Balochi – Pashto ($p < 0.001$), Balochi – Punjabi ($p < 0.001$), Balochi – Saraiki ($p < 0.001$), Balochi – Shina ($p < 0.001$), Balochi – Sindhi ($p < 0.001$), Balochi – Urdu ($p < 0.001$), Balti – Punjabi ($p = 0.014$), Pashto – Punjabi ($p < 0.001$), Pashto – Urdu ($p = 0.013$); For Question = UR should replace EN: Hindko – Punjabi ($p = 0.041$), Hindko – Saraiki ($p = 0.047$), Pashto – Punjabi ($p < 0.001$), Pashto – Saraiki ($p < 0.001$), Pashto – Sindhi ($p < 0.001$), Pashto – Urdu ($p < 0.001$);	0.0306

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

(cont.)

Predictor	ANOVA p value	Significant post hoc pairs	Nagelkerke Pseudo R-Squared Contribution
		For Question = Prefer UR mostly: Balochi – Balti ($p = 0.030$), Balochi – Hindko ($p < 0.001$), Balochi – Other ($p < 0.001$), Balochi – Pahari/Poth ($p = 0.038$), Balochi – Pashto ($p < 0.01$), Balochi – Punjabi ($p < 0.001$), Balochi – Saraiki ($p < 0.001$), Balochi – Shina ($p = 0.011$), Balochi – Sindhi ($p < 0.001$), Balochi – Urdu ($p < 0.001$), Balti – Punjabi ($p = 0.045$), Pashto – Punjabi ($p < 0.001$), Pashto – Urdu ($p < 0.01$), Punjabi – Sindhi ($p = 0.020$); For Question = Prefer UR on MT w/ kids: Balochi – Hindko ($p < 0.001$), Balochi – Pahari/Poth ($p = 0.031$), Balochi – Punjabi ($p < 0.001$), Balochi – Saraiki ($p < 0.001$), Balochi – Sindhi ($p < 0.01$), Balochi – Urdu ($p < 0.001$), Balti – Punjabi ($p < 0.001$), Balti – Saraiki ($p < 0.01$), Balti – Urdu ($p < 0.001$), Hindko – Punjabi ($p < 0.001$), Hindko – Urdu ($p < 0.001$), Other – Punjabi ($p < 0.001$), Other – Saraiki ($p < 0.001$), Other – Urdu ($p < 0.001$), Pahari/Poth – Punjabi ($p < 0.01$), Pahari/Poth – Urdu ($p = 0.028$), Pashto – Punjabi ($p < 0.001$), Pashto – Saraiki ($p < 0.001$), Pashto – Urdu ($p < 0.001$), Punjabi – Shina ($p < 0.001$), Punjabi – Sindhi ($p < 0.001$), Saraiki – Shina ($p = 0.013$), Saraiki – Sindhi ($p < 0.001$), Shina – Urdu ($p < 0.001$), Sindhi – Urdu ($p < 0.001$); For Question = Prefer MT mostly: Balochi – Balti ($p < 0.01$), Balochi – Other ($p = 0.038$), Balochi – Pahari/Poth ($p < 0.01$), Balochi – Shina ($p < 0.01$)	

Appendix F

Reading Passage and Diphthongs Plot

On a warm Thursday evening, a young university teacher named Zara walked through the wide, white hall towards her office. She was writing a review on the value of communication and the way people around Pakistan usually speak English. In her notebook, she wrote that many visitors wondered whether the weather in Lahore was ever this humid. While waiting for a colleague, she quietly began reading a long article about cultural change, global challenges, and modern technology. At the end of the page, she noticed a short question: 'Should we think about the future, or should the future think about us?' Then she laughed gently, knowing that the answer wasn't simple at all.

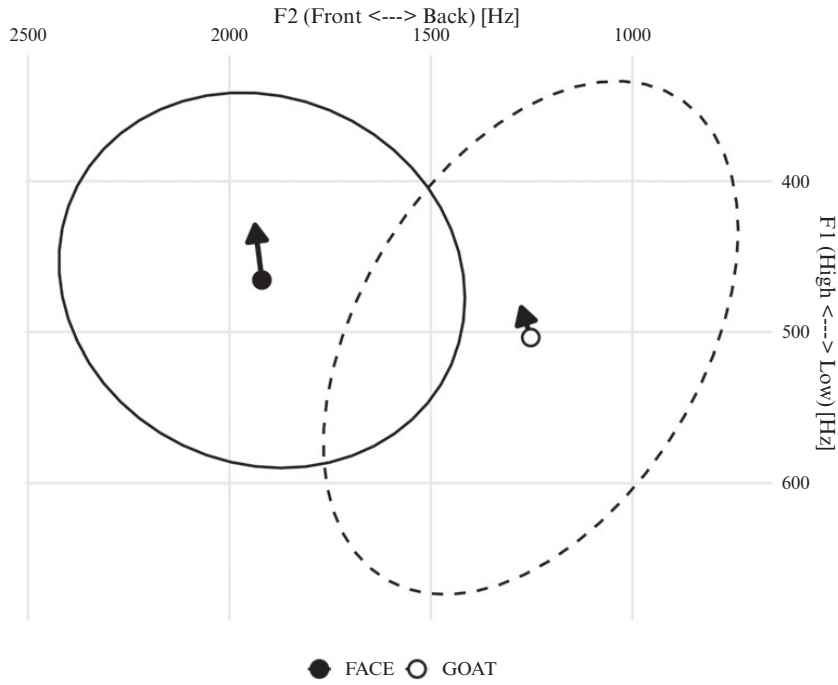


Figure 19 Vowel onset distributions and formant trajectories (20% to 80%) for FACE and GOAT vowels in the PakE YouTube data

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

References

Aarts, F. G. A. M. 1971. On the distribution of noun-phrase types in English clause-structure. *Lingua* 26(3), 281–293. https://doi.org/10.1016/0024-3841(71)90013-1.

Abbas, Furrakh & Zafar Iqbal. 2018. Language attitude of the Pakistani Youth towards English, Urdu and Punjabi: A comparative study. *Pakistan Journal of Distance & Online Learning* IV(I), 199–214.

Abbas, Shemeem. 1993. The power of English in Pakistan. *World Englishes* 12(2), 147–156. https://doi.org/10.1111/j.1467-971X.1993.tb00017.x.

Ali, Dr. Syed Shujaat, Muhammad Ishtiaq & Muhammad Shahid. 2020. Perceptions of Pakistani learners of English about standard British and American English: An exploratory approach. *sjesr* 3(2), 99–106. https://doi.org/10.36902/sjesr-vol3-iss2-2020(99-106).

Anonymous. 2002. AJK fails to adopt English as official language. *Dawn*. www.dawn.com/news/62283/ajk-fails-to-adopt-english-as-official-language.

Anonymous. 2017. *National Education Policy*. Ministry of Federal Education and Professional Training Government of Pakistan.

Anonymous. 2023. *National Census Report*. Pakistan Bureau of Statistics. www.pbs.gov.pk/wp-content/uploads/2020/07/National-Census-Report-2023.pdf.

Anonymous. 2025. *Ethnologue*. Pakistan. www.ethnologue.com/country/PK/.

Anthony, Laurence. 2024. AntConc. Tokyo, Japan: Waseda University. www.laurenceanthony.net/software/AntConc.

Arel-Bundock, Vincent, Noah Greifer & Andrew Heiss. 2024. How to interpret statistical models using marginaleffects for R and Python. *Journal of Statistical Software* 111(9), 1–32. https://doi.org/10.18637/jss.v111.i09.

Asgher, Tahira, Asma Kashif Shahzad & Anum Hanif. 2023. Unveiling Pakistani English: A linguistic and cultural Odyssey of features, evolution, and distinctiveness. *Qlantic Journal of Social Sciences and Humanities* 4(3), 299–311. https://doi.org/10.55737/qjssh.466147122.

Azam, Aram Lubna, Faria Shaheen & Mudasar Jahan. 2024. Comparing Pakistani graduates' preferences for American versus British English. *Journal of Asian Development Studies* 13(2), 1631–1646. https://doi.org/10.62345/jads.2024.13.2.128.

Baker, Paul. 2023. *Using Corpora in Discourse Analysis*. 1st ed. London: Bloomsbury. https://doi.org/10.5040/9781350083783.

Bashir, Soha & Fizza Batool. 2017. English as medium of instruction in Punjab: The 2009 experiment. *Lahore Journal of Policy Studies* 7(1), 13–45.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Baumgardner, Robert J. 1995. Pakistani English: Acceptability and the norm. *World Englishes* 14(2), 261–271.

Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman Publications Group.

Bobda, Augustin Simo. 2007. Patterns of segment sequence simplification in some African Englishes[1]. *World Englishes* 26(4), 411–423. https://doi.org/10.1111/j.1467-971X.2007.00520.x.

Boersma, Paul & David Weenink. 2026. Praat: doing phonetics by computer. https://praat.org.

Bohmann, Axel, Julia Müller, Mirka Honkanen & Miriam Neuhausen. 2025. *Linguistic Data Science and the English Passive: Modeling Diachronic Developments and Regional Variation*. London: Bloomsbury.

Brato, Thorsten. 2020. Noun phrase complexity in Ghanaian English. *World Englishes* 39(3), 377–393. https://doi.org/10.1111/weng.12479.

Brunner, Thomas. 2014. Structural nativization, typology and complexity: noun phrase structures in British, Kenyan and Singaporean English. *English Language and Linguistics* 18(1), 23–48. https://doi.org/10.1017/S1360674313000269.

Buschfeld, Sarah. 2020. Synopsis: Fine-tuning the EIF model. In Sarah Buschfeld & Alexander Kautzsch (eds.), *Modelling World Englishes: A Joint Approach to Postcolonial and Non-Postcolonial Varieties*, 397–415. Edinburgh: Edinburgh University Press. www.jstor.org/stable/10.3366/j.ctv1453j9w.23.

Clarke, Isobelle, Tony McEnery & Gavin Brookes. 2021. Multiple correspondence analysis, newspaper discourse and subregister: A case study of discourses of Islam in the British press. *Register Studies* 3(1), 144–171. https://doi.org/10.1075/rs.20024.cla.

Collins, Peter C. 1996. *Get* -passives in English. *World Englishes* 15(1), 43–56. https://doi.org/10.1111/j.1467-971X.1996.tb00091.x.

Dang, Thi Ngoc Yen & Stuart Webb. 2016. Making an essential word list for beginners. In I. S. P. Nation (ed.), *Making and Using Word Lists for Language Learning and Testing*, 153–167. Amsterdam: Benjamins. https://doi.org/10.1075/z.208.

Deuber, Dagmar, Stephanie Hackert, Eva Canan Hänsel, et al. 2022. The Norm Orientation of English in the Caribbean. *American Speech* 97(3), 265–310. https://doi.org/10.1215/00031283-8791736.

Deuber, Dagmar & Eva Canan Hänsel. 2019. The English of current Caribbean newspapers: American, British, in between or neither? In Viola Wiegand &

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Michaela Mahlberg (eds.), *Corpus Linguistics, Context and Culture*, 43–74. De Gruyter. https://doi.org/10.1515/9783110489071-003.

Egbert, Jesse & Doug Biber. 2019. Incorporating text dispersion into keyword analyses. *Corpora* 14(1), 77–104. https://doi.org/10.3366/cor.2019.0162.

Egbert, Jesse & Shelley Staples. 2019. Doing multidimensional analysis in SPSS, SAS and R. In Tony Berber-Sardinha & Marcia Veirano Pinto (eds.), *Multidimensional Analysis: Research Methods and Current Issues*, 125–144. London: Bloomsbury.

Fox, John & Sanford Weisberg. 2019. *An R Companion to Applied Regression*. 3rd ed. Thousand Oaks CA: Sage. www.john-fox.ca/Companion/.

Gries, Stefan Th. 2025. Cultural keywords in varieties research. *Journal of Research Design and Statistics in Linguistics and Communication Science*, 7, e20240001. https://doi.org/10.3138/jrds-2024-0001.

Grieve, Jack, Douglas Biber, Eric Friginal & Tatiana Nekrasova. 2010. Variation among blogs: A multi-dimensional analysis. In Alexander Mehler, Serge Sharoff & Marina Santini (eds.), *Genres on the Web: Computational Models and Empirical Studies*, 303–322. Springer.

Gul, Sanum & Liaquat Ali Channa. 2023. Dialectal preferences: A mixed methods study of ESL students' attitudes towards Englishes in Pakistan. *Asian Englishes* 25(3), 452–467. https://doi.org/10.1080/13488678.2022.2073096.

Haque, Anjum R. 1983. The position and status of English in Pakistan. *World Englishes* 2(1), 6–9.

Hinrichs, Lars, Benedikt Szmrecsanyi & Axel Bohmann. 2015. Which -hunting and the standard English relative clause. *Language* 91(4), 806–836. https://doi.org/10.1353/lan.2015.0062.

Jabeen, Farhat, Muhammad Asim Mahmood & Sobia Rasheed. 2011. An attitudinal study of Pakistani English. *Interdisciplinary Journal of Contemporary Research in Business* 3(5), 109–119.

Jadoon, Naeem Khan & Mustanir Ahmad. 2022. Pakistani English: A glimpse of phonetics and phonological features. *Pakistan Journal of Social Research* 4(1), 931–945.

Jadoul, Yannick, Bill Thompson & Bart De Boer. 2018. Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics* 71, 1–15. https://doi.org/10.1016/j.wocn.2018.07.001.

Khalid, Aamna. 2016. A study of the attitudes and motivational orientations of Pakistani learners toward the learning of English as a second language. *Sage Open* 6(3), 2158244016665887. https://doi.org/10.1177/2158244016665887.

Labade, Sachin, Claudia Lange & Sven Leuckert. 2020. English in India: Global aspirations, local identities at the grassroots. In Sarah Buschfeld &

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

AlexanderEditors Kautzsch (eds.), *Modelling World Englishes: A Joint Approach to Postcolonial and Non-Postcolonial Varieties*, 85–111. Edinburgh: Edinburgh University Press.

Le Foll, Elen & Muhammad Shakir. 2023. MFTE Python. https://github.com/mshakirDr/MFTE.

Le Foll, Elen & Muhammad Shakir. 2024. The multi-feature tagger of English (MFTE): Rationale, description and evaluation. *Research in Corpus Linguistics* 13(2), 63–93. https://doi.org/10.32714/ricl.13.02.03.

Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in Contemporary English: A Grammatical Study*. 1st ed. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511642210.

Lenth, Russell V. 2025. emmeans: Estimated Marginal Means, aka Least-Squares Means. https://doi.org/10.32614/CRAN.package.emmeans.

Mahboob, Ahmar & Nadra Huma Ahmar. 2004. Pakistani English: Phonology. In Bernd Kortmann & Edgar Werner Schneider (eds.), *A Handbook of Varieties of English*, vol. 1, 1003–1016. Berlin, New York: De Gruyter.

Mahmood, Muhammad Asim. 2009. *A Corpus Based Analysis of Pakistani English*. Multan: Bahauddin Zakariya University PhD Thesis.

Mahmood, Rashid. 2009. *A Lexico-Grammatical Study of Noun Phrase in Pakistani English*. Multan: Bahauddin Zakariya University PhD Thesis.

Mair, Christian. 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.

Mangiafico, Salvatore S. 2025. *Rcompanion: Functions to Support Extension Education Program Evaluation*. New Brunswick, New Jersey: Rutgers Cooperative Extension. https://CRAN.R-project.org/package=rcompanion/.

Memon, Natasha & Sumera Umrani. 2016. The impact of IELTS on the test preparation industry of pakistan. *International Journal of Humanities and Social Science* 6(4), 138–146.

Merejkowsky, Dimitri. 2025. pyenchant. https://pypi.org/project/pyenchant/.

Neumann, Stella & Stephanie Evert. 2021. Chapter 6. A register variation perspective on varieties of English. In Elena Seoane & Douglas Biber (eds.), *Corpus-based Approaches to Register Variation*, vol. 103, 143–178. Amsterdam: Benjamins. https://doi.org/10.1075/scl.103.06neu.

Parveen, Sabahat & Asim Mehmood. 2013. A study on attitudes towards varieties of spoken English in Pakistani context. *Language in India* 13(3), 652–667.

Pew Research Center. 2012. *How the Presidential Candidates Use the Web and Social Media*. www.pewresearch.org/journalism/2012/08/15/how-presidential-candidates-use-web-and-social-media/.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Python Core Team. 2025. *Python: A Dynamic, Open Source Programming Language*. Python Software Foundation. www.python.org/.

Qamar, Zoha & Khurram Khurshid. 2018. Machine learning based forecasting for analyzing correlation between social attention and study abroad trend for Pakistani students. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 1–6. https://doi.org/10.1109/ICOMET.2018.8346324.

Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton & Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. arXiv. https://doi.org/10.48550/ARXIV.2003.07082.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Arnold.

R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. www.R-project.org/.

Rahman, Tariq. 1991. Pakistani English: some phonological and phonetic features. *World Englishes* 10(1), 83–95.

Rahman, Tariq. 1995. The Pashto language and identity-formation in Pakistan. *Contemporary South Asia* 4(2), 151–170. https://doi.org/10.1080/09584939508719759.

Rahman, Tariq. 1999. Language, politics and power in Pakistan: The case of Sindh and Sindhi. *Ethnic Studies Report* XVII(1), 21–34.

Rahman, Tariq. 2015. *Pakistani English*. Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University.

Raj, Ali. 2017. The case for Urdu as Pakistan's official language. *Herald*, sec. People & Society. https://herald.dawn.com/news/1153737.

Rasool, Ghulam & Paula Winke. 2019. Undergraduate students' motivation to learn and attitudes towards English in multilingual Pakistan: A look at shifts in English as a world language. *System* 82, 50–62. https://doi.org/10.1016/j.system.2019.02.015.

Sardinha, Tony Berber & Shannon Fitzsimmons-Doolan. 2025. *Lexical Multidimensional Analysis: Identifying Discourses and Ideologies*. 1st ed. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009335683.

Schilk, Marco & Steffen Schaub. 2016. Noun phrase complexity across varieties of English: Focus on syntactic function and text type. *English World-Wide: A Journal of Varieties of English* 37(1), 58–85. https://doi.org/10.1075/eww.37.1.03sch.

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

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Shahid, Ayesha, Hafiz Ahmad Bilal, Hafiza Muarfa Masood Mirza & Muhammad Umar Raza. 2024. A Survey Of Phonological Characteristics Of Pakistani English. *Migration Letters* 21(S10), 957–971.

Shakir, Muhammad. 2020. *A Corpus Based Comparison of Variation in Online Registers of Pakistani English Using MD Analysis*. Münster: University of Münster PhD Thesis.

Shakir, Muhammad. 2024. An exploratory investigation of functional variation in South Asian online Englishes. *English Language and Linguistics* 28(2), 371–400. https://doi.org/10.1017/S1360674324000017.

Shakir, Muhammad. 2025. Modal verbs in South Asian Online Englishes: Exploring the use of MUST, (HAVE) GOT TO, HAVE TO and NEED TO. *World Englishes* weng.12726. https://doi.org/10.1111/weng.12726.

Shakir, Muhammad & Dagmar Deuber. 2019. A multidimensional analysis of Pakistani and U.S. English blogs and columns. *English World-Wide* 40(1), 1–23. https://doi.org/10.1075/eww.00020.sha.

Shakir, Muhammad & Dagmar Deuber. 2023. Compiling a corpus of South Asian online Englishes: A report, some reflections and a pilot study. *ICAME Journal* 47(1), 119–139. https://doi.org/10.2478/icame-2023-0007.

Shamim, Fauzia. 2008. Trends, issues and challenges in English language education in Pakistan. *Asia Pacific Journal of Education* 28(3), 235–249. https://doi.org/10.1080/02188790802267324.

Singh, Atamjit. 1997. The language divide in Punjab. *South Asian Graduate Research Journal* 4(1). https://apnaorg.com/book-chapters/language/. (1 January 2026).

Soomro, Niaz H. 2016. *Towards an Understanding of Pakistani Undergraduates' Current Attitudes towards Learning and Speaking English*. Glasgow: University of Glasgow PhD Thesis.

Syed, Nasir Abbas, Shah Bibi & Sehrish Shafi. 2024. Context-specific variance in PakE coronal stop: An investigative study of Pakistani English speakers. *Linguistics and Literature Review* 10(1), 67–84. https://doi.org/10.32350/llr.101.04.

Tottie, Gunnel. 2008. *An Introduction to American English* (The Language Library). Nachdr. Malden, MA: Blackwell.

Venables, William N. & Brian David Ripley. 2002. *Modern Applied Statistics with S*. 4th ed. New York: Springer. www.stats.ox.ac.uk/pub/MASS4/.

Web Desk. 2011. Imran Khan's 'tsunami' sweeps Lahore. *The Express Tribune*. https://tribune.com.pk/story/285058/pti-rally-in-lahore-live-updates.

Xu, Jiajin & Maocheng Liang. 2009. A tale of two C's: Comparing English varieties with Crown and CLOB (the 2009 Brown family corpora). *ICAME Journal* (37), 175–183.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Yasir, Hafiz Syed M., Ahsan Bashir, Jahanzeb Jahan, Noreen Zamir & Maryam Ahsan. 2021. Phonological shifts in Pakistani English (PakE): A comparative study under standard British English. *Elementary Education Online* 20(5), 795–806.

Zhang, Yu, Zhenghua Li & Zhang Min. 2020. Efficient second-order TreeCRF for neural dependency parsing. In Dan Jurafsky, Joyce Chai, Natalie Schluter, Joel TetreaultAssociation for Computational Linguistics. *Proceedings of ACL*, 3295–3305. www.aclweb.org/anthology/2020.acl-main.302.

Zhang, Yu, Houquan Zhou & Zhenghua Li. 2020. Fast and accurate neural CRF constituency parsing. In Christian Bessiere International Joint Conferences on Artificial Intelligence. *Proceedings of IJCAI*, 4046–4053. https://doi.org/10.24963/ijcai.2020/560.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Cambridge Elements

World Englishes

Edgar W. Schneider
University of Regensburg
Edgar W. Schneider is Professor Emeritus of English Linguistics at the University of Regensburg, Germany. His many books include *Postcolonial English* (Cambridge, 2007), *English around the World, 2e* (Cambridge, 2020) and *The Cambridge Handbook of World Englishes* (Cambridge, 2020).

Editorial Board
Alexandra D'Arcy, *University of Victoria*
Kate Burridge, *Monash University*
Paul Kerswill, *University of York*
Christian Mair, *University of Freiburg*
Christiane Meierkord, *Ruhr University*
Raj Mesthrie, *University of Cape Town*
Miriam Meyerhoff, *Victoria University of Wellington*
Daniel Schreier, *University of Zurich*
Devyani Sharma, *Queen Mary University of London*
Sali Tagliamonte, *University of Toronto*
Bertus van Rooy, *University of Amsterdam*
Lionel Wee, *National University of Singapore*

About the Series

Over the last centuries, the English language has spread all over the globe due to a multitude of factors including colonization and globalization. In investigating these phenomena, the vibrant linguistic sub-discipline of "World Englishes" has grown substantially, developing appropriate theoretical frameworks and considering applied issues. This Elements series will cover all the topics of the discipline in an accessible fashion and will be supplemented by on-line material.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308

Cambridge Elements

World Englishes

Elements in the Series

Inheritance and Innovation in the Evolution of Rural African American English
Guy Bailey, Patricia Cukor-Avila and Juan Salinas

Indian Englishes in the Twenty-First Century: Unity and Diversity in Lexicon and Morphosyntax
Sven Leuckert, Claudia Lange, Tobias Bernaisch and Asya Yurchenko

Language Ideologies and Identities on Facebook and TikTok: A Southern Caribbean Perspective
Guyanne Wilson

Transnational Korean Englishes
Sofia Rüdiger and Alex Baratta

Multiscriptal English in Transliterated Linguistic Landscapes
Chonglong Gu

World Englishes as Components of a Complex Dynamic System
Edgar W. Schneider

Optimizing Decision Trees for the Analysis of World Englishes and Sociolinguistic Data
Sarah Buschfeld and Claus Weihs

Philippine Englishes-in-Motion: The Case of Filipino Migrants in Japan
Julius C. Martinez

Contemporary Nigerian English
Foluke Olayinka Unuabonah, Rotimi Taiwo, Rotimi Olanrele Oladipupo, Folajimi Oyebola

Young Singaporean English: Current Developments in CSE Morphosyntax in a Digital World
Jakob R. E. Leimgruber, Wilkinson Daniel Wong Gonzales, Mohamed Hafiz, Jun Jie Lim, Nick Huang, Mie Hiramoto

The Life-changing Impact of English at the Grassroots in Cambodia
Evangeline Lin

Pakistani English in the Twenty-First Century: Attitudes, Features, and Uses
Muhammad Shakir, Dagmar Deuber, Muhammad Asim Mahmood and Muhammad Rashid Mahmood

A full series listing is available at: www.cambridge.org/EIWE

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 06:26:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009665308