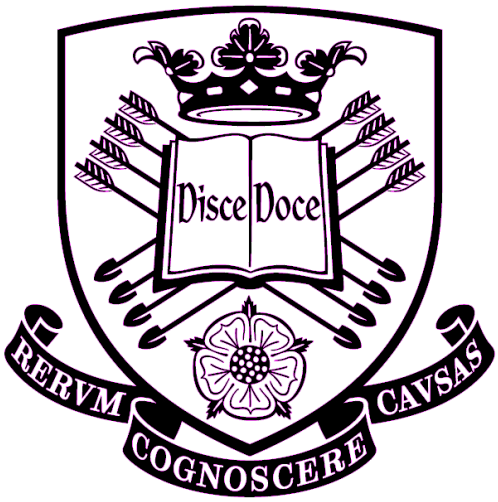
DEPARTMENT OF ENGLISH LANGUAGE AND LINGUISTICS

THE UNIVERSITY OF SHEFFIELD



From Speaking to Being English:

Exploring the acceptability of morphosyntax in Polish-born and LGBTQ+ speakers of English

Nathaniel Alexander Dziura

August 2023

Submitted for the degree of Doctor of Philosophy
Department of English Language and Linguistics
The University of Sheffield

ACKNOWLEDGEMENTS

Firstly, thank you to my supervisors, Prof Emma Moore and Dr Robyn Orfitelli. Your enduring guidance, patience, and kindness, as well as your insight and feedback throughout the project, has been invaluable. Emma, you have now completed an academic hat-trick of supervising three of my dissertations! I have learnt a lot from you both - It has meant a great deal to me and will stay with me throughout my career. I would also like to thank my examiners Prof Jennifer Smith and Dr Kook-Hee Gil, whose feedback has helped me reframe some key aspects of the thesis and further clarify my ideas.

I am incredibly grateful to all the participants who chose to take part in this research, especially the Polish-born participants who kindly offered their time - this research would not be possible without their generosity. This project was funded by the White Rose College of Arts & Humanities, to whom I am grateful for giving me the opportunity to travel, network, and share my ideas. I am very lucky to have encountered so much cool research throughout my time doing the PhD! I would like to especially thank the LavLang community for the much needed help in consolidating my ideas in the early stages, and the UCL SSEES Polish Migration group for help connecting with the Polish migrant community. I am also incredibly thankful to have had unwavering dedication and empathy from my mentors, Amy and Tim, without whom I never could have completed the final crucial stages of the writeup.

I have felt incredibly welcomed into the School of English during the PhD, by those who make it feel like a second home; Thank you especially to the Floor 5 gang (we really need a better name) - Sophie W, Beth, Irene, Starlina, Sophie H, and Alexis. Also, a special shoutout to the one and only Leesa for all your care, banter, and very much appreciated wellbeing biscuits!

I couldn't have done the PhD without the support of those around me. In particular, thank you to Bramall Lane Convenience Store (Jo, Nikki and Hielke) for everything that went down in that flat and all the #reflecting on what matters most (and thank you, Hielke, for our countless instalments of Therapy Wars!), and also to the pub crew, especially El, for your continued love and silliness. Thank you to Yi (The Best) and Sam (also pretty great) for all of your support, adventures, and shenanigans. Thanks to Holly and Theo, I really appreciate your kindness, dog walks, and the fact that you never give up on trying to educate me on proper food! Thanks to Jake for your patience, and everything you did to help me through the PhD. Last but not least, thank you to my mum and my gran for being there for me, and for being proud of what I've accomplished.

ABSTRACT

Many sociolinguistic studies of morphosyntactic variation primarily focus on highly socioindexical variants (ideologically linked to social meanings), and it is rarer to see comparison across variants of differing degrees of social salience and discourse-pragmatic utility (cf. Moore, forthcoming; Cornips & Corrigan, 2005: 87). Consequently, the present study is interested in the acceptance (ie. perception) of the following morphosyntactic constructions: **Argument movement** (raising-to-subject with the verb *seem*); ii. **Optional discourse-based movement** (left dislocation, right dislocation, topicalisation), and iii. **Nonstandard agreement** (past-tense BE: nonstandard *was*, nonstandard *were*, and nonstandard *weren't*).

Using an Acceptability Judgement Task, acceptance of these constructions is measured in Polish-born migrants - newer members to the population, compared with English-born participants. An additional dimension of comparison is by membership to a specific non-hegemonic community - the LGBTQ+ community. Factors on acceptance are considered: i. macro-social demographic factors; ii. L2 constraints on migrants' acquisition; and iii. meso-social (community embeddedness) factors.

Results from ordinal logistic regression models reveal a clear pattern in acceptability: Argument movement $>$ optional discourse-based movement $>$ nonstandard agreement. Migrants also follow this pattern, though have lower acceptance than English-born participants, with acceptance of nonstandard agreement diverging more than optional discourse-based movement. Higher English proficiency and education lower acceptance while other aspects of migrants' identities, such as age, and acculturation level (embeddedness in British culture) increase acceptance. Being LGBTQ+ increases acceptance, particularly in migrant participants. However, *level* of LGBTQ+ community embeddedness *lowers* acceptance of optional discourse-based movement.

This study challenges us to consider which speakers and communities 'matter' for variation, given the social diversity and varied lived experiences across both migrants and non-migrants. Methodologically, this study is an exemplar for accessing information about lived experiences while operating at a wide-angle 'macro' scale (without access to participants' instantiated use of language, or to participants themselves).

CONTENTS

1. Introduction

This thesis explores factors on the acceptance (ie. perception) of English morphosyntactic variation in a non-L1 (ie. a non-native language). Specifically, the research focuses on exploring sociolinguistic patterns in migrants according to their membership to, and level of embeddedness within, both the national community (ie. England), and a non-hegemonically oriented community (the LGBTQ+ community). Participants' acceptance of three types of morphosyntactic construction will be considered: Argument movement (raising-to-subject with the verb *seem*); ii. Optional discourse-based movement (left dislocation, right dislocation, topicalisation), and iii. Nonstandard agreement (past-tense BE: nonstandard *was*, nonstandard *were*, and nonstandard *weren't*). These three types represent morphosyntactic constructions that are, respectively, (i) widely considered grammatically acceptable across British-English varieties and relatively unlikely to be subject to social and stylistic constraints; (ii) less widely considered grammatically acceptable across British-English varieties and subject to social and stylistic constraints; and (iii) only considered grammatically acceptable in specific British-English varieties and subject to social and stylistic constraints. Firstly, I will provide a brief introduction to the key issues explored within this thesis, which will motivate the five research questions to be considered (see Section 1.3 for a formal statement of these research questions).

As outlined above, linguistically, the focus of the present study is at the level of morphosyntax. Much early variationist research (eg. Labov, 1964; 1966) has primarily focused on alternate phonetic realisations of phonemes, and how these pattern according to social factors. However, many variationist scholars (including, for example, Labov, 1968; Cheshire, 1972; Tagliamonte, 1998; Tagliamonte & Baayen, 2012; Snell, 2008), have since also turned their focus to analysing how sociolinguistic patterns can be measured in other levels of the linguistic system, such as in the abstract, structural elements of language (ie. relating to the grammatical system). As overviewed by Moore (forthcoming), although much of the work on morphosyntactic variation has uncovered interesting findings with regards to the social meanings of variants according to associations with particular macro-social categories - ie. ones which have strong *social type* meanings - there has been comparatively less focus on what Moore calls *referential meaning* (ie. the 'on paper' meaning), and on how variants come to accrue social meaning. Moore posits that referential and social meaning exist on a continuum, and that different types of morphosyntactic constructions accrue social meaning differently. For instance, Moore argues that the social meanings of certain morphosyntactic constructions, such as the discourse-based movement construction of right dislocation, are mitigated by their discourse-pragmatic utility (in this case, providing end-focus and emphasis), which is linked to the information structure of the construction, and this, then, provides the potential for social meanings to develop. Other morphosyntactic constructions, on the other hand, have more straightforward socioindexical links between the linguistic form and the social types they index. There is still much to be learned about how social meanings function across this proposed continuum, which motivates the consideration of a range of types of morphosyntactic variants in the present study. Using an Acceptability Judgement methodology, this study aims to investigate how speakers pattern in their acceptance of three different types of morphosyntactic construction which vary in terms of the socioindexical links (or lack thereof) associated with them across British Englishes. Relatedly, they also vary in terms of standardness, regional distribution, and social salience.

The field of sociolinguistics has made great headway in understanding how patterns of language variation correlate with social identities across varieties of speakers' L1s. However, comparatively much less work focuses on non-L1 speakers'

patterns of sociolinguistic variation, despite evidence that speakers do demonstrate an awareness of sociolinguistic rules and patterns in their non-L1s (which will be discussed later in this section). Given that community-oriented analyses can have a tendency to exclude migrant speakers (Bucholtz, 1999: 208), the present study aims to help counteract this, furthering our understanding of how non-L1 migrant speakers (ie. newer members of the population) fold into existing patterns of English morphosyntactic variation.

Given this, the present study will investigate how Polish-born migrant participants' acceptability judgements compare to those of English-born speakers. I will now outline the range of factors that will also be measured and discussed in explaining participants' patterns of acceptance towards the morphosyntactic constructions under investigation.

Formal Second Language Acquisition (SLA) studies have traditionally focused on speakers' acquisition of obligatory linguistic forms, typically in the standard form of the language in question, and have brought to light a number of factors found to be influential for learners' acquisition of a language. For instance, it has been argued that, if acquisition begins, even in a naturalistic environment, after a learner has reached puberty - ie. they are beyond the *critical period* (Lenneberg, 1967) - their acquisition will be "constrained by age-related maturational factors" (Jiang et al., 2009: 481) that younger learners are not encumbered by. Given that participants in this study are adult non-L1 speakers of English who have acquired the English language at various stages of their lives, previous research in SLA draws our attention to the importance of taking into account such factors, if we are interested in measuring the range of constraints on their acceptance of variation. Given this grounding from the field of SLA research, the present study will consider the extent to which factors relating to formal linguistic accounts (such as the degree of prior English linguistic input, and the degree of English language proficiency) can accurately capture the patterns of variation found in the Polish-born migrants' acceptability judgement ratings.

Although factors such as the degree of input and the degree of language proficiency are important to consider, sociolinguistic research has built on this by pointing to the importance of sociolinguistic constraints on variation in non-L1 contexts, modelling variation as "not a by-product of the learning process, but an integral element of overall language acquisition" (Roberts, 2005: 153–154). Sociolinguistic research (eg. Bayley & Regan, 2004) has shown that migrants are able to acquire *sociolinguistic competence* in their target language - ie. an understanding of the social and pragmatic mechanisms of variation. Given the "profound sociolinguistic consequences" (Kerswill, 2006: 1) that migration has been argued to have on individuals' variation, it is possible that a range of sociolinguistic constraints may factor into the patterns of acceptance of morphosyntactic variation in these newer arrivals, and this is not just limited to those native to the country. These include, but are not limited to: age, given that previous work has argued there to be a critical threshold for the acquisition of morphosyntactic variation in early adulthood (eg. Tagliamonte & D'Arcy, 2009: 63), and, in the migrant context, age also ties in to the age-related L2 factors discussed earlier in this section; socioeconomic status, which previous research (on L1 English speakers) has associated with adherence to local versus wider community norms and use of nonstandard morphosyntactic constructions (eg. Snell, 2008); and gender, which has been found to constrain the acquisition of native-like variation in Polish adults in Manchester (Drummond, 2010). Therefore, taking into account prior sociolinguistic accounts aiming to explain patterns of variation, this study will consider the extent to which participants' patterns in acceptance of the morphosyntactic constructions under investigation can be explained by considering these macro-social factors (ie. demographic factors such as age; gender; and socioeconomic class).

Further to the consideration of these macro-social factors as constraints on the acceptance of morphosyntactic variation, the present study aims to delve deeper into aspects of participants' lived experiences, going beyond these broad demographic identities and taking into account participants' community memberships, doing so by measuring participants' depth of embeddedness within the communities under investigation. We know that speakers orient linguistically to their sociolinguistic surroundings (eg. Labov, 1966; 1972), and these sociolinguistic surroundings are made up of the various communities and networks that speakers belong to. A key part of community membership, particularly in the case of communities which are non-hegemonic in their orientation, is the set of "culturally constructed valances" (ie. "cultural expectations, norms, and expectations") that binds community members together and informs their established behaviours (Snell, 2018: 19). Since identities are "continually constructed through interactional practices" (Snell, 2018: 13), if we are to further understand the relationship between language and identity, it is important to consider the communities to which speakers belong, and how these more 'meso-social' factors, as well as macro-social factors, influence variation. It has been pointed out that migrants have often been excluded from consideration in community-oriented analyses (Bucholtz, 1999: 208), therefore, the present study aims to investigate how community embeddedness factors into migrants' acceptance of English morphosyntactic variation. Two types of community membership (and levels of embeddedness in the respective community) are considered: membership to the broader national community and culture (namely, the English national community, though migrants' ties to the Polish community and culture are also explored), and membership to a non-hegemonically oriented community (the LGBTQ+ community).

Turning to the first of these, level of embeddedness in the broader national community will be considered via the theory of *acculturation*. It has been argued that, despite non-L1 speakers' acquisition potentially being constrained by maturational factors such as the *critical period*, we should also take into consideration "*post*-maturational factors" (Jiang et al., 2009: 481). Acquisition of a language does not occur in a vacuum, but occurs in tandem with the acquisition of the norms, behaviours, and lifestyle associated with the culture of the respective community. As such, non-L1 speakers' levels of *acculturation* - change in cultural orientation following contact with a new cultural environment (Berry, 1980) - have been demonstrated to positively affect the process of their language acquisition (Jiang et al., 2009: 481; Masgoret & Gardener, 1999) as well as their acquisition of sociolinguistic competence (Schleef et al., 2011). Relatedly, though not using the term *acculturation* specifically, results from several sociolinguistic studies indicate the importance of migrants' integrative motivations to their acquisition of variation, finding comparatively more native-like variation in migrants who are more embedded in networks where they are exposed to a higher degree of native speaker variation (Drummond, 2010: 219) as well as finding more native-like variation in migrants who are more settled (Howley, 2015). Given the importance of this post-maturational factor as a constraint on variation, as well as the utility of the *acculturation* measure in informing us about migrants' degree of embeddedness in their national communities, the present study will measure whether acceptability judgement ratings pattern differently according to participants' British or Polish acculturation levels.

Now, turning to the second meso-social factor related to community embeddedness - that of LGBTQ+ community involvement. All types of communities are sites for the production of identity (Eckert & McConnell-Ginet, 2007: 29) and, as I have discussed earlier in this section, shared sets of values, attitudes, and orientations are what hold a community together, and what allow individuals to derive a sense of identity from membership to that particular community. Particular communities may have reason to orient away from hegemonic norms, instead constructing their shared identity around

different sets of values, attitudes, and orientations that bring the communities together, compared to the wider population. We know that speakers' identities reflect in their use of language, and scholars (eg. Bucholtz, 1999: 208) have argued the importance of considering identities beyond those represented by hegemonic norms into our understanding of variation. This is something the present study aims to do, using the LGBTQ+ community as a case study for such a community. Studies (eg. Zimman, 2021; Becker & Stoddard, 2018) have found that certain members of the LGBTQ+ community, notably transgender and non-binary speakers, define the 'envelope of variation' (ie. set the limits for the upper and lower parameters of use) of the variables studied. However, despite recent uptake in sociophonetic and discourse analytical approaches to LGBTQ+ identities, there have been comparatively fewer approaches to morphosyntactic variation with a focus on LGBTQ+ identities. A notable example is Konnelly and Tagliamonte (unpublished manuscript), who compare the use of quotatives and intensifiers within an LGBTQ+ adolescent Community of Practice in Toronto with that of the ambient Toronto speech community, and find that the LGBTQ+ speakers not only lead the ambient community in use of the incoming *be like* quotative, but also participate in the adaptation of the feature into new contexts of usage, unlike the ambient community. Findings from non-linguistic studies indicate that the bonds tying one to the LGBTQ+ community can provide resilience, strength, and mobilisation in resistance to anti-LGBTQ+ discrimination (eg. Ross, 2012). Furthermore, the LGBTQ+ community has also been theorised as having an 'imagined centre' (Winer, 2020), with individuals relatively more or less connected to the core of this community. Therefore, in taking membership to this non-hegemonically oriented community into consideration, the present study will explore how the patterns of acceptance of the morphosyntactic constructions under investigation vary according to participants' membership to, and level of embeddedness within, the LGBTQ+ community, compared with the overall Polish-born and English-born populations.

Migrants, as members of their target communities, are no exception to the effects of the factors discussed in this section. Often, the various factors relating to identities, community ties, and other lived experiences - which have been shown in L1 speakers to influence or constrain the use of language variation - are overlooked in studies on L2 migrants' acquisition of language, or are not considered in depth, in favour of focusing on language through the lens of speakers' proficiency. While these approaches have their merits, it is known that extralinguistic factors play a complex and crucial role on the patterns of variation that we find. Therefore, it is important to take into consideration migrant identities and the communities within which they are embedded when studying their language variation. Given the broad range of factors that have been implicated in previous research on migrants' sociolinguistic variation, the present study will explore migrants' patterns of acceptance of morphosyntactic constructions across several dimensions, as has been discussed throughout this section: from the perspective of the macro-social (demographic) level; in terms of L2 constraints on acquisition; and from the perspective of the meso-social (community-embeddedness) level. In doing so, the relative effects of these factors can be clustered together and compared, to build up a richer picture of the complex interplay of factors on migrants' acceptance of morphosyntactic variation.

The next section of this chapter will consist of an overview of the field of variationist sociolinguistics, which chronicles the work that has been carried out, and on which the present study builds. In doing so, this section will introduce micro-, meso-, and macro- approaches to variationist sociolinguistics, and argue for the 'top-down' approach taken in the present study, whereby macro-social correlations are considered in tandem with relatively more meso-social factors regarding participants' community embeddedness and lived experiences. I then briefly introduce the participants that will be the focus of

this study - Polish-born migrants, both LGBTQ+ and non-LGBTQ+. After this, I state the research questions for the present study, and outline the structure of this thesis.

1.1 Variationist sociolinguistics

Sociolinguistics is the study of language use in its social context, ie. how different speakers utilise language as a tool to enact their identities during social communication and form judgments about others' identities based on their language use. More specifically, variationist sociolinguistics - the framework within which the present study is primarily situated - is concerned with analysing how language features undergo systematic variation among different populations of speakers (Young & Bayley, 1996: 254). The study of variationist sociolinguistics is predicated upon the concept of the linguistic variable, that is, any element of language that is able to alternate among two or more competing forms ('variants'), without altering the basic truth value of the speaker's utterance. Therefore, a speaker's use of a particular variant of a linguistic variable over another can be understood, in simpler terms, as the speaker saying the same thing in a different way (Labov, 1964: 166).

A prototypical example of such a linguistic variable comes from Labov's study (1966) of realisations of the /r/ phoneme in different department stores in New York. Labov systematically observed store workers' pronunciations of the words "fourth floor", observing whether they used rhotic pronunciations (overtly vocalising the /r/, as in Standard General American English - [fɔrθ flɔr]), or non-rhotic (not overtly vocalising the /r/, common among New York speakers - [fɔːθ flɔː]). The three chosen department stores were stratified according to the general socioeconomic standing of their typical clientele bases. Labov's findings revealed "structural and ordered" heterogeneity (Weinreich et al. 1968: 100-101) to the pattern of /r/ variation across these stores; Workers in the upper-class store, Saks, used more rhotic /r/'s than those in other stores - the variant associated with Standard American English (SAE), and, therefore, the less stigmatised variant. Workers in the middle-class store, Macy's used fewer rhotic /r/'s than those in Saks, and workers in the most working-class store, Macy's used even fewer. Furthermore, even within each store, the patterns of /r/ vocalisation were also stratified hierarchically according to the store worker's level of employment; ie. managers were more likely to use the standard, non-stigmatised variant, overtly vocalising their /r/'s, while floor workers were found using the more stigmatised non-standard variants more often. Labov also found this pattern to be influenced by attention paid to speech; Speakers were more likely to use the SAE variant upon repetition of the utterance, compared to the first time they had said it. Labov followed up this observation by studying New Yorkers' /r/'s while they were telling personal, emotional stories (Labov, 1972), and found that overt vocalisation of /r/ decreased as people became absorbed in the act of storytelling, and a noticeable increase in /r/ correlated with shifts in the topic of conversation to matters such as politics. These findings illustrate the multifaceted nature in which (socio)linguistic variation operates; Patterns can be seen not only across the more fixed, macro-level categories we may divide speakers into, such as sociodemographic factors (eg. socioeconomic status), but also, across micro-level dimensions occurring throughout specific points in interactions (eg. with variation occurring according to the topic of conversation, and the register, style, or level of emotional affect a speaker is currently employing). The field of variationist sociolinguistics has undergone several observable 'waves' of approach, as chronicled by Eckert (2012). What follows is a brief and non-exhaustive overview of these 'waves' of the changing focus of sociolinguistic inquiry, which foregrounds the context upon which the present study is built.

Studies belonging to the first wave (within which Labov's early studies were situated, together with studies including: Wolfram, 1969; Trudgill, 1974; and Macaulay, 1977) primarily focus on analysing patterns of non-standard variation according to measurable, stratifiable "macrosociological categories" (Eckert, 2012: 88) such as socioeconomic class or gender, viewing speakers, themselves, as "human tokens" (Eckert, 2012: 88) - ie. representing clusters of measurable characteristics, which could be used as points of comparison. This level of focus, of course, has its limitations regarding the depth of inquiry that can be made into within-speaker variation - we gain an overall snapshot of a person's rate of standard versus non-standard usage of a particular variant, at the expense of fine-grained detail of how that person's rate of non-standardness might vary from moment-to-moment during the course of their daily interactions.

The second wave marked a collective shift towards this latter notion, with the influx of the ethnographic approach to the study of variation. Here, the methodological focus was on the active integration and participation of the researcher in the daily experiences of the participants in order to gain a deeper understanding of the interaction between speakers' group memberships and their navigation of interspeaker communication at specific points in interaction. Consequently, these studies revealed how the interpersonal and inter-group dynamics that speakers are embedded within result in observable patterns of sociolinguistic variation. This second wave of variationist study began with a focus on the roles that social network structure plays on phonological variation. An example is Milroy's (1980) study of Belfast speakers. Milroy defines an individual's social network as the "sum of relationships [they] have contracted with others" (Milroy, 2000: 217). In this study, she observes the connection between the density of working class community networks, as well as their joint participation in shared practices and behaviours, and the speakers' positive use, at this local community level, of linguistic variants which, on a global societal level, are typically stigmatised. The study uncovers a phenomenon which Milroy terms *vernacular maintenance* (1980: 60-1), which can be thought of as a kind of centre of gravity that arises within more closely-knit communities, with speakers orienting their language use towards the 'pull' of the values and norms of the community, overcoming the draw of overall hegemonic norms. Milroy theorises this as a kind of "local norm-enforcing power" (Eckert, 2012: 91) generated within such networks. Central to this work on social networks is the notion of "multiplexity" (eg. Milroy, 2000: 218), a way of classifying networks according to the relative strength and interconnectedness of individuals' social bonds with others in the network; A social network in which all or most individuals have strong mutual bonds with each other are considered to be multiplex. Milroy's (1980) study of working class Belfast speakers indeed found a correlation between the use of non-standard variants and the individuals' embeddedness in dense, multiplex community networks. Consequent second wave studies verified this observation with a shifting towards ethnographic methodologies, where researchers embedded themselves within the communities under study. Cheshire's (1982) study into the nonstandard variation of adolescents in Reading, UK found a link between speakers' practices, behaviours, and cultural values, and their use of certain nonstandard morphosyntactic variants (that is, those rooted in the underlying structural, morphological and syntactic elements of language, as opposed to comparatively more 'surface' elements such as the speech sounds produced - an example is nonstandard *was* in "You *was* angry". These findings about the key role that speakers' practices play in conditioning their patterns of sociolinguistic variation motivated further investigation. The Community of Practice (CofP) framework was introduced into variationist sociolinguistics during the second wave by Eckert and McConnell-Ginet (1992), and provided a theoretical approach for analysing variation according to the dynamics of social power that arise amongst groups of individuals undertaking regular mutual engagement in a social practice, such as the social categories of "jocks" and "burnouts" in Eckert's (2000)

ethnographic study of a high school in Detroit. Eckert found that adolescent speakers from these ideologically opposed CofPs diverged linguistically; The middle-class, suburban, college-oriented jock girls produced below average rates of nonstandard negation and of the nonstandard phonetic variants under study, whereas the working-class, urban, anti-establishment, vocationally-oriented burnout girls comparatively demonstrated above average rates of many of these nonstandard variants. "Jocks" and "burnouts" were not found to be discrete categories; Both entail continuums, with individuals expressing different degrees of affiliation with either label, and some identifying as 'in-betweens'. A particular sub-group of burnouts, for instance, known as "burned-out burnouts" who, of all the burnouts, identify particularly strongly with this label, and consistently demonstrate far above average use of all non-standard variants under study. This demonstrates the importance of analysing individuals' strength of orientation towards both local and wider values and community identities when analysing language variation. Many other ethnographic studies have reinforced this connection between speakers' enacted social practices and their use of language variation (eg. Moore, 2004; Jones, 2012; Nance, 2013; Howley, 2015). Crucially, Eckert and McConnell-Ginet's reflection on the CofP framework (2007) signalled the importance of anchoring the analysis of variation and social practice within the wider world. Although understanding the effects of individuals' practices on their instantiated uses of language variants can be insightful on the 'ground level', Eckert and McConnell-Ginet stress that the overall focus should ultimately be on relating the effects of specific practices to their wider relevance, as these practices, through their structured and repeated enactment, inherently embed the individual within the wider societal paradigms (and the macro-social categories which we apply to parameterize them) that the first wave of variation focused on, such as the ordered societal structures of gender or socioeconomic status.

Another important product of the second wave was the theorisation of social meaning. This term refers to a property of linguistic variants which are *marked* (ie. significant) within the "social landscape" (Bender, 2001: 200) - ones which are distinguished particularly as carrying out social functions. Variants of the same variable, therefore, despite conveying the same referential meaning (ie. meaning the same thing 'on paper'), can vary in their social or stylistic significance and, therefore, carry different social meanings, and even work to facilitate differing discourse pragmatic functions. The functions that linguistic alternants can carry out will be explored in more detail in Section 2.6. For the time being, it is important to note the role of social meaning in the second wave theorisation of sociolinguistic variation, and the view that the transmission and interpretation of linguistic social meaning is not necessarily enacted on a conscious basis, rather, it is a process that very much underpins social communication (Ochs, 1992: 338).

In sum, the second wave marked a 'zooming in' of researcher focus from the level of macro-level categories down to the meso-level of the individual and the groups and practices they are embedded in. The third wave of variationist study, in turn, marked a further focusing in on the micro-level individual speakers' stylistic practice over generalisable community norms (Sorace, 2011), acknowledging how speakers' awareness of social meaning and their agency during communicative acts plays a causal role in their instantiated use of language variation (Nycz, 2018: 176). A key element of this wave was the theorisation of the social meanings tied to linguistic variants not as fixed connections, but as a phenomenon arising from speakers' continuous and repeated anchoring of themselves within the social space, and an unavoidable phenomenon within language. Studies within this wave reveal that speakers' use of language variation is not merely as a result of their belonging to a particular social identity, but it is something that is deliberately enacted by speakers and it is this enactment, in turn,

that anchors them within wider group and community dynamics, and then, on a more global level, to the broadest level of macro-social categories that we can see emerge.

The social-semiotic framework of indexicality is an important theoretical underpinning within this wave of research. This is the notion that certain elements of language can be used to 'index' (or 'point to') broader, socially meaningful information - for instance, the use of nonstandard *were* can index, at least among British Englishes, a speaker's Northern and working class identit(ies), and the qualities associated with these identities. Crucially, the third wave took the more static interpretation of indexicality put forth in earlier research and reinterpreted the indexical links between variables and their broader social meanings as mutable, multi-purpose, and continuously under recombination and reinterpretation (Eckert, 2012: 94) through a process of bricolage (Hebdige, 1984). For instance, Podesva's (2007) study reveals how a speaker draws upon the indexical value of the hyper-articulated aspiration of intervocalic /t/ and its enregisterment within school-teacher speak to construct different personas in workplace versus casual settings, through the use of varying styles (ie. 'socially meaningful clusterings of features' (Campbell-Kibler et al., 2006). The speaker in Podesva's study specifically used /t/ as part of a style indexing "hyperarticulateness or prissiness" (Eckert, 2012: 96), however, as Eckert outlines, depending on its usage, this same variant can be used as part of different constructed styles to index a wide range of qualities, from politeness to anger.

This mutability that Eckert (2012: 94) describes comes about as a result of the ordered process of indexicality (eg. Silverstein, 2003), originating from 'on the ground' direct indexes (Ochs, 1991) between individual speakers from a salient population and a particular trait or set of traits associated with them. If these speakers also use a marked linguistic variant frequently enough - particularly if for the purpose of meeting a specific interactional goal that can be ideologically linked to the associated trait(s) - this can result in indexical links between the associated trait(s) and that variant (and, in turn, the associated population). Through this process, over time, use of the variant becomes indexically associated with membership to that population. That variants' use can then develop broader ideological dimensions, being used either to articulate one's association or affinity with that population, or being encoded as part of stereotypical or caricatured portrayals, serving to distance oneself from that population. Because the order of indexicality is non-linear, and parallel indexical links can evolve simultaneously and be utilised to varying ends by different speakers or groups, we can think of the range of possible indexical links to one variable as its *indexical field* (Eckert, 2008), ie. a multi-layered "constellation of [its] ideologically linked meanings" (Eckert, 2012: 94), which can be drawn upon in differing ways at different times.

This overview of the three waves of variationist focus, and the ways in which we conceptualise social meaning through the framework of indexicality, reveals a feedback loop of sorts; our focus has moved from the macro-level socio-demographic dimension, through the meso-level community dimension, to the micro-level individual dimension, however, when we arrive at this point, we discover that these individual micro-level instantiations of direct indexicality are both conditioned by and, with enough repetition, establish themselves *in*, the meso- (community) level (Du Bois 2002). Similarly, the social meanings indexed through communities' behaviours and practices at the meso-level both generate, and are constructed by, macro-level societal constructs (Eckert and McConnell-Ginet, 2007: 28). These orders of magnitude through which we analyse variation are inextricably interconnected, whereby the variation we can observe is both a result of, and a key mechanism *in the construction of*, social meaning. Because waves of variation are not unidirectional or linear, but recurse back and forth between, and simultaneously co-occur across, the largest and smallest orders of magnitude at which

we have observed them, the rationale within the present study is that one can capture a snapshot of these patterns of variation across a *subset* of these orders of magnitude (for instance, analysing patterns of variation across macro–level categories such as age or gender), providing us with a static impression of the fluid process taking place across the wider system. This is similar to knowing the mean value of a set of numbers but not every number involved in its calculation - it does capture an overall pattern. Similarly, knowing what variants a set of working class speakers (for instance) use, can help us capture abstract patterns of association between certain variants and this hypothetical, average working class speaker (who doesn't actually exist but is the combination of all the patterns of variation among all the working class speakers in the set).

Although lateral analysis across one level can be informative, if we are interested in how variation patterns across a community of speakers, it is arguably necessary to interweave relevant elements of focus from each of the various waves of variationist study and analyse variation across multiple levels (eg. both meso *and* macro) in tandem, in order to more deeply inform our understanding. Eckert & McConnell-Ginet (2007: 29), for instance, highlight the importance of anchoring meso-level community-oriented analysis within the relevant, broader context of the social order that these communities operate within. Their discussion focuses on the analysis of variation with regards to the social factors of gender and sexuality but, since the same social systems and orders underpin all macro-level categories (eg. age, socioeconomic status, etc.), it follows that we should aim to apply this same principle to any sociodemographic factors we are interested in. Moreover, although Eckert & McConnell-Ginet's discussion focuses specifically on Community of Practice approaches, these communities directly influence, and are directly influenced *by*, the wider network structures they are embedded within, therefore, the points made here are relevant beyond the Community of Practice framework. Most notable to the present study, Eckert & McConnell-Ginet point to the consideration of broader social structures when carrying out community-oriented analyses of variation. This includes not only the wider social networks that communities are embedded within, but also even larger and more abstract dimensions of social structures, such as the consideration of 'imagined' elements of communities. These 'imagined' elements of communities refer to the fact that communities can be, at least in part, based around a shared identity, for instance, national identity (eg. the Polish community) or another characteristic, such as a minority identity (eg. the LGBTQ+ community). These elements of communities do not directly map on to a shared location in physical space but, nonetheless, contribute towards a shared sense of community identity amongst members. We are currently experiencing an ongoing, rapid increase in transnational mobility, with community dynamics operating in increasingly globalised, superdiverse and multilingual contexts (eg. Blommaert, 2010; 2014; Mesthrie, 2014: 277) and, with the widespread accessibility of digital communication, these community dynamics are no longer limited by speakers' physical proximity, but also abstractly enacted in intangible spaces. These ongoing contextual and modal shifts in the ways speakers and groups interface with each other necessitates a consideration of these elements that bring communities together, beyond traditional physical and spatial elements of community membership, if we aim to broaden our understanding of how variation is influenced by speakers' orientations toward, and embeddedness within, the communities to which they belong. Different communities, and particularly those that have reason to orient away from hegemonic norms (such as the LGBTQ+ community), have different sets of values, attitudes and orientations that bring the communities together, compared to the wider population. Therefore, if we think about the system of variation as a feedback loop, enacted on the micro-, meso- and macro- scale simultaneously, then, folding in these additional aspects of community and group structures to our current

understanding of variation at its various orders of magnitude represents a *widening* of this loop to be more fully representative of the different functions and dynamics of communities and networks that exist within our current social context.

Crucially, because, as third wave research has uncovered, multiple indexical values can be linked to any one variant, and these indexical links are not fixed, if one is interested in how a population (or a community within a population) of speakers might enact socially meaningful variation, we must ask ourselves not 'what does x variant index across this population?' (because we cannot isolate just *one* meaning), but 'which variants become socially meaningful across this population, and how are these social meanings accrued?'. This is especially important if we are interested in how variation might be enacted by a new population (eg. a migrant population) merging with an existing population, and when considering non-hegemonic communities which may have reason to pattern differently from the broader population. Since variants can be socially meaningful across multiple indexical orders (Silverstein, 2003; Snell, 2018: 7), Moore (2011: 366) argues that a variant's sociolinguistic vitality may be more likely to be maintained if a variant is socially meaningful (ie. can be purposefully exploited by speakers to enact communicative goals) across several *different* levels of meaning. The broader a variant's indexical field, the more potential options that variant has to be incorporated into socially meaningful variation, and the more likely speakers are to recognise it as socially salient. This suggests, then, that if we identify variants that have been shown to have a broad indexical field among a population (for instance, L1 English speakers), we can use these to compare how newer members to this population (eg. migrants) might enact variation relative to the dominant community. As we have seen, there is no 1-to-1 direct index between any social meaning and any one variant, but, instead, speakers incorporate specific variants into their speech while enacting their communicative goals. This means that, when we take a top-down approach to the analysis of variation (ie. one that does not have access to 'on the ground' micro-level information, such as variation arising from instances of stance-taking, but focuses on macro- or meso-level patterns of variation across a community or population), we lack overt evidence of the micro-level indexical links constraining these variants' usage. In absence of this information, we can instead consider how we might *categorise* the different communicative purposes that the dominant (eg. L1 English-speaking) population might desire to enact through variation. For instance, we can identify variants which have broad indexical potential within the L1 population (based on prior findings), and which we can evidence as being particularly likely to be used by the population to facilitate these communicative purposes that have been identified. By investigating how L1 English-speakers orient towards different morphosyntactic variants, we can then assess how *newer* migrant populations (such as the Polish-national community), as well as non-hegemonically oriented communities (such as the LGBTQ+ community) pattern in this regard (and the extent to which they do so), compared to the dominant population. Do we see all communities orienting similarly to the variants according to their different underlying communicative purposes, or do we see evidence that these communities pattern differently? If we find the latter, this suggests that membership to one of these particular communities is associated with enacting different communicative purposes, perhaps because members of these communities (especially in the case of migrant members) have not established indexical links with the variants in question. Given the importance of the communities under investigation to the present study, the following section consists of an introduction to the non-L1 participants who have informed this study.

1.2 Introduction to the participants

The present study focuses on LGBTQ+ Polish-born migrants to England, and the following section explores the key sociocultural context underpinning the lived experiences of this group.

There is a relatively large number of Polish-born people living in the UK and, consequently, there exist throughout the country robust Polish networks and communities; Statistics show that, of migrants from EU8 countries (those that entered the EU in 2004) living in the UK between 2013-2015, the vast majority (over 800k people) were from Poland. Of these, an overwhelming majority (around 760k) were adults (ONS, 2017; Figure 2). Therefore, Polish-born adults constitute a significant proportion of England's migrant population and an important part of the job market and overall societal structure.

Although some English cities with large numbers of migrant inhabitants, such as Sheffield, have been assigned 'Cities of Sanctuary' whereby there are policies implemented to provide a safe environment for incomers (Kirkham, 2013: 48), migrants face many additional challenges navigating cultural divides in England, such as the current post-Brexit climate and increase of anti-immigrant sentiment. Szulc (2019: 4) found that 76% of LGBTQ+ Polish-born respondents were against Brexit, including 60% strongly disapproving Brexit. It is important to take into account the possibility that these additional pressures faced by migrants are likely to influence their acculturation pathways (ie. the ways in which they adapt to their new culture) as they navigate this process. Due to the prevalence of Polish communities around the country, Polish migrants have the option of only acculturating to English culture in a limited way, if they wish to, or they have the ability to be highly Polish-acculturated even if their English acculturation is also significant (ie. a bicultural orientation). This makes this group an ideal test case for researching the interplay between migrants' motivations and their acculturation levels.

We must also consider Polish-born migrants' place within the LGBTQ+ community, and I begin by defining the LGBTQ+ label.

1.2.1 Defining LGBTQ+

Identity labels can be difficult to accurately capture and define, and the LGBTQ+ label is no exception (Vincent, 2016: 3). The acronym stands for Lesbian, Gay, Bisexual, Transgender (or, for some, Transsexual[1]), and Queer/Questioning. The plus (+) symbolises that this category is not limited to this handful of labels, but also consists of an ever-expanding and highly nuanced set of labels for gender and sexual orientation identities. Theoretically, 'LGBTQ+' subsumes any person with a non-normative sexual orientation or gender (ie. anyone who isn't exclusively heterosexual and/or cisgender). The category inherently lends itself towards nuance over black and white categorisation, in part due to the blurring of, and recognition of the interplay between hegemonically categorical distinctions such as those between male and female, or gender and sexuality.

It is clear that there exist stark differences in the sociocultural and political treatment of non-normative sexualities and gender identities in Poland compared to that in Britain. One notable measure of this is ILGA Europe's benchmarking tool

[1] This term is generally dispreferred, but, because some LGBTQ+ people do identify with this label, it has been included here

'Rainbow Europe', which compares the civil and political rights of LGBTQ+ people according to the laws and policies across 49 European countries, taking into account 6 categories: equality and non-discrimination; family; hate crime and hate speech; legal gender recognition and bodily integrity; civil society space; and asylum. Rainbow Europe (2020) measured the level of LGBTQ+ human rights in Britain at 66%, compared to only 16% in Poland. It is particularly important to note that, in 2020, Poland was the lowest rated EU country for LGBTQ+ rights. With the openly anti-LGBTQ+ 'PiS' party in governing power, combined with the branding of almost a third of the country with 'LGBT ideology-free zones', LGBTQ+ Poles are facing increasing discrimination (Golebiowska, 2014: 141), with LGBTQ+ rights being considered an imposition on national values (Stychin, 2003). These recent events mark a pinnacle in the treatment of LGBTQ+ issues in Poland, and, therefore, exploration of the motivations and identities of LGBTQ+ Poles is more relevant now than ever before.

According to Szulc (2019), LGBTQ+ Polish-born migrants in the UK are a cohesive group, sharing many values, attitudes and opinions. Given the situation for LGBTQ+ identities in Poland, it is not surprising that LGBTQ+ status impacts migrants' motivations. For instance, Szulc found that queer Polish-born people living in the UK are more likely to be out to their friends in the UK than those in Poland. The intersection between LGBTQ+ identity and migrant identity is a relevant factor for LGBTQ+ migrants, with Szulc finding that LGBTQ+ issues factor into the rationale for moving to the UK for more than a quarter of his 767 participants. Although practical matters, such as work opportunities, were listed as the most common reasons for migrating, Szulc's participants demonstrably perceived the UK as having a high level of freedom and acceptance of LGBTQ+ people, compared to the increasing anti-LGBTQ climate that persists in Poland. Szulc found that LGBTQ+ Polish-born migrants in the UK had in common similar struggles with anti-LGBTQ and anti-migrant discrimination. This is something that is explored by Acosta (2008), who argues that migrants bear the additional challenge of negotiating otherness in LGBTQ+ spaces due to their migrant identities. Nonetheless, despite this, 75% of all Szulc's respondents planned to stay in the UK and, for many people, LGBTQ+ issues were a key factor in this.

Now that I have discussed the participant sample of interest to the present study, and overviewed some key contextual aspects of their lived experiences, I turn to outlining the research questions for the present study.

1.3 Research questions

The aim of the present study is to explore factors on the acceptance (ie. perception) of morphosyntactic variation in a non-L1, and to do so by considering the effects of macro-social (ie. demographic) factors; practical L2 acquisition constraints relating to degree of linguistic input and proficiency; and meso-social factors (relating to participants' degrees of community embeddedness).

The present study is concerned with answering the following research questions:

1. How do participants pattern in their acceptability judgement ratings of British-English morphosyntactic features, according to type of morphosyntactic construction - argument movement; optional discourse-based movement; nonstandard agreement? How do Polish-born participants' acceptability judgements compare to those of English-born participants?

2. To what extent can factors relating to formal linguistic accounts (eg. degree of prior linguistic input; linguistic proficiency) capture the patterns of variation found in Polish-born participants' acceptability judgement ratings?

3. What more can we learn about participants' acceptance of morphosyntactic variants by considering macro-social factors (eg. age; gender; socioeconomic class)?

4. Do participants' acceptability judgement ratings pattern differently according to participants' British or Polish acculturation levels?

5. How does participants' acceptance of morphosyntactic variation vary according to participants' membership to, and level of embeddedness within, the LGBTQ+ community, compared with the overall Polish-born and English-born populations?

1.4 Thesis structure

The following is an overview of the structure of this thesis. A summary of the contents of each of the thesis chapters is provided here.

The Background and Literature Review (Chapter 2) outlines the theoretical groundwork that the present study builds upon. It covers the sociolinguistic underpinnings which motivate this research, including how sociolinguistic theory has previously been applied to inform our understanding of the patterning of variation at different social scales. In doing so, I review previous studies' findings about the significance of macro-social factors (eg. age, socioeconomic factors, region, gender, and sexuality), as well as how these factors may interact with each other. Following this, I discuss community-oriented approaches to variation, and the types of communities relevant to the present study. I discuss communities at the national level and also the LGBTQ+ community, how level of embeddedness in these communities is relevant in the context of migrant language users, and how the acquisition of language variants by non-L1 speakers has been approached in traditional second language acquisition studies. I then focus particularly on the theory of acculturation and the sociocultural relevance of the participant demographics chosen. Following this, I turn to the literature relating to morphosyntactic variation, outlining the linguistic variants of interest in the present work, and why they have been selected as the subjects of analysis. Finally, I bring all of these aspects together, restating the research questions of this study after having explained why they are pertinent.

The Methodology (Chapter 3) outlines the methodological structure of the present study. In order to tailor the method to this study's research aims, several distinct methodological elements were required. I outline the design of these elements, which are as follows: (i) an Acceptability Judgement Task, to measure participants' acceptance of British-English morphosyntactic variants; (ii) the Versant English Language Speaking Test, to provide a formal measure of Polish-born participants' English language proficiency; (iii) Sociological surveys, to measure participants' Polish and English acculturation levels, and their LGBTQ+ Community Involvement levels, and to collect demographic information. I outline the participant sample characteristics and explain how participant recruitment was undertaken. Then, I explore the study design and justification for use of the given methods. I reflect on the pilot study process and the changes made following this. I also outline how the data was collected, cleaned and analysed, and how the statistical modelling procedure was conducted.

The Descriptive Results (Chapter 4) are covered next, and these illustrate the patterning of the dependent variable (AJT Response) according to each of the AJT test conditions (ie. the linguistic variants of interest in this study).

Correlations are investigated between each factor and the AJT Response. For each factor, the distribution of participants across its categories is first considered, to assess whether the sample is reasonably equally distributed according to country of birth and/or LGBTQ+ status. AJT Response results are then considered according to the type of morphosyntactic construction: (i) argument movement (raising-to-subject with *seem*); (ii) optional discourse-based movement (left dislocation, right dislocation, topicalisation); (iii) nonstandard agreement (past-tense BE: nonstandard *was*, nonstandard *were*, nonstandard *weren't*). These results are visualised using stacked bar charts, which show the percentage distributions of AJT Responses across each increment of the AJT scale. Exploration of these effects begins with factors related to key sample characteristics (ie. birth country and LGBTQ+ status), followed by macro-social (ie. demographic) factors, namely age, gender, region, and the socioeconomic factors of socioeconomic status and education status. This is followed by an analysis of the second language (L2) factors influencing Polish-born participants, namely their age of arrival to England and their English language proficiency. Finally, I explore the effects of the meso-social factors considered in this study (ie. those related to community embeddedness). These include participants' levels of English and Polish acculturation, as well as their levels of LGBTQ+ Community Involvement.

In the Inferential Results (Chapter 5), I present an exploration of the statistical findings from four proportional odds (ordinal logistic regression) models. I take a nested approach in order to segment the sample according to the key characteristics under investigation (birth country and LGBTQ+ status): Model 1 incorporates all participants (both English-born and Polish-born, and both LGBTQ+ and non-LGBTQ+); Model 2a incorporates only Polish-born participants (both LGBTQ+ and non-LGBTQ+); Model 2b incorporates only Polish-born participants who completed the Versant English Speaking test of linguistic proficiency; and Model 3 incorporates only LGBTQ+ participants (both English-born and Polish-born). I discuss significant predictors in each model output and other main effects on the dependent variable of AJT Response (if applicable), as well as the interaction effects found between significant model predictors and the focal predictor of AJT Condition. These discussions have been structured similarly to the descriptive results (Chapter 4). I then summarise the findings for the four statistical models, comparing equivalent results between models. This is in order to explore how different factors influence participants' acceptability judgements, and will enable us to compare the differences in effects found across Polish-born participants and LGBTQ+ participants with those found across all participants.

The Discussion (Chapter 6) discusses the findings from the descriptive and inferential analysis. The first point of discussion is the patterns of acceptability across the three types of morphosyntactic construction under investigation. I then discuss how non-L1 Polish-born migrants integrate into these patterns, in the English cultural context, and how their patterns of acceptance of the different types of morphosyntactic construction might relate to these participants' lived experiences and motivations. As part of this, I explore Polish-born migrant's potential acquisition of indexical associations, as well as meso-level factors which might be inhibiting their acceptance of nonstandard morphosyntactic variation. Next, I consider the meso-social factors of cultural embeddedness (acculturation), and how this interacts with Polish migrants' acquisition of sociolinguistic variation. I then explore patterns of acceptance of morphosyntactic variation, as well as divergence from sociolinguistic norms, with a focus on LGBTQ+ membership and embeddedness in the LGBTQ+ community. Finally, I explore how existing at the confluence of the two communities under investigation in the present study - Polish-born LGBTQ+ migrants - impacts morphosyntactic variation.

Finally, in the conclusion (Chapter 7), I sum up the key findings from this study and review some potential wider implications.

2. Background & Literature Review

2.1 Introduction

The following chapter outlines the theoretical groundwork that the present study builds upon. I will cover the sociolinguistic underpinnings which motivate this research, including how sociolinguistic theory has previously been applied to inform our understanding of the patterning of variation at different social scales. In doing so, I will review previous studies' findings about the significance of macro-social factors (eg. age, socioeconomic factors, region, gender, and sexuality), as well as how these factors may interact with each other. Following this, I will discuss community-oriented approaches to variation, and the types of communities relevant to the present study. I will discuss communities at the national level and also the non-hegemonic community under investigation here - the LGBTQ+ community. In doing so, I will discuss how level of embeddedness in these communities is relevant in the context of migrant language users, and how language variation in non-L1 speakers has been approached in traditional second language acquisition studies. I then focus particularly on the framework of acculturation and the sociocultural relevance of the participant demographics chosen. Following this, I turn to the literature relating to morphosyntactic variation, outlining the linguistic features of interest in the present work, and why they have been selected as the subjects of analysis. Finally, I will bring all of these aspects together, outlining the research questions of this study after having explained why they are pertinent.

2.2 Sociolinguistic factors

Variationist sociolinguistics models the phenomenon of sociolinguistic variation as an integral and non-optional element of speakers' instantiated use of language in its social context (Eckert, 2019), and of the overall acquisition of language itself (Roberts, 2005: 153–154). As Milroy and Milroy (1985: 345) point out, some linguistic innovations come from the *speakers*, and not always the language itself. Identity factors play a part in how speakers use socially meaningful variation. However, this is a nebulous concept that can be difficult to capture and dissect, and has been approached in several ways by variationists. Identity has been theorised as a form of categorisation of people, either by themselves, or by others, into particular social positions (eg. Cameron & Kulick, 2003: 104). This view focuses only on the macro-level of sociodemographic categories that people end up identifying or being associated with, and does not take into account how speakers' 'on the ground' stylistic practices feed into, and, are also conditioned by, these identities. Because of this, some scholars have instead approached the notion of identity with the view that these macro-social categories (or 'social types') that we can categorise people into are macro-level "ideological instantiations or interpretations", generated from the reification of micro-level stylistic practices (Moore & Podesva, 2009: 449). As some scholars have called for ideological factors on language variation to be more integrally incorporated into variationist research (eg. Snell, 2018: 19), it is beneficial to our understanding of the driving forces of language variation to acknowledge the part that the ideological and interpretative nature of these macro-social categories plays, and not treat them as concrete, objective monoliths. Importantly, the 'reification' interpretation of identity fits into the aforementioned micro-, meso-, and macro- social orders across which we

can model variation (as suggested by, eg. Coupland: 2007, 13–14; Bucholtz & Hall, 2005: 592). Variation across each of these orders is intimately intertwined with identity; The construction of identity (and the use of linguistic tools as a means to do so) has been theorised as a series of agentive acts (eg. Le Page & Tabouret-Keller, 1985). Through speakers' ongoing construction and (re)interpretation of social meanings during their engagement in communicative acts, they are constantly situating themselves relative to identity frames at these differing orders of magnitude (Moore & Podesva, 2009: 449).

To recap, the present study takes the approach that the mechanism of variation is present and influential at every social order of magnitude (ie. the macro sociodemographic level, the meso group/community level, and the micro individual/instantiated level), and that, as Moore & Podesva (2009: 449) state, macro-social categories can be generated from the reification of instantiations of style. Therefore, if we compare laterally across, for instance, the macro-level layer, we can capture a 'snapshot' of the ways in which people are using variation across this subset of the social order hierarchy (ie. how variation patterns with regard to macro-social identities - middle-aged, working class, woman, etc). An additional element of fluidity worth noting is that the macro-level ideological interpretations that we categorise ourselves or others into are, themselves, not static, but reflective of fleeting states or currently dominant cultural ideologies. A person's identification with a particular category will almost certainly evolve over time, and this happens more predictably for some categories than others - for instance, a person's age will certainly increase. Other categories, such as gender, are also fluid; During their lives, transgender people come to identify with a gender category other than the one they were assigned at birth, and even cisgender people will likely re-interpret their own gendered identity over time (not to mention that the ways in which we divide up these categories also changes over time, eg. the presence or absence of gender labels beyond the binary of male and female throughout history and cultures). Therefore, comparing language variation against macro-level categories not only provides us with a static impression of the fluid process of *language variation* taking place across the wider system, but we are then comparing this against a static ideological impression of how (at least, at the time of data collection) participants are orienting towards the macro- identity frames under investigation. Capturing the information at the intersection of these axes of variation affords us a glimpse into the parallel factors that are influencing the linguistic variation process in that particular moment in time. This approach allows us to compare and rank the relative influence of these factors.

Our next focus is to establish which factors, which interact with the system of sociolinguistic variation, warrant inclusion in this study. The system of linguistic alternants, with their differing social and stylistic functions, can be seen to pattern in a socially stratified way (eg. Weinreich et al, 1968: 162), meaning that we must take into account social factors if we are to accurately model how variation is conditioned. In Section 1.2, I have argued the benefits that first-wave macro-social approaches can bring to the study of variation, and how they can be strengthened by implementing elements from second-wave approaches - for instance, by additionally taking into consideration speakers' community identities and the degrees to which they are embedded within these communities (see Section 2.5.1.1 for discussion). Consequently, the following section outlines the macro- and meso- sociolinguistic factors relevant to the present study, and how they have been treated in prior variationist research. The macro-level sociodemographic factors which I will be taking into consideration here consist of: age, socioeconomic factors, region of residence, gender, sexuality, and LGBTQ+ status. This will then lead us into a discussion of current approaches towards meso-level factors relating to community identities and degrees of embeddedness within these communities, and I will outline how consideration of migrant speakers helps inform our perspectives on these factors.

2.3 Macro-social factors

2.3.1 Age

Speaker age has long been considered a critical constraint on the acquisition and use of sociolinguistic variation, and this is additionally stratified by variant type. Previous research has shown that younger people more readily adopt incoming variants, such as the reuptake of standard *was* in New Zealand English (Hay & Schreier, 2004) and the confinement of nonstandard *weren't* to negative contexts in Fenland English (Britain, 2002: 17). Lexical variation, for instance, which occurs beyond any grammatical constraints, is generally considered as able to be acquired relatively easily, even by older speakers (Kerswill, 1996: 179). The potential for variation within the grammatical system, on the other hand, is said to reach a critical threshold in early adulthood (Tagliamonte & D'Arcy, 2009: 63). Therefore, in the present study, it will be tested whether age indeed has an effect on participants' acceptance of the morphosyntactic constructions under investigation.

Certain life stages, in particular, the adolescent years, represent a period of linguistic changes and stabilisation. Children undergo a process of socialisation, acquiring sociolinguistic norms through contact with their local communities, learning which variants have socioindexical value and meaning (Ochs, 1993), and some patterns of sociolinguistic variation can be acquired from a very young age (eg. Smith et al., 2013). However, during the adolescent years, speakers typically undergo rapid development of their social identities, and become increasingly self-aware about their position on a "linguistic market" (Sankoff & Laberge, 1978), where their use of language comes to function as a tool to both mediate social power relations and socially enact their identities (see Section 2.3.5 for more discussion of this with regards to gender and sexuality). Adolescence is associated with the enregisterment, abandonment, or entrenchment of sociolinguistic variants (Chambers, 2003: 195, referring to gender-oriented variation). Essentially, this period represents an expansion in speakers' "sociolinguistic competence'' - ie. their ability to use language variation to "engage with the social world" (Labov, 1972: 86). As such, the macro-social factor of age is inherently connected to other aspects of identity, in particular, gender (eg. Tagliamonte & Baayen, 2012: 139; Hay & Schreier, 2004), which I will explore in Section 2.3.5.

It has been disputed at which age this process of "vernacular reorganisation" comes to an end, as this is complex to definitively measure (eg. Tagliamonte & D'Arcy, 2009; Kirkham & Moore, 2013). The process of stabilisation has been associated by some with the ages of 14 to 17 (Labov, 2001: 447), in late adolescence (Tagliamonte & D'Arcy, 2009: 66), and, as a result, some (eg. Brook et al., 2018) have turned to the label of "emerging adulthood" (Arnett, 2000) to better encapsulate this life stage. Kirkham & Moore (2013) nuance the assertion that this stage marks the cementing of speakers' patterns of variation, given that we have evidence of linguistic change in adults (citing, eg. Sankoff, 2004). This is also supported by Fruehwald (2017: 4), who notes that "intraspeaker volatility" has been observed across various age groups, and by Tagliamonte & D'Arcy (2009: 62-3), who state that adults can shift their frequency of use of incoming variants, despite not seeming to participate in the process of grammatical readjustment itself. For these reasons, the present study is concerned with acceptance of variation in adult speakers of English.

The notion of a "critical-stage" in adolescence also applies to second dialect acquisition (Chambers, 2009: 181-84). Age of exposure to the dialect has been clearly shown to impact speakers' acquisition of variation (eg. Payne, 1980; Trudgill, 1986; Chambers, 1992) with speakers acquiring a dialect after the age of around 14 having their extent of second-dialect acquisition almost certainly affected (Chambers, 2003: 179). Similar age-related factors have also been shown to constrain

second language (L2) speakers' acquisition of variation, namely their age of migration and age of L2 acquisition. These will be explored in Section 2.4, however, I also forewarn here that evidence suggests this is not a straightforward matter, and a number of factors have been found to influence migrants' ability to acquire nonstandard patterns of variation in their target language. But first, I turn to socioeconomic factors on speakers' use of sociolinguistic variation.

2.3.2 Socioeconomic factors

Another macro-social factor which the present study is concerned with, in terms of its effects on acceptance of morphosyntactic variation, is that of socioeconomic status. This factor is particularly crucial to the present study due to the association between socioeconomic status, community norms, and use of particular nonstandard morphosyntactic constructions, which will be discussed later in this section.

Socioeconomic status has been the topic of much investigation within variationist sociolinguistics because social class is one of the core dimensions by which we can observe patterns of ordered heterogeneity in sociolinguistic variation. The consideration of socioeconomic factors on language variation is crucial for a number of reasons. For instance, they correlate with patterns of ongoing change; Labov (2001: 31) argues that language change undergoes a *curvilinear principle*, whereby changes from below the level of conscious awareness originate in social groups more central within the socioeconomic class hierarchy (ie. upper working- or lower middle-class communities) rather than more peripheral class groups. Additionally, the *types* of social network structures associated with different socioeconomic groups influence the ways in which language variation patterns among these groups. We can see this in phenomena such as *vernacular maintenance* (Milroy, 1980: 60-61), whereby the more closely-knit nature of working-class social networks generates a "local norm-enforcing power" (Eckert, 2012: 91) that bolsters the retention of nonstandard variants within these communities - variants which, while hegemonically stigmatised, amass a covert sense of prestige within working-class networks. Milroy found speakers' level of embeddedness within such networks to correlate with greater use of nonstandard morphosyntactic variants and argued these nonstandard forms serve an important function within working-class culture - to orient towards ideologies of solidarity and adherence to local norms.

A prominent pattern of socioeconomic stratification in agreement morphology is in the variation between nonstandard *was* and standard *were*, where use of nonstandard *was* (eg. '*You **was** here yesterday*') has been found to be heavily associated with working-class speakers, particularly in more informal interactional contexts (Tagliamonte & Baayen, 2012: 138). Cheshire et al. (2005: 3) note that, in many cases, middle-class English-speaking communities have been found to show avoidance of non-standard morphosyntactic variants. Levon & Buchstaller (2015: 323) argue that this is because such variants, which include non-standard agreement morphology, have become linguistic shibboleths - ie. they have become inextricably associated with certain regional and working-class social groups (and their associated aims and values) which middle-class speakers may wish to distinguish themselves from. The position taken in the present study is that this codification comes about as a result of the process of indexicality.

Another way in which morphosyntactic variation has been found to pattern by socioeconomic status is in the different linguistic strategies that speakers of working- and middle-class status have been observed using to reach similar interactional goals; According to Macaulay (2005), for example, to achieve the goal of making evaluations, Scottish middle-class speakers were argued to be more likely to employ adjuncts (adverbs and adjectives) to modify elements within

the matrix clause than working-class speakers, who instead made greater use of morphosyntactic tools such as right dislocation, that involve reduplication of structures outside the matrix clause. Much like with the use of standard versus nonstandard variants, this phenomenon of alternate utilisation of morphosyntactic strategies may also, at least partially, be caused by a similar desire among speakers from either class to differentiate from the other. It could be the case that, because the language structure permits multiple strategies for achieving the same goal, maybe it just so happens that salient members within each group have tended towards the use of a certain strategy over another and, because people tend to ideologically orient their language use towards others who share their values and, following the 'reification' interpretation of identity (see Section 2.2), these patterns of linguistic choices have then resulted in these emergent patterns of differentiation by social class.

This, and the previous discussion of nonstandard *was*, once again return us to the key role played by speakers' instantiated usage of variants as part of their enacted practices. As such, the variation in use of certain morphosyntactic features among different socioeconomic classes has been demonstrated to pattern according to the associated values and practices that these variants facilitate. For instance, right dislocation - the reduplication of a clausal subject or object *after* the clause, forming a dislocated element known as a 'tag' or 'tail', eg. "*Is* ***it*** *brown or blond,* ***your hair***" (Snell, 2018: 10) - has been observed occurring more often in working class than middle class speakers (Moore, 2003; Macaulay, 2005; Sorace, 2011: 10; Moore, 2020: 3). Right dislocated tags perform affective, attitudinal, or evaluative functions that aid in the management of discourse and facilitate interpersonal communication (Timmis, 2010: 11), therefore, it is hardly surprising that members of different socioeconomic classes - who typically have differing ideologies, goals, and values - might diverge in their usage of these. Not only have working-class speakers been found to employ right dislocation more frequently than middle-class speakers, the contexts in which right dislocation is used differs by socioeconomic class when the dislocated tag is a personal pronoun. Moore (2003) and Snell (2018) both find that 1st person pronoun tags (eg. "*I want that one,* ***me***" - Snell, 2018: 11) are used by middle-class as well as working-class schoolchildren. 2nd and 3rd person pronoun tags, on the other hand (eg. "*She's horrible,* ***her***" - Snell, 2018: 13), are used almost exclusively by working-class children, and fulfil more interpersonal discourse-management functions. These include evaluation (both positive and negative) of people, places or events, and function as stance-taking devices, used by speakers to align themselves (either with or against) the subjects of their evaluation, and to otherwise mark the speaker's status within the group . While first person tags refer to oneself, and are therefore described by Snell (2018: 14) as "less risky" evaluations, 2nd and 3rd person tags directly draw upon intra-group dynamics between members. Snell (2018: 22) argues that the use of such evaluations is multivalenced, in that, despite the prevalence of negative peer evaluations observed among the working-class children, which seem to conflict with the traditional ideology of working-class solidarity, speakers' "candid and unmitigated" use of right dislocation to construct these oppositional status-based stances is in itself emblematic of the enduring nature of bonds between speakers in the community.

Macaulay (2005: 86) argues that the differences in the communicative *functions* carried out by such discourse features across the socioeconomic spectrum are perhaps of more importance than the mere differences in frequency of their usage by different socioeconomic groups. For one, these functional differences are present even when the raw frequency of use of a feature is similar across socioeconomic groups - such as in the case of the discourse marker *you know* (Macaulay, 2005: 86) - and, importantly for variationist analysis, variants that fulfil important communicative functions (in the case of

right dislocated pronoun tags, appealing to both solidarity- and status-based ideologies) are more likely to maintain sociolinguistic vitality and endure ongoing use within these communities as part of the phenomenon of vernacular maintenance (Snell, 2018: 22).

In this section, I have discussed many interactions between socioeconomic class and use of morphosyntactic language variation, however, it is important to note that social class is a broad category that is itself made up of several dimensions. Factors such as occupation, income, place of residence, and education all feed into socioeconomic status (Labov, 2001: 57-61), and are, themselves, also interconnected - for instance, the social prestige (and, often, the associated salary) of a given occupation is strongly associated with the level of education required to attain that position. Occupation is generally considered to be the factor most directly correlated with social class, and we can see this through the prominent class associations of different job types, such as the association between blue collar jobs and working-class communities (Labov, 2001: 57). Often, we find that the stratification of variation by these socioeconomic sub-dimensions aligns with overall socioeconomic class category. This is the case with nonstandard *was*, which, as shown earlier in this section, is heavily stratified by social class. It has also been observed to be stratified by education level in Ottawa, Canada, with higher educated speakers employing more of the standard *were* variant, and less highly educated speakers using more nonstandard *was* (Meechan and Foley, 1994).

Despite these socioeconomic sub-dimensions typically feeding into individuals' overall class category, it is sometimes necessary to investigate some of these sub-dimensions in their own right, depending on the specific factors pertinent to the variants and speech communities under investigation. This is a particularly important consideration when investigating the acquisition of nonstandard variants in non-L1 migrants, who may have different classed identities and ideologies, and for whom nonstandard variation may be less stratified by social class than for speakers embedded in said country and culture from birth. Labov (2001: 114) suggests that educational level is more fluid than occupation, arguing that the latter is most strongly correlated with family background and circumstances of birth, and, therefore, with early patterns of variation, while the former is associated with variation acquired at later stages of life. Education level is described as having a "cumulative, socialising effect" and its potential for conditioning the stratification of variation strengthens over the lifespan (Labov, 2001: 185), suggesting that the analysis of variation in adults warrants special consideration of education level. Education level is also an important factor when investigating nonstandard morphosyntactic variation, which has been the subject of much overt proscription from "standard ideology" (Eckert, 2019: 758) in the educational setting (Moore & Spencer, 2021: 2). As such, some nonstandard morphosyntactic features, such as nonstandard *were* (eg. "*She **were** late*"), which have been found to be relatively unconstrained by socioeconomic class boundaries (eg. Moore, 2011: 364), can, in fact, be found to be stratified by level of education. Although the effect was comparatively minimal compared to the macro-social factors of age and gender, level of education was a factor in Tagliamonte's (1998: 177) analysis of constraints on nonstandard *were* production in York, UK, with less highly educated speakers preferring the nonstandard variant. Therefore, due to the confluence of these factors, it is necessary to take into account education level as well as socioeconomic status when investigating variants such as nonstandard *were* in adult migrants. Not only are these speakers subject to the metalinguistic judgements and stigmatisation of these nonstandard variants that L1 speakers are, but have likely been overtly proscribed the "right" and "wrong" ways to conjugate the past tense of the copula verb in their formal English instruction.

I can now summarise our discussion of socioeconomic factors on variation. As Eckert (2012: 92) notes, and as I have previously argued in our discussion of language and identity, the macro-social correlations that we find between socioeconomic factors and use of standard versus nonstandard variants are not *directly* due to individuals' socioeconomic status, or its sub-dimensions of education, occupation, or income. Rather, because of the multivalanced nature of social meaning (Sorace, 2011: 6), Snell (2018: 22) explores the possibility that variants' local *use-value* (term from Skeggs, 2004) is conditioned by the broader ideologies prevalent within different socioeconomic groups (such as working-class 'solidarity' and 'toughness'). Furthermore, Moore (2003; 2020: 3) finds the use of right dislocated structures to be inextricably linked to the *practices* enacted by groups, specifically featuring in the language use of communities oriented around working-class practices. Papapolydorou (2010) suggests that social-class is a core factor that brings such Communities of Practice (discussed in Section 1.2) together, further strengthening the link between the macro-social factor of class and speakers' everyday lived practices. Therefore, due to the association between socioeconomic status, community norms, and use of nonstandard morphosyntactic constructions, socioeconomic status is an important factor to consider in the present study.

As mentioned previously, not all nonstandard variation is necessarily primarily constrained by socioeconomic class. This is the case with nonstandard *were*, which has been found to pattern with place identity, namely parental region of birth (eg. Moore, 2011: 364). Since Snell (2018: 19) also draws upon the interconnectedness of factors such as place identity to local meaning-making, I now turn to our final macro-social consideration: that of region, or place.

2.3.3 Region of residence

An integral aspect to the system of dialectal variation is that of a speaker's region of residence. This is a particularly important factor to consider in the present study given the regional distribution of some of the morphosyntactic constructions under investigation (namely, past-tense BE agreement constructions). Sociolinguists have been interested in regional variation since the field's inception; Large-scale surveys have been conducted to map regional dialect variation across England (eg. the Survey of British Dialect Grammar, Cheshire & Edwards, 1998), as well as work being conducted to map patterns of levelling and spread over time of individual variants (for instance, of nonstandard variants of the *was/were* paradigm - eg. Schilling-Estes & Wolfram, 1994; Smith & Tagliamonte, 1998; Anderwald, 2001; Britain, 2002; Britain, 2010). The patterns found for *was/were* variation will be explored in more detail in Section 2.6.3, but the aim of the current section is to outline how region of residence has been theorised with regards to language variation. Many approaches have tracked how individual variants which correlate, at least somewhat, with geographical area (such as nonstandard variants of *was* and *were*) also correlate with other macro-social factors, including speaker age (eg. Britain, 2002: 31; Cheshire & Fox, 2009: 15; Schilling-Estes & Wolfram, 1994: 291), gender (eg. Tagliamonte, 1998: 176; Cheshire & Fox, 2009: 23), education level (eg. Meechan & Foley, 1994; Tagliamonte, 1998: 176), and socioeconomic status (eg. Feagin, 1979; Cheshire, 1982, Meechan and Foley, 1994)[2].

Other approaches by variationists have come to centre identity in their interpretations of findings about regional variation as well as also measuring patterns of variant usage against macro-social factors. This is partially following the shifting focus of variationist research (as has been overviewed in Section 1.2), and in the ways the notion of 'region' itself has

[2] The studies cited here all focus on *was/were* variation, but similar studies have been conducted on many other regionally- constrained variants and their patterning with regards to these macro-social factors

been theorised. It is well known that mapping the precise area of effect of a specific dialectal variant is an impossible task as there exists no discrete, objective boundary for any specific dialect region (eg. Johnstone, 2010), and many nonstandard dialect variants are not constrained to just one specific dialect. Furthermore, rates of usage of variants across different areas are constantly changing as part of ongoing processes of language change, including geographical *diffusion* of forms outwards from a particular area, *levelling* of particular forms as a result of competing dialects, *simplification* of grammatical paradigms, and *reallocation* of social, stylistic, or grammatical purposes to different competing forms (Britain, 2002: 16). This is further complicated by rapidly increasing mobility among populations (Britain, 2010: 198), both intra-nationally and internationally, meaning that many people do not remain in their region of birth, or even country of birth their whole lives, and may immigrate or re-settle multiple times.

A region or any geographical area alone does not inherently condition the system of variation of its associated language variety, but the system arises as a result of the 'culturally constructed valances' (Ochs 1996: 417) reflective of the processes occurring within the social network structures of the people residing in that area. Instantiated uses of particular regional variants as part of communicative acts, during which speakers orient towards or against wider values - for instance the notion of solidarity (eg. Snell, 2018) - while carrying out their interactional goals, both generate (Ochs, 1996: 417) and draw upon indexical associations between the variants and the region (and its ideological associations) they are orienting to or against with their use of those particular variants. For instance, phenomena such as *vernacular maintenance* (as discussed in Section 1.2) arise in close-knit networks due to the individuals within those networks ideologically orienting *away from* hegemonic pressures from outside the network, and *towards* the internal values of their network(s) and the community/ies in which they operate (Milroy, 1980: 60-61). As these individual processes (which inherently occur as speakers engage in practices throughout their daily routinised activities, moving through time and physical space) accumulate, the networks created through these processes can be seen to manifest in physical space as groups and communities and, when projected to a larger scale, as places and regions (Britain, 2010). Therefore, at a macro-level, the ideological notion of 'region' can be theorised as a representation of this ongoing 'push' and 'pull' of individuals' orientations towards the power structures present in the world they experience around them (Johnstone, 2010), mapped onto physical space (Britain, 2010: 196).

This perspective on region, although interesting even in a purely theoretical sense, becomes especially relevant when we study speakers' orientations towards *marked* variants - ie. those which are significant within the "social landscape" (Bender, 2001: 200) and are distinguished particularly as carrying out social functions - in this case, those which have place associations. As I have discussed, regions themselves are fluid, and continually being created and re-created according to the identities and practices of people living within them. Because of the subtle complexities in the process of region-formation, one approach to measuring how speakers from different groups orient to a sense of regional identity is to zoom out, and compare across regions in a broad sense. The rationale for this approach has backing in the literature; Watt (1998, 2002), for instance, presents evidence of speakers orienting towards a 'modern' pan-Northernness to signal their regional identity (rather than orienting specifically to their more local region of Newcastle) (1998: 7), patterns which would be captured by a more zoomed out approach to the classification of different regions. Britain (2010: 202) comments that Watt's participants achieve this pan-Northern orientation through their use of *supralocal* forms - ie. more widespread, but still regionally distributed, variants which have superseded more localised variants due to having more "socio-spatial currency" (Britain, 2010: 193). This is likely consistent with the phenomenon of dialect *accommodation*, whereby speakers in situations of

dialect contact, may avoid more marked localised forms in order to facilitate more comprehension in interaction (eg. Trudgill 1986: 25) - the use of supralocal forms here, instead, could enable speakers to orient to a broad sense of regional identity while minimising alienating their interlocutors. Finally, the use of such supralocal forms has been found to correlate with macro-social factors such as gender (Britain, 2010: 201), suggesting that a supralocal approach to regional patterns of variation is well suited to research that focuses on patterns of variation on a macro-social scale.

As already mentioned, the increasing mobility and transnationality of speakers is a crucial factor to take into account into our understanding of the patterning and use of regionally distributed variants (eg. Britain, 2010: 198). We have learned a lot from analysing the interactions between usage of regional variants and identity formation in people who have *always* lived in a certain region in which these particular linguistic forms are salient. However, a further dimension to our understanding is to be gained from folding in *newer* residents of a region into our analyses, and assessing how their presence in a certain region might fit within, or otherwise impact, the systems and processes of sociolinguistic variation. As I have discussed, salient regional variants can be utilised by residents, in part, to index their status as authentic members of that region and the communities embedded within it. From the existing literature, we know that mobile speakers are able to adapt to and exploit socially-salient indexical information in their new dialect while simultaneously negotiating the place identities associated with their original variety (eg. Nycz, 2018). We also know that migrants, despite not being native speakers of a language, are able to develop an understanding of and ability to exploit the system of variation present in their new area of residence in a way that draws upon indexical links to express aspects of their identities. This relies on them building an understanding of intergroup linguistic differences. For example, Doran (2004) documents a link between migrant youths' use of a nonstandard multiethnolectal variety and their ideological opposition to hegemonic power structure of French society. Drawing upon the notion of vernacular maintenance, the importance of group solidarity to these migrant youths, and their resistance of bourgeois ideology (of which the dominant language variety is representative), was found to motivate their use of the nonstandard variety.

In this section, I have discussed the importance of considering the macro-social factor of region of residence in the present study. It is important to take region into consideration given the regional distribution of some of the morphosyntactic constructions under investigation, and, in this section, I have argued the value of a supralocal approach to regional patterns of variation, as well as one which takes into account newer members to regional communities as well as those who are native to the region.

I will now discuss current approaches to the final macro-social categories relevant to the present study - those of gender, sexuality, and LGBTQ+ status.

2.3.4 Gender, Sexuality, and LGBTQ+ status

Sections 2.3.4.1-2.3.4.3 outline the importance of the macro-social factors of gender, sexuality, and LGBTQ+ status within the present study - particularly important factors given this study's focus on the LGBTQ+ community. Section 2.3.4.4 will then discuss the significance of the migrant context with regards to gendered patterns of language variation.

2.3.4.1 Gender

Sociolinguists have been interested in gender since the field's inception, with early work discussing the tendency (though, by no means, general principle) for women to lead in linguistic change (Labov, 1972: 303), while simultaneously adhering more strongly to prestige norms (Trudgill, 1972: 179). Labov outlines gender as playing a crucial role in his "Fundamental Principles of Change" (Labov, 2001: 272). To briefly summarise, these principles posit that women tend to have higher sensitivity towards overtly proscribed norms, and tend away from violating these norms compared to men, though do not show this same tendency towards nonstandard variants whose 'correct' usage is *not* overtly proscribed. Simultaneously, women are said to be readier adopters than men of innovative variants (regardless of whether these new forms are incoming from above or below a level of conscious awareness), while men tend towards retaining local norms. Although these early approaches highlighted some general trends, they also came with problems, for instance the "Gender Paradox" (Labov, 2001: 266) which suggests that women simultaneously avoid violating norms while also readily incorporating innovative forms.

This paradox is clear from the mixed results of studies evaluating use of nonstandard forms by gender. For instance, in Tagliamonte & Baayen's (2012: 138-9) overview of patterns of *was/were* variation, they comment on the irregularity in gender-related trends in variation. Hay & Schreier's (2004: 216) study, for instance, finds that, across seven bands of birth year, each spaced 20 years apart, women lead men in use of the nonstandard feature of plural existential *was* (eg. "*There* ***was*** *cows*" - Hay & Schreier, 2004: 211) in four of these seven birth year bands, and that, actually, birth year itself is a far stronger factor than gender on the patterning of this nonstandard feature. Tagliamonte & Baayen, therefore, question the expectation that male speakers favour more nonstandard forms given Hay & Schrier's results, arguing that this, along with many other studies, have found women to be more frequent users of nonstandard forms than men. Regarding the other side of the 'Gender Paradox' - that women tend to be readier adopters than men of incoming trends in variation - Tagliamonte & Baayen (2012: 139) note that studies have found an increasing reuptake of **standard** *was*, citing, for instance work on New Zealand English (Hay & Schreier, 2004), and Australian English (Eisikovits, 1991). According to these studies, this shift is being led by women and, to once again demonstrate the relevance of birth year and speaker age - these studies find that younger people are leading this trend.

Due to the inconsistencies in findings on gender-based language variation, the general discussion amongst scholars in the field has moved away from the more deterministic first wave approaches to gender. Not all members of any one macro-social demographic category ever pattern entirely homogeneously (eg. Ochs, 1992: 340 - referring specifically to gender), so we can be fairly certain that the gendered stratification that is often found across patterns of variation does not exist directly as a result of speakers' sex or gender. Rather, connections have been observed between gender-correlated language variants, the types of social information they can be used to transmit, and the attitudes and motivations of the speakers using them. This links together patterns of gender-stratified language variation and individuals' "patterns of social interaction in everyday life" (Labov, 1972: 303), as well as their broader lived experiences, such as their socialisation (eg. Ochs, 1992). Variationists interested in gender, such as Eckert (2008), have consequently turned to attitudinal and practice-based interpretations of gendered patterns of variation. Eckert compared the use of phonological variants associated with the Northern Cities Shift (and also one morphosyntactic variant - negative concord) amongst white adolescents in Detroit. She found two distinct Communities of Practice - ie. groups whose members undertake regular mutual engagement

in shared social practices. Use of the older, more stabilised variants of the shift was found to be stratified by gender, with girls leading the change. However, Community of Practice membership, rather than gender, was the key predictor in the use of the newer urban-associated variants, including the nonstandard negative concord. Here, while the anti-establishment urban-oriented 'burnouts' dominated the education-oriented 'jocks' in frequency of use, burnout girls, specifically, led in the use of two of the three newer phonological variants (Eckert, 2008: 459). Furthermore, a subcategory of even more urban-oriented 'burnt-out burnout' girls dominated all other groups, burnout boys included, in their use of all nonstandard forms. This demonstrates that, although it might not have otherwise appeared that gender is an important factor here, the added dimension of orientation-based community membership reveals a crucial interaction, within which gender most certainly plays a part. Eckert's findings, along with numerous other studies incorporating the Community of Practice framework (eg. Eckert & McConnell-Ginet, 1992; Bucholtz, 1999; Moore, 2011; Jones, 2012) assert the importance of considering speakers' lived practices, motivations, attitudes and orientations within the speakers' social contexts as a core component of variationist analysis, and not only the macro-social categories to which they belong. As part of this, Eckert's observation concerning the differences among level of 'burnt-out-ness' and its interaction with the macro-social category of gender highlights the value of measuring the level of embeddedness (ie. centrality versus peripherality) of speakers within the communities under study, if we are to more fully understand the constraints on variation. Therefore, gender is an important factor to consider when analysing nonstandard variation, especially when doing so in relation to community norms.

2.3.4.2 Sexuality

Following, and in parallel with, variationist focus on gender, patterns of sociolinguistic variation have also been explored with regards to sexuality. Although gender and sexuality are distinct phenomena (eg. Cameron & Kulick, 2003: 55; Livia and Hall, 1997: 5), the ways in which they interact with the system of linguistic variation are difficult and, arguably, impossible, to completely disentangle (Motsenbacher, 2011: 150; Levon, 2011: 69). For instance, Eckert (2011: 85) argues that the heterosexual norm of cross-categorical partnerships (ie. with the 'other' gender, in binary terms) represents an economic exchange rooted in power dynamics and "commodification of the self". During preadolescence, when speakers (in this case, girls) become aware of their place within the "social market" (2011: 89) - ie. the transactional system based around social value - they can use this to exploit the social value of heterosexual opposition to boys. Through their interactions with boys, and their entering into strategic status-boosting relationships, these girls socially engineer personae and styles rooted in values such as 'trendiness'. This is reflected in and, in part, achieved through, their use of language, resulting in the girls who are most invested in these processes leading the cohort in the use of incoming sound changes. Other findings also speak to the ideological interconnectedness between gender and sexuality in people's perceptions. For instance, speakers in Ontario, Canada, showed sensitivity to the markedness of the variables under analysis - specifically, whether the variants had gendered indexical associations - with heterosexual men avoiding variants associated with femininity (Hazenberg, 2016). One of these variants was the fronted realisation of /s/, which is associated with gender non-normativity and queer identities including, but not limited to, gay men (eg. Podesva & Van Hofwegen, 2014; Levon, 2015), and Hazenberg (2016: 270) argues this feature's avoidance by straight men is a stylistic measure to orient away from the potentially undesirable associations of femininity or queerness. This further supports the notion that gender and sexuality are interconnected in their indexical associations, and that this social information is able to be manipulated by speakers according

to their underlying attitudes and motivations. Given previous findings about the effect of sexuality as a constraint on sociolinguistic variation, this macro-social factor will be considered in the present study.

2.3.4.3 LGBTQ+ status

The following section outlines the relevance of considering LGBTQ+ status in sociolinguistic research, with a particular focus on incorporating identities beyond the gender binary into research, and with regards to research on community norms and their effects on sociolinguistic variation. All findings that have been discussed so far in Section 2.3.4 concern cisgender, heterosexual speakers, however, scholars have long been commenting on the importance of considering identities beyond these hegemonic norms into our models of language variation (eg. Bucholtz, 1999: 208). Consequently, many variationists have since turned their attention to queer identities - both in terms of gender and sexuality. Given that these non-normative identities can arguably be defined as existing external (and in opposition to) dominant societal structures and discourses (eg. Levon, 2011: 71), and also the fact that we know that speakers' identities (more specifically, their values and attitudes) reflect in their use of language variation, it is of interest to variationists to understand how LGBTQ+ communities fit into our paradigms of sociolinguistic variation (eg. Calder, 2020: 5; Konnelly & Tagliamonte, unpublished manuscript: 19). We know, for instance, that transgender groups discursively construct their identities and stances of opposition towards normative power structures (eg. Jones, 2022), and that queer communities have made use of cryptolects and argots - ie. codified, secret language varieties (eg. Baker, 2003), using techniques such as speech and gender play (Barrett, 2018) as part of their identity creation. Queer speakers' use of language is versatile and multifaceted; Podesva's (2004) sociophonetic findings demonstrate how stereotypical aspects of 'gay-sounding' speech form part of complex constellations of resources used to enact a variety of personae in different settings.

In particular, speakers whose gender identities lie beyond the boundaries of exclusively male or female present "challenges to the linguistic status quo" (Zimman, 2016, referring to transgender people's use of pronouns). Some sociophonetic studies (eg. Podesva & Van Hofwegen, 2016; Hazenberg, 2012) have found that cisgender, heterosexual people define the 'envelope of variation' (ie. set the limits for the upper and lower parameters) for the frequencies of /s/ realisation, with LGBTQ+ people operating within these parameters. These results have, however, been countered by other findings (eg. Zimman, 2021; Becker & Stoddard, 2018), which suggest that non-binary speakers set the envelope of variation that binary speakers operate within - specifically, binary transgender speakers in Zimman's study, and both cisgender and transgender binary speakers in Becker & Stoddard's study. Becker & Stoddard tentatively suggest that we are potentially seeing this 'flip' due to the relatively progressive nature of the speech communities their participants (and Zimman's, too) are embedded within. The studies discussed provide a persuasive rationale for variationists to continue to update the ways in which we conceptualise and measure gender and sexuality in our quantitative research (as discussed by, eg. Becker et al., 2022), in order for us to better understand how identities beyond hegemonic norms interact with the system of variation.

Despite many sociophonetic and discourse analytical approaches to queer identities, there have been comparatively fewer approaches to morphosyntactic variation with a focus on non-binary, or other LGBTQ+ identities. Examples include Konnelly & Cowper's (2020) theorisation of non-binary 3rd person singular *they* as being part of the most recent wave of a grammatical change in progress, which demonstrates another link between non-binary speakers and linguistic innovation.

Another example is Konnelly and Tagliamonte's (unpublished manuscript) 'small-within-large' study comparing the use of quotatives (eg. "*He* ***was like***: ...") and intensifiers (eg. "***so*** *good*") within an LGBTQ+ adolescent Community of Practice in Toronto and in the ambient Toronto speech community (ie. the general population of Toronto "who share [...] linguistic norms" (Labov, 1972: 158). Konnely & Tagliamonte's use of the *small-within-large* methodology (see Tagliamonte, 2012: 356) allows for the comparative analysis of variation across multiple scales - that of the smaller-scale local community and, simultaneously, of the wider overall speech community, providing nuance to otherwise overarching large-scale analysis. This comparison proves fruitful as the two participant groups show clear divergence in patterns of variation. Within the ambient speech community, use of both quotatives and intensifiers is stratified by gender, whereas members of the LGBTQ+ sub-community either pattern homogeneously regardless of gender, or all pattern independently from each other and from the ambient community. Some stratification by age is potentially present, with younger LGBTQ+ speakers having higher usage of incoming quotatives, though this is not statistically significant in the mixed effects model, while individual performance is. Furthermore, findings suggest the LGBTQ+ speakers not only lead the ambient community in use of the incoming *be like* quotative, but are also found to participate in the adaptation of the feature into new contexts of usage, unlike the ambient community. Therefore, it is pertinent to explore non-hegemonic communities such as the LGBTQ+ community if we are interested in how community norms may influence sociolinguistic variation.

2.3.4.4 Migrant context

Finally, I now turn to the contributions that migrant speakers make to our understanding of gendered sociolinguistic variation. Understanding the social mechanisms of variation is a key component of being able to use the language in real-world social contexts (Bayley & Regan, 2004: 325). 'Sociolinguistic competence' - ie. learning which forms can be used to carry out sociopragmatic functions (eg. Nycz, 2018) - is a core concept to the discussion around migrants' acquisition of variation. Howley's (2015: 217) study, for instance, finds both that female Roma children display patterns of sociophonetic variation aligned more closely with native Manchester residents compared to their male counterparts, and that the female children tend to have more open friendship networks than male children. Additionally, although the process of vernacular stabilisation occurs during adolescence, adult speakers can also be sensitive to patterns of variation, and migrant adults are no exception to this. Drummond's (2010: 219) findings show female Polish adults in Manchester of acquiring native-like sociophonetic variation to a greater extent than their male counterparts, which Drummond (2010: 149) links to the relative employment circumstances, with women more embedded in networks where they are exposed to a higher degree of native speaker variation. Tagliamonte & D'Arcy (2009) argue that speakers can increase their rates of usage of nonstandard variants (Tagliamonte & D'Arcy, 2009) and the studies I have mentioned here, along with many others, point to the relevance of migrants' own attitudes and motivations, and wider structures that form around these, such as the types of communities they become embedded within, all play a part in the patterns of variation we observe in these speakers. These findings reinforce the importance of incorporating comparison between more central and more peripheral members to our analyses of communities. Not all speakers within a community necessarily share the same norms, and scholars such as Bucholtz (1999: 208) note that community members who are migrants are often excluded from consideration in community-oriented analyses.

Despite these findings of gender stratification among migrants, the factors influencing variation in migrant adults have also been found to *diverge* from that of L1 English speakers. For instance, Adamson & Regan (1991: 325), investigating male Vietnamese and Cambodian L2 speakers of English, observed the use of nonstandard realisations of (ing) to *increase* in contexts where more attention was being paid to speech, with gender factoring heavier into this than the speech style itself. This suggests that, especially when it comes to migrant communities, we should be careful to consider the relative weightings of the various constraints and factors influencing variation, and observe how these rank compared with the ambient speaker population.

The findings discussed in this section motivate the consideration of not only how the gender of the Polish-born participants in the present study might condition their acceptance of the morphosyntactic constructions under investigation, but also their membership to, and levels of embeddedness in, the communities under investigation.

2.3.5 Interaction between factors

Scholars argue the interconnected and intersectional nature of extralinguistic factors on patterns of variation (Tagliamonte, 1998: 175). For instance, the factors of age and gender have been found to be intersecting constraints on the patterning of nonstandard *were* (Tagliamonte & Baayen, 2012: 139) as well as right dislocation (Durham, 2011). As explored in Section 2.3.4, the construction of gender and sexuality identities are closely intertwined (Motsenbacher, 2011: 150), and both have been linked to social class (eg. Moore, 2010; Moore & Podesva, 2009). As argued throughout this chapter, identities are fluid and their emergence is driven by practice (Bucholtz, 1999: 209), and enacted through interaction (Austin, 1962). Given this, the use of variants which are associated with particular macro-social identity categories are so only *indirectly*. Because we often see this overlap of several macro-social categories stratifying with a variable, this suggests that the social meaning related to these categories is not direct, but is derived from "something that is related to all of [these categories]" (Eckert, 2008: 455), ie. to some more core sense of identity, of which multiple aspects can be drawn upon in unison (Levon, 2015: 300). I have argued previously the importance of considering the relative weightings of factors influencing variation in the various communities under study. Given Bucholtz's (1999: 209) claim that multiple identity practices are enacted simultaneously (and speakers can fluidly index combinations of these identities), and Eckert's claim regarding the relatedness of macro-social categories, an additional aspect to take into account to determine the sociolinguistic vitality of particular variants of interest within different communities seems to be how many macro-social factors the variant is stratified by.

This interconnected approach to the measurement of macro-social correlations is relevant to studies on communities. Eckert & McConnell-Ginet (2007: 29) note that all types of communities are sites for the production of identity. We already know this is the case for Communities of Practice (see Section 1.2), with, for instance, Moore (2004; 2010) finding that high school girls in Bolton use variants linked to both class and gender (eg. tag questions) as part of positioning their identities within the community dynamics. However, it is not yet clear how interconnected indexical links pattern in all types of communities, for instance in migrant and in non-hegemonic communities. The communities of interest to the present study are discussed in the following sections.

2.4 Migrant L2 acquisition

Throughout this chapter, I have already discussed the contributions that migrant speakers make to our understanding of variation. A key element of mastering a language is *communicative competence* (Hymes, 1972), ie. a grasp not only of the grammatical structure, but an overall ability to use the language for a range of communicative purposes. A related notion is *sociolinguistic competence*, ie. understanding the social and pragmatic mechanisms of variation. Sociolinguistics models variation as "not a by-product of the learning process, but an integral element of overall language acquisition" (Roberts, 2005: 153–154), and this kind of competence is core to the discussion around migrants' acquisition of variation (Bayley & Regan, 2004). The focus of the present study is on the acquisition of variation in relation to acculturation, so, although some of the patterns of variation (particularly in the case of optional discourse-based movement) are found in the L1 of the Polish-born participants in this study (discussed in Sections 2.6.2 and 2.6.3), this variation is beyond the scope of this research, and only factors on the acquisition of L2 English language variation will be explored here.

Age- and time-related factors have been shown to constrain second language (L2) speakers' acquisition, namely their age of migration and age of L2 acquisition, as well as their degree of formal language instruction and, relatedly, their overall proficiency in the language. Acquisition of communicative competence and related linguistic practices requires time, therefore, it is no surprise that length of residence in the country associated with the language being acquired is a core constraint on degree of acquisition, as this is expected to directly correlate with the amount of *comprehensible input* - ie. high quality instances of language in the learner's environment that they can meaningfully interpret - available to the learner (Krashen, 1985). Some, such as Howley (2015: 213), do not find acquisition of variation to pattern with length of residence in the target country, and suggest that age of acquisition or age of migration may instead be more relevant. Howley (2015: 217) found that children with lower ages of migration tended to have more open friendship networks, perhaps because they had had longer to acclimatise and establish themselves within these networks relative to their starting age. This is supported by Jiang et al. (2009: 481), who state that learners who begin acquisition, even in a naturalistic environment, after puberty are "constrained by age-related maturational factors" that younger learners are not encumbered by. This brings us to the *critical period* hypothesis (Lenneberg, 1967), which posits that native-like acquisition is constrained by biological factors, such that, once a learner is beyond a critical period in their maturation (usually considered to be around the time of puberty), they may not fully master native-like proficiency in a target language. This hypothesis is supported by studies such as Johnson & Newport (1989), who found that Chinese and Korean L2 English speakers' performance on a Grammaticality Judgement Task (GJT) - which tested obligatory linguistic rules rather than variation or variability - declined as age of acquisition increased. The notion of a critical period has since been recontextualised as more of a continuum rather than a hard boundary. McDonald (2000: 397) notes that, given results from many studies comparing age of acquisition with proficiency in speakers well into adulthood, there is evidence that age of acquisition continues to correlate with proficiency in learners even over the age of 20. Furthermore, these formal SLA approaches do not take into account post-maturational factors (discussed further in Section 2.5.2) which may boost a learner's linguistic proficiency, and their acquisition of 'advanced' aspects of language like sociolinguistic variation.

Formal SLA studies have traditionally focused on speakers' acquisition of obligatory linguistic forms and overall competence (eg. Mougeon, et al., 2004). As will be explored further in this section, there is also ample evidence to motivate

exploring the language use of comparatively more proficient speakers and their sociolinguistic competence, as this can give us additional information about the extent of the acquisition process. Drawing from traditional SLA research, however, we can better inform variationist analysis by taking into consideration two key factors and whether they influence non-L1 learners' acquisition of variation: degree of linguistic proficiency and degree of formal instruction in the language. *Comprehensible input* is defined by Krashen (1985) as occurring in either a formal classroom environment or an informal 'real-world' environment, however, he also brings into question the advantage that formal grammar instruction (as tested by, eg. Robinson, 1995; Regan, 1995) and correction of speech by an instructor have on acquisition of variation as, although both are common elements of formal language instruction, they do not aid language acquisition, or only do so peripherally (Krashen, 1994: 302). Robinson's (1995) findings suggest that the level of 'difficulty' of grammatical rules is something to take into consideration, as learners with formal instruction were found only to have a significant advantage for judging the grammaticality of 'easy' rules regarding the English article system, however they did not have an advantage over non-instructed learners for correctly judging more obscure grammatical rules. Krashen (1985) therefore argues that conscious learning occupies different areas of cognition to subconscious acquisition. This suggests that we can expect comprehensible input obtained from informal environments to have more relevance than formal instruction on the social and discourse-pragmatic contexts associated with sociolinguistic variation. Regarding overall linguistic proficiency, we may also expect this to influence acquisition of variation. Previous research has used formal methods to measure individuals' language proficiency, such as Orfitelli & Grüter (2013), who used the Versant English Speaking Test (Pearson, 2011) to measure participants' English proficiency according to four metrics: Sentence Mastery, Vocabulary, Fluency, and Pronunciation. By assessing Versant proficiency scores against age of arrival in the US, as well as length of exposure to English input, they found that both longer length of residence in the US (ie. length of exposure to English input) and earlier age of arrival were associated with higher English proficiency. Proficiency, in turn, has been associated with the notion of *acculturation* - ie. change in cultural orientation after an individual moves to the country associated with their target language; Jiang et al. (2009: 481) measured the proficiency levels of Chinese L2 English speakers in the US, and found higher levels of US acculturation to be associated with higher English speaking proficiency. This points us to the relevance of migrants' own attitudes and motivations and how these impact their language use, which will be discussed further in Section 2.5.2.

We also know that newcomers to a region are able to exploit salient socioindexical links within their new dialect (eg. Nycz, 2018) and that acquisition of grammatical variation is likely to be more complex compared to something comparatively more superficial such as lexical variation (Kerswill, 1996: 179). When it comes to migrant speakers, we know that migration has "profound sociolinguistic consequences" (Kerswill, 2006: 1), and that situations of migration result in language contact situations which can "create complex syntheses of features" (Cornips, 2014). We have prime motivation to investigate how members of migrant communities acquire variation; Knowledge of native speaker patterns of variation - known as 'Type 2' variation (Bayley & Regan, 2004) - is a requirement if a second language (L2) learner is to become fully proficient in the target language (if they so desire). Furthermore, traditional Second Language Acquisition (SLA) approaches have tended to prioritise variation in obligatory linguistic forms - ie. 'Type 1' variation - and variation in 'expert' language users at the sociolinguistic level has often been relatively overlooked (Howley, 2015: 68). Studies have assessed patterns of acquisition among younger migrants (eg. Ryan, 2021), however, adult migrants can also be sensitive to patterns of variation, and we know that some can acquire a greater breadth of native speaker-like variation than others (eg. Lybeck, 2002; Schleef et al.,

2011). As we have already seen, this can sometimes be seen to pattern with macro-social factors, for instance, Drummond's (2010: 219) findings, which show female Polish adults in Manchester acquiring native-like sociophonetic variation to a greater extent than their male counterparts.

Given that migrants have also often been excluded from consideration in community-oriented analyses (Bucholtz, 1999: 208), an insight into how community embeddedness influences these communities' knowledge of variation is a valuable contribution to our understanding. As such, in the following section, I will discuss the importance of incorporating comparison between more central and more peripheral members to our analyses of communities. Among the non-L1 migrant participants of the present study, this will be explored using the theory of *acculturation*, and among the LGBTQ+ participants, using a similar measure of LGBTQ+ Community Involvement. I will now turn to these meso-social measures of community embeddedness, and suggest indicators that are pertinent to factor into these measures.

2.5 Meso-social factors

Although social networks and Communities of Practice have received much attention in variationist work, some scholars have questioned how exactly we define a speech community, and what sorts of communities we should focus on (eg. Bender, 2001: 263). It is imperative to understand the patterning of language variation across communities within which speakers are embedded if we are to understand the mechanics of variation (Labov, 2001). It is expected that speakers orient linguistically to their sociolinguistic surroundings (eg. Labov, 1966; 1972), and the various communities and networks that speakers belong to make up the backdrop for this. There is extensive evidence that variation patterns with group affiliations and the relative degrees to which speakers are embedded within them. The "culturally constructed valances" (ie. "cultural expectations, norms, and expectations") that underpin community dynamics also bind community members together (Snell, 2018: 19), and it has been argued that, throughout daily interactions, it is *these* community dynamics that inform speakers' behaviours over more broad societal dynamics (Snell, 2018: 20). We can directly compare amongst different communities, particularly if one is embedded within another. For instance, following a "small-within-large" methodology (Tagliamonte, 2012: 356), Konnely & Tagliamonte (unpublished manuscript) conducted such an analysis, finding it to provide valuable nuance. This study has already been discussed in Section 2.3.4, but to briefly reiterate here, the LGBTQ+ speakers (in this case belonging to a Community of Practice) were found to diverge in their patterns of variation from the ambient Toronto speech community. The LGBTQ+ community resist ambient community gender norms, patterning homogeneously regardless of gender, or patterning independently from each other *and* from the ambient community, and the LGBTQ+ group also lead the ambient community in contexts of usage of the incoming *be like* quotative.

2.5.1.1 Community embeddedness

Beyond physical community ties, members of communities can be 'imaginarily' linked around a shared sense of identity and associated values (Anderson, 1986), for instance, national identity (eg. being Polish) or another characteristic, such as a minority identity (eg. being LGBTQ+). These 'imagined' dimensions of communities do not map onto a shared location in physical space and most members will likely never meet each other. These relatively abstract elements of community structures, therefore, challenge our perceptions of space as being about more than just physical proximity, but also intangible social and perceptual distance (Britain et al., 2020). Indeed, many non-physical factors are pertinent to

fostering a sense of community, including feeling a sense of membership, feeling a sense of having influence on the community, having a sense of integration and support, and a shared emotional and experiential connection (McMillan & Chavis, 1986).

This reconsideration of communities is increasingly relevant to today's world of increasing transnational mobility and digital communication, it is important for us to incorporate how community membership might play into members' use of variation. As I have discussed throughout this chapter, we know that identities are "continually constructed through interactional practices" (Snell, 2018: 13), however, these more abstract elements of community membership forces us to consider the ways in which we define 'practices', and whether we may incorporate the intangible shared connections between community members into our definition of this.

A good way of determining whether community membership does indeed influence members' use of language variation is through the consideration of community members' degrees of involvement or embeddedness with the communities in question. Milroy (1987) considered members' degrees of community (in this case, social network) embeddedness, through the use of a 'network strength scale' which took into account indicators such as membership to groups and clubs with fellow community members, and degree of socialising with fellow community members outside enforced workplace contexts. Milroy found significant positive correlation between membership to denser, multiplex networks within these communities, and speakers' use of nonstandard variants associated with said communities. Although this study focused on the differences between social networks, it does suggest interesting implications regarding the importance of level of embeddedness within a community structure to the patterning of variation.

2.5.2 Acculturation

In Section 2.4, I mentioned "maturational factors" (Jiang et al., 2009: 481) that can constrain acquisition. I now turn to the "*post*-maturational factor" (Jiang et al., 2009: 481) of *acculturation*, which is defined as a change in cultural orientation following contact with a new cultural environment (Berry, 1980). This is in contrast to *enculturation* (Miller, 2007), which refers to the level of adherence by an individual to their original cultural background. Acquisition of a language occurs in tandem with that of the associated culture, and that the process of acculturation positively affects acquisition (Jiang et al., 2009: 481; Masgoret & Gardener, 1999) and sociolinguistic competence (Schleef et al., 2011). Acculturation is an ongoing process throughout a migrant's lifetime (eg. Golden, 1988) and migrants can take many pathways, depending on their attitudes and motivations, such as avoiding integration (separatism); integrating into the target culture at the expense of their original culture (assimilation); or achieving a cultural balance (biculturalism) (eg. Farver et al., 2002).

Many studies (eg. Moore, 2004; Podesva, 2007; Nycz, 2018) highlight the role of individuals' motivations in their use of language variation, and this notion of speaker agency is equally important when it comes to acculturation. *Integrative motivation* - ie. migrants' identification with and attitudes towards integrating into the target culture's speech community - has been found to have an impact on overall attainment in that language (eg. Gardner & Lambert, 1959: 267; Masgoret & Gardner, 2003). This is likely because higher integration results in a larger degree of comprehensible input and, as Nycz (2018: 175) states, people talk like those around them. Comparatively more native-like variation has been found in speakers more embedded in networks where they are exposed to a higher degree of native speaker variation (Drummond, 2010: 219) as well as those with more social network connections, and who are more settled (Lybeck, 2002; Howley, 2015), potentially

due to the quality of interaction these more integrated speakers are exposed to. Since cultural immersion is associated with a sense of belonging and security, and the mitigation of psychological stressors such as culture shock, it has been suggested that successfully navigating this process "may be necessary for successful L2 learning" (Lee, 2001). Hammer (2017) supports this, finding that intention to remain in the country long-term is closely linked to internalisation of the L2 and its use in cognitive functions. It is plausible that speakers with higher integrative motivations have more reasons to want to engage in a larger range of cultural contexts (Schrauf, 2014: 27) as well as enact a range of discourse purposes - eg. clarifying, elaborating, exemplifying, etc. (Tagliamonte & D'Arcy, 2009: 76) - throughout their daily interactions, using sociolinguistic variation to do so. It is also important to consider that macro-level sociodemographic factors might interact with acculturation level and integrative motivation. For instance, Howley's (2015) more settled and open-networked participants tended to be female, and other literature also backs up this association between gender and integrative motivation (eg. Gardner & Lambert, 1972; Mori & Gobel, 2006).

These findings, as well as discussion by Schwartz et al. (2014) suggests that language is a significant enough domain of the acculturation process to warrant its own methods of measurement. Lopez-Class et al. (2011: 1557) warn against *only* assessing linguistic acculturation and using it as a proxy for overall acculturation, as this misses other key elements of acculturation: social network (ie. quality of interaction)-related, value-based, behavioural, attitudinal, and identity-related factors. Therefore, we must ensure to take into consideration the full range of acculturation indicators so as to achieve the greatest detail in our measurement of this phenomenon. To add a further aspect to our measurement of acculturation, we know that acculturation can affect language use relatively quickly even if, on paper, immersion seems to be happening quite slowly; Jiang et al. (2009: 488-9) find that, although participants had only been in the US for five years, and still maintained very strong Chinese cultural attachments, American acculturation still clearly patterned with higher English proficiency. In fact, Howley (2015) finds that those with stronger migrant identities use more native-like variation in Manchester English, and Lee (2001) argues that a positive attitude towards one's ethnic background is associated with successful target language acquisition. Given that acculturation does not only entail integration into the *target* culture, but also one's relation with the *original* culture, we must consider how to go about incorporating this bilinearity into our measurement of acculturation.

2.5.3 LGBTQ+ Community Involvement

'LGBTQ+' is more than just a category, and many scholars have defined the LGBTQ+ community as being associated with specific practices, values, and a sense of shared solidarity (eg. Winer, 2020) - aspects which I have argued throughout this chapter serve to reinforce aspects of identity and, hence, interact with linguistic variation. Queer linguistic approaches (eg. Motschenbacher, 2011) problematise the use of rigid identity labels due to their "symbolic boundaries" (Lamont & Molnár, 2002) excluding members - in the case of the LGBTQ+ label, this includes closeted individuals who are not heterosexual or cisgender, but might not outwardly identify with the label or associate with the community (Motschenbacher, 2011: 152). However, because I have shown throughout this chapter that it is an individuals' sense of agency and motivations through which they enact their memberships to communities, and which influences their use of variation, by the very nature of the perspective of the present study, I will be considering only LGBTQ+ individuals who

expressly identify with this label, and will be measuring community membership according to individuals' depth of orientation towards this community structure.

Much like embeddedness within a culture can be measured along a scale of acculturation, the LGBTQ+ community has also been theorised as having an 'imagined centre' (Winer, 2020), with individuals relatively more or less connected to the core of this community. Findings indicate that the bonds tying one to the LGBTQ+ community can provide resilience, strength, and mobilisation in resistance to anti-LGBTQ+ discrimination (eg. Ross, 2012). Relatively little work exists on the LGBTQ+ community and its implications for variation, but such shared values of solidarity can provide motivation to bolster the use of nonstandard variation (eg. Snell, 2018 - though, in this case, in local networks). Because of this, and the complex interplay of socio-cultural factors influencing LGBTQ+ migrants' acculturative pathways (and potentially, in turn, their language use), LGBTQ+ communities are a worthwhile focus for sociolinguistic research.

2.6 Morphosyntactic features

Section 1.2 outlined the importance of the linguistic variable to the study of linguistic variation. A variable has been defined as any element of language that alternates among two or more forms ('variants'), without altering the utterance's basic truth value, essentially equating to "different ways of saying the same thing" (Labov, 1964: 166). Seminal variationist research (eg. Labov, 1964; 1966) primarily focused on alternate phonetic realisations of phonemes, and how these pattern according to social factors. In the decades that followed, exactly which dimensions of language have the potential to undergo variation, and how such variation arises, has been the subject of much discussion within the field. As well as focusing on variation in the sound system, Labov, and many scholars since, have also turned to analysing how variation manifests itself in abstract, structural elements of language (ie. relating to the grammatical system), as systematic variation in the grammar demands further attention in variationist research. The following section will discuss the treatment of morphosyntax within variationist research, and will conclude by outlining the three types of morphosyntactic features under investigation in the present study.

Morphosyntax (a term combining *morphology* and *syntax*) encompasses any element of the underlying structure of word and morpheme ordering within and across languages, and accounts for structural relationships between grammatical elements of any magnitude. An example of this is the nonstandard variation we see in the auxiliary verbs *was* and *were* (eg. *You was* (were) *afraid* and *He were* (was) *afraid*). Importantly, Fasold (1996: 88-9) draws attention to the fact that morphosyntactic variation concerns deep underlying structural relations within the grammar, meaning that variant morphosyntactic forms are not simply one structure converted into another, but, instead, a re-arrangement of the same underlying information structure (Fasold, 1996: 88-9) - eg. the process governing the dislocation of elements originating within the matrix (ie. main) clause: "*For school, I read all the books ~~for school~~*".

While not a core aspect of this thesis, it is worth mentioning that this thesis has been informed by a Minimalist approach to morphosyntax. This approach takes the morphosyntactic system to be blind and autonomous (Kroch, 1989) but permits 'free', non-deterministic (ie. allowed but not obligatory) variation in structure or form (Adger, 2006: 527). Hence, this has been called morphosyntactic *variability* rather than variation (Adger, 2006: 504). Social or psychological factors can affect which options arising from this variability are selected by speakers, but these aspects are not built into the grammatical

system - they are secondary and epiphenomenal (Adger, 2006: 525). This variability, in turn, leaves speakers with options, the choice of which can be influenced by discourse or social factors, "some of which will have syntactic effects" (Adger & Smith, 2005: 164). Other approaches to grammatical variation have also been taken by variationist sociolinguists, such as theorising the existence of variable syntactic rules (eg. Labov, 1969).

When it comes to morphosyntactic variation, there is room for interpretation as to how we define and identify variants. Scholars have suggested that we can consider alternating forms to be morphosyntactic variants if they are either semantically vacuous (eg. function words) (Lavandera, 1978: 176), or the variation consists of changes in word or constituent order (Bender, 2001: 191). Furthermore, (socio)syntactic variant forms should be functionally equivalent (eg. Weiner & Labov, 1983: 33), but, because variants never have precisely identical meaning (Tagliamonte & D'Arcy, 2009: 74), subtle differences in meaning (Lavandera, 1978) or epistemic value (Tagliamonte, 2006: 76) can be permitted.

The extent to which sociolinguistic variation occurs within the morphosyntactic system has been under debate; Smith, Durham, and Fortune (2007: 91), for instance, find that in Buckie, Scotland, speakers avoid nonstandard phonological forms but are less aware of, and therefore, less avoidant of, use of nonstandard morphosyntactic variants. This suggests that morphosyntactic variation exists further below speakers' levels of awareness compared to other types of variation, and is argued by eg. Labov (1993) and Meyerhoff & Walker (2013) to be less socially stratified than less 'structural' variation. The argument here lies in the assertion that the morphosyntactic system receives less overt examination by speakers, resulting in less stratification. This has also been problematised, however (eg. Levon & Buchstaller, 2015: 322-3), as we have evidence of nonstandard morphosyntactic features (eg. variation in agreement morphology) functioning as linguistic shibboleths, and thus being avoided by speakers of certain macro-social (ie. demographic) categories. We have seen examples throughout this chapter that speakers can alter their frequency of use of variants based on their individual motivations (eg. Trudgill, 1986: 11), which brings into question the effect of speaker agency and self-awareness on the patterning of variation. Generative approaches have distinguished between the *i-language* - ie. the individual's cognitive interpretation of the grammar, and the *e-language* - ie. a hypothetical, though intangible construct of the speech community's grammar (Cornips, 2015; Bender, 2001: 264). Since i-languages are formed based on input from and interpretation of the e-language, this process relies on more than simply the individual's degree of linguistic competence, but also on factors pertaining to their lived experiences and the social contexts they inhabit (Bender, 2001: 263). Cheshire (1996) suggests that, because morphosyntactic forms are less frequent than, for instance, phonological ones, this means they may be less likely to come to form socioindexical links. Chambers (1995: 51), on the other hand, claims that "grammatical variables tend to mark social stratification more sharply [than phonological ones]."

Throughout this chapter, I have argued the importance of considering which variants become socially meaningful across the speaker populations under study. A key aspect of this has been through the consideration of a range of linguistic forms, with differing degrees of sociolinguistic vitality - a property of socially meaningful variants which can be purposefully exploited by speakers to enact communicative goals across several different levels of meaning. Although attention has historically been paid to the sociolinguistic patterning of morphological processes, scholars (eg. Moore, 2021) highlight the value of also incorporating morphosyntactic phenomena which are relatively more syntactic into variationist research. In the present study, I investigate three types of morphosyntactic constructions: argument movement (raising-to-subject with the

verb *seem*); optional discourse-based movement (topicalization, left dislocation, right dislocation); and nonstandard agreement (past-tense BE: nonstandard *were*, nonstandard *was*, nonstandard *weren't*). Many sociolinguistic studies of morphosyntactic variation primarily focus on highly socioindexical variants (those ideologically linked to salient social categories), but it is rarer to see comparison across variants of differing degrees of social salience (cf. Cornips & Corrigan, 2005: 87). Therefore, this research aims to expand knowledge of how social factors influence morphosyntactic variation across a range of types of morphosyntactic constructions. I will now discuss each of these types of morphosyntactic construction in turn.

2.6.1 Argument movement

The morphosyntactic processes underlying the argument movement construction considered in this study function to facilitate the underlying grammatical mechanism, and they carry relatively very minimal discourse-pragmatic or socioindexical import through instances of their use. Therefore, these types of features are unlikely to be constrained by social identity or discourse-related factors, and would be widely considered to be grammatical structures across British English varieties. This type of construction is acquired and can be used by all speakers of a language, and it is these types of features that have been the typical subject of interest to formal syntactic inquiry. The construction of interest to the present study is the process of raising-to subject with the verb *seem* in sentences such as "*John seems to like cake*". *Seem* does not pattern like most English verbs as it does not select for the subject *John*. Instead, the subject *John* is semantically linked to the verb phrase *to like cake*. This can be determined by the fact that *seem* can take an expletive pronominal subject - ie. one that exists to satisfy grammatical requirements but is not coreferential with any semantically meaningful entity (***It*** *seems John likes cake*), and also by the ungrammaticality of its pseudoclefted structure (**What John seems is to like cake* - compared to the grammaticality of *What John likes is cake*). In generative syntactic approaches, raising structures such as this motivate the theory of movement, whereby, in order to meet its semantic requirements, the subject *John* originally merges in specifier (ie. subject) position of the verbal predicate *like* within the embedded clause, and then remerges in specifier position of the matrix clause in order to fulfil the overall sentential grammatical requirements such as case assignment: *John seems to* ~~*John*~~ *like cake*.

The frequency of raising structures in English is something that is cross-linguistically marked (Givón, 2001), and the patterning of verbs like *seem* are rarely, if ever, explicitly taught, even to non-L1 acquirers of English. It is well documented that such raising structures are notoriously difficult for L1 English-speaking children to acquire (eg. Orfitelli, 2012; Choe & Deen, 2016), and for adult L2 learners of English, too (Callies, 2008: 201). Callies found Polish-born L2 English speakers struggled with target-like use of raising constructions, therefore, this feature is a suitable proxy for benchmarking advanced learners' acquisition of English.

2.6.2 Optional discourse-based movement

The three optional discourse-based movement constructions selected for the current study are right dislocation, topicalisation, and left dislocation. Right dislocation consists of a clause followed by a tag[3] which can be a noun phrase or a personal or demonstrative pronoun (Snell, 2018: 10) - eg. "*I've not got an accent*, ***me***" (Moore, 2020). Topicalisation, or topic-bearing structures - eg. "***Bill***, *I can't stand*" (Greenberg, 1984: 283) - involve the fronting or isolation of a particular element of the morphosyntactic structure - usually a noun phrase or prepositional phrase, and also possible, though less naturalistic, a verb phrase - as the topic of the sentence (Fillmore, 1968: 57). Left dislocation - eg. "***The people***, *they got nothing to eat*" (Mesthrie, 1996: 265) - involves the fronting of a noun phrase, or noun phrase plus prepositional phrase (Jespersen, 1928: 71), which also co-refers with an antecedent in the matrix clause (underlined). Before exploring the patterning of optional discourse-based movement constructions across British Englishes, I note that, in the L1 of the migrant participants of this study, optional discourse-based movement constructions such as topic dislocation in the left periphery also occur (eg. Mokrosz, 2022), though other constructions do not occur in Polish as they do in English (such as right dislocation). However, because acceptance of optional discourse-based movement in migrants' Polish L1 is beyond the scope of the present study, the English language will be focused on throughout the remainder of this section.

Certain variants are selected by speakers to aid in achieving discourse-pragmatic aims, for example, to position oneself relative to the interlocutor and facilitate interactional flow (Giles & Powesland, 1975), or perform functions such as clarification, elaboration, exemplification, etc. (D'Arcy, 2007). In using optional discourse-based movement constructions, speakers draw upon not only the referential content of the construction, but may also additionally draw upon relevant socioindexical associations, as these types of features exist at the intersection of pragmatic, social, and grammatical function (Moore, 2020). Discursive choices are sometimes subject to existing socioindexical relations within the community - though only indirectly, according to Eckert (2018: 190) - with the discursive purposes of variants such as right dislocation with personal pronoun tags rooted in community identity and associated values of solidarity (eg. Snell, 2018: 19). The referential meaning and structure of optional discourse-based movement constructions can facilitate their functionality, as is the case with tag questions - eg. "*They're happy, aren't they?*" (Moore & Podesva, 2009: 456) - which Moore & Podesva argue, in part due to their right dislocated structure, are discourse-pragmatically conducive (ie. lending the interlocutor towards agreement). As well as right dislocation, the left periphery (or C-domain) of the morphosyntactic structure - ie. the functional projection that dominates the matrix clausal structure (D'Arcy, 2005: 75) - has been linked to discourse-pragmatic variation via other forms of dislocation such as topicalisation and left dislocation (Henry, 1995: 135). The fact that the C-domain allows for discourse-oriented variability has been demonstrated in variationist work, for instance, the left periphery has been argued to have been a point of entry for incoming discourse variants, such as the discourse marker *like* (D'Arcy, 2008). Variation within the left periphery is linked to the illocutionary force of the utterance - ie. the speaker's intention (Cornips & Corrigan, 2005: 22), and also with emphasis (Macaulay, 1989; Cornips & Corrigan, 2005: 97), and with affective meaning (Cheshire, 2005; Timmis, 2010). Right dislocated tags perform discourse management functions of emphasis, clarification,

[3] Right dislocated tags have also been referred to in the literature as *amplificatory tags* (Quirk, 1985), *emphatic tags* (Petyt, 1985), and *tails* (Timmis, 2010), among other names. See Snell (2018: 10) for an overview of this.

or focus (Snell, 2018: 10). The nature of the tag element plays a role in its discourse purpose. For instance, noun phrase tags perform a clarifying function (Huddleston and Pullum, 2002: 1411–1412), but where the dislocated tag is a personal pronoun, these carry interpersonal functions (Carter & McCarthy, 1995: 151), conveying affective or attitudinal information (Snell, 2018: 10), or performing evaluative functions (Timmis, 2010: 11). Left dislocation has been argued to perform as a focusing construction (Gough, 1986) and a way of emphasising the sentential subject (Piatt et al., 1984: 120), and is considered to be a syntax-discourse interface phenomenon (Yuan & Dugarova, 2012). Therefore, all three of these optional discourse-based movement constructions have been argued in the literature to primarily serve as a means to facilitate discourse between interlocutors.

Use of these variants has been found to pattern with macro-social categories. In Ayrshire English, Macaulay (1989, 1991) found right dislocation with pronoun tags, noun phrase topicalisation (fronting), and left dislocation patterning with speakers of working-class socioeconomic status (but not their middle-class counterparts), and particularly in personal pronoun constructions, whose function is argued to convey intensity. Right dislocation is associated with informal speech styles (Quirk, 1985) and is a common feature of non-standard English (Huddleston & Pullum, 2002: 1408), specifically in spoken language (e.g. Carter & McCarthy, 1995: 150). Snell (2018: 10) notes that right dislocated personal pronoun tags are especially absent from standard English but more frequent in Northern British-Englishes (eg. Moore & Snell, 2011). As Macaulay found, these types of tags are associated with working-class speakers (Moore & Snell, 2011; Snell, 2018) and communities which orient towards working-class practice (Moore, 2020). Mesthrie (1996: 266), notes that topicalisation has historically been suppressed in formal and standard repertoires, and that it is not unexpected for these structures to persist in informal, or nonstandard dialectal speech, even in L2 speakers. Others have also suggested that the distribution of some optional discourse-based movement constructions, such as right dislocation, additionally patterns with other macro-social factors such as gender and age (Durham, 2011).

We have seen that the use of optional discourse-based movement constructions draws upon not only the referential meaning of the construction but can also additionally draw on socioindexical links to, for instance, macro-social categories such as social class. Overall, however, optional discourse-based movement constructions are nonetheless usually able to be acquired across a range of communities and most linguistic varieties. Their usage hinges on interpersonal motivations in communication, and an awareness of interactional norms and behaviours.

2.6.3 Nonstandard agreement

Our final category is that of nonstandard agreement constructions, specifically nonstandard past-tense BE agreement: Nonstandard *were* - eg. "*My Dad **were** up there*", nonstandard *was*, eg. - "*We **was** living there*", and nonstandard *weren't* - eg. "*It **weren't** very satisfactory*" (Tagliamonte, 1998: 155). There are notable differences between nonstandard past-tense BE agreement and optional discourse-based movement. In using these variants, speakers draw upon the referential content of these constructions *and* these variants can also be drawn upon as part of fulfilling particular discourse-pragmatic functions, much like for the optional discourse-based movement constructions. However, in addition to this, nonstandard past-tense BE constructions have particularly high levels of sociolinguistic vitality in British Englishes, and are socially meaningful within specific communities, which is something that can be exploited by speakers during communication to present themselves in particular ways. While optional discourse-based movement constructions facilitate interactional flow or

inter-speaker positionality (ie. the 'focus' of the variant is on the interactional context), nonstandard past-tense BE agreement features are more directly and saliently indexical of particular macro-social identity categories. Therefore, their use hinges on the speaker's perception of these indexical associations, and a desire to communicate these associations linguistically. Acceptance of such variants is often associated with some extent of embeddedness within, and identification with, a community within which the variant in question is socially meaningful. They are associated with specific (non-hegemonic) communities and identities and, as such, often violate prescriptive norms and can be met with pejorative judgments. For these reasons, it is proposed that these variants' acceptance is the most restricted of the three types of morphosyntactic constructions investigated here.

Nonstandard alternants of the *was/were* paradigm are socially salient variants in English and, hence, have received much attention in variationist literature, particularly in work on the sociolinguistics-syntax interface (eg. Adger, 2006; Adger & Smith, 2005; Cornips & Corrigan, 2005). In Minimalist Syntax, *was/were* variability is governed by a mental feature matching algorithm, and is derived from a syntactic relationship between the uninterpretable features of the verb forms (regarding person and number), and the interpretable features required by the subject (eg. Adger, 2006). Agreement between the features must be met for the structure to be a potential grammatical option for a variety of English to acquire, and the choice of the grammatical subject can condition which verb form(s) can pattern with it (eg. Hudson, 2007). This variability means that not all varieties of English pattern the same way with regards to *was/were* agreement, and, because agreement phenomena are an obligatory part of acquisition, the acquisition of these patterns is built into the process of acquisition of the variety in question. Some (eg. Moore, 2011: 354) have also suggested that non-standard *were* perhaps performs specific discourse-pragmatic functions, given that it has been found to occur infrequently with other variants that we know perform discourse-pragmatic functions, such as quotatives. Nonetheless, *was/were* variation has been found to pattern persistently and complexly with social factors, representing a high degree of sociolinguistic vitality and many different degrees of indexical meaning (Moore, 2011).

Before exploring the socioindexical links that these nonstandard agreement constructions have across British Englishes, I note that, in the L1 of the migrant participants of this study, the past-tense form of the Polish verb BE ('był') also undergoes agreement inflection for number (as well as person and gender), however, this verb does not have nonstandard equivalents across Polish varieties in the way that nonstandard past-tense BE agreement is found across varieties of English. Moreover, as the present study is concerned with acceptance of nonstandard past-tense BE agreement constructions by migrants' in their non-L1 English, the English language will be focused on throughout the remainder of this section.

2.6.3.1 Regional distribution

Although nonstandard *were* has been referred to as a "minority levelling option" (Wolfram & Sellers, 1999: 94), it is found across many varieties of English: It is historically more frequent in Northern Englishes, having been reported, for instance, in the northwest (Cheshire, Edwards, and Whittle, 1993: 71–72), Yorkshire (Wright, 1892), Lancashire (Ellis, 1869–1889), Bolton (Shorrocks, 1999: 168; Moore, 2011), but also in Norfolk, Suffolk (Ellis, 1869–1889), and the East Midlands (Britain, 2002: 23–24). Variation in the *was/were* paradigm has undergone some interesting splits across varieties of British English. Some varieties permit nonstandard *was* but not nonstandard *were* (Labov et al., 1968), and in existential contexts (eg. "*There* ***was*** *dogs at the park*"), nonstandard *was* even arguably exists in standardized varieties of English -

found, for instance, in London (Cheshire & Fox, 2009: 23). It has also been reported that some non-British Englishes are seeing a reuptake in standard *was* among younger people - for instance, New Zealand English (Hay & Schreier, 2004) and Australian English (Eisikovits, 1991). Other British English varieties permit nonstandard *were* but only with negative polarity (ie. nonstandard *weren't*), which has also been reported frequently in Northern Englishes (Cheshire, Edwards, & Whittle, 1993: 72). This study found that, in ten schools across the country, nonstandard *weren't* was permitted, but the positive polarity form of nonstandard *were* was not - for instance, in Reading, Berkshire (Cheshire, 1982). Some varieties permit both nonstandard *was* and nonstandard *weren't* - for instance, York English (Tagliamonte, 1998). Moore (2011) found an unusual pattern in Bolton, with speakers permitting both nonstandard *were* and nonstandard *weren't*, and argues that nonstandard *were* is also exhibiting revitalisation in its usage within certain communities of social practice.

2.6.3.2 Grammatical constraints

The patterning of *was/were* variation is conditioned by internal grammatical constraints, including the grammatical subject, clause type, and as already mentioned, the polarity of the structure (eg. Moore, 2011: 355; Tagliamonte & Baayen, 2012: 135). In terms of grammatical subject, nonstandard *were* occurs with 1psg *I*, in Bolton (Moore, 2011: 357, the East Midlands (Britain, 2002) and York (Tagliamonte, 1998: 179). It can be found with 3psg existential *it*, though, in Bolton, this is less common than with other 3psg pronouns *she/he* (Moore, 2011: 359). It is generally disfavoured with existential *there* (Moore, 2011: 350; Britain, 2002), perhaps influenced by the collocation of *there* + *was*. Moore (2011: 257) additionally finds nonstandard *were* to pattern very strongly with compound indefinite pronouns, such as *everyone*, perhaps due to their inherent plurality (Moore, 2003). No consistent effect of grammatical subject was found on non-standard *weren't* in London (Cheshire and Fox, 2009) but, in York, nonstandard *weren't* occurred more often with with 1psg *I* than with 3psg pronouns, as well as exhibiting increased frequency with existential constructions but decreased frequency with noun phrase subjects (Tagliamonte, 1998). Britain (2008) finds the opposite pattern in the East Midlands, with noun phrases patterning most strongly with nonstandard *weren't*, followed by pronouns. Nonstandard *was* has been found to pattern strongly with existential *there* followed by a plural noun phrase, for instance, in Bolton (Moore, 2003: 72-3) and in London (Cheshire and Fox, 2009: 16–17). It can occur with 1ppl *we*, 2psg *you*, 3psg *they*, and with plural noun phrase subjects but not 3ppl *they* (Adger & Smith, 2005: 154).

Regarding clausal structure, nonstandard *were* commonly patterns in interrogative (as opposed to declarative) structures (Moore, 2011: 356; Cheshire & Fox, 2009), especially within tags and, specifically, with *it* tags in York and London Englishes (Tagliamonte, 1998; Cheshire and Fox, 2009: 25). In fact, tag structures are a particularly significant factor on the use of nonstandard *weren't* (Tagliamonte, 1998; Cheshire and Fox, 2009). Nonstandard *were* have been found to be disfavoured in quotative constructions and in nonstandard constructions with progressive verbs such as *sit*, *stand*, etc. (eg. "*I were sat*"), as well as within regular wh- questions (Moore, 2011: 354-6). In London, nonstandard *was* has been found to occur more frequently in interrogatives (and, specifically, those with existential *there*) than in declaratives without existential *there* (Cheshire and Fox, 2009:16–17).

A significant and consistent polarity effect has been found on nonstandard *were* realisation across British Englishes, with negative contexts bolstering use of non-standard *were* - ie. *weren't* (Cheshire, 1982; Schilling-Estes & Wolfram, 1994; Moore, 2011: 353), especially negative tags (Tagliamonte, 1998: 177). *Was/weren't* levelling is the phenomenon in which

weren't is realised in negative constructions, and *was* in positive constructions, and, at more advanced stages of this levelling, *was* is always realised in positive constructions, whether standard or nonstandard (eg. Britain, 2002) - eg. both "*I* ***was***, ***weren't*** *I?*" and "*You* ***was***, ***weren't*** *you?*" (Beal, 2008: 382). Such extension of *was* and *weren't* to pattern according to polarity rather than person and number has been argued to constitute a realignment within the *was/were* paradigm (Schilling-Estes & Wolfram, 1994: 280). This pattern of levelling, and the different grammatical constraints on use of nonstandard *were* versus *weren't*, motivates the treatment of nonstandard *were* in negative constructions (ie. nonstandard *weren't*) as a separate variable for analysis.

2.6.3.3 Sociolinguistic variation

The variability in the English *was/were* paradigm gives rise to variation, and it is these variants that come to pattern socioindexically. *Was/were* variation has been found to be constrained by many combinations of macro-social factors, though no one factor can single-handedly account for the patterns of distribution (Moore, 2011: 347). As seen throughout this chapter, morphosyntactic variants can act as markers of identity (eg. Cornips, 2014: 2), and we know this is the case for nonstandard *were* (eg. Schilling-Estes & Wolfram, 1994: 287–8; Moore, 2011: 349). Nonstandard *were* patterns with socioeconomic factors, though not ubiquitously. Petyt (1985: 196) finds nonstandard *were* use to be confined to working-class speakers in West Yorkshire, however, in Bolton, the patterning of nonstandard *were* "largely transcends class boundaries" (Moore, 2011: 364). Nonstandard *was* is also more frequent among working class speakers (Feagin, 1979; Tagliamonte & Baayen, 2012: 138). Speakers of lower education status have been found to be more common users of both nonstandard *were* in York English (Tagliamonte, 1998: 178) - though with education only being influential to a minimal degree - and nonstandard *was* in Ottawa, Canada (Meechan & Foley, 1994). Relatedly, formality of the interactional context results in more probable use of standard *was* (Tagliamonte & Smith, 2000). These patterns are not surprising as *was/were* variation is strongly exposed to prescriptive judgement (eg. Tagliamonte & Smith, 1999), and the standard forms of *was* and *were* are overtly proscribed in educational settings.

Age is another key macro-social factor constraining *was/were* variation; In York, use of nonstandard *were* and nonstandard *weren't* has been found to pattern with younger, and also middle-aged (30-50 year old) speakers (Tagliamonte, 1998: 177). In the East Midlands, young people were found increasingly adopting patterns of *was/weren't* levelling (Britain, 2002: 17). Age also interacts with other social factors such as gender (eg. Tagliamonte & Baayen, 2012: 139) and ethnicity (Cheshire and Fox, 2009) in constraining *was/were* variation. However, Tagliamonte & Baayen do also note mixed results for patterning by gender (see Section 2.3.4 for discussion of *was/were* variation with regards to gender, and Section 2.3.5 on the interplay between macro-social factors). Gender has been argued to exhibit "a moderate ... effect", with nonstandard *were* and nonstandard *weren't* used more by men than women in York (Tagliamonte, 1998: 177), while, simultaneously, the spread of ongoing *was/weren't* levelling has been found to be female-lead (Tagliamonte, 1998; Cheshire and Fox, 2009: 23). In York, less educated speakers do not pattern significantly differently in their nonstandard *was* use by gender, however, more educated women use far more nonstandard *was* (especially in existential constructions) than their educated male peers (Tagliamonte, 1998: 183). Finally, and unsurprisingly, non-standard *were* correlates with factors related to regional identity, with Moore (2011) noting that parental place of birth plays a key role in speakers' high non-standard *were* use. This has interesting implications, particularly when examining the acceptance of morphosyntactic variation with regards to

acculturation level in non-L1 migrant speakers who may have exposure to multiple dialects and patterns of variability over their time living in England.

2.7 Conclusion

This chapter has outlined the theoretical groundwork that the present study builds upon. I have covered the sociolinguistic underpinnings which motivate this research, including how sociolinguistic theory has previously been applied to inform our understanding of the patterning of variation at different scales. I have discussed the significance of macro-social factors (eg. age, socioeconomic factors, region, gender, and sexuality), and how these factors interact with each other. I have also discussed community-oriented approaches to variation, and the types of communities relevant to the present study - at a national level, as well as with a focus on the LGBTQ+ community. I have also explored how level of embeddedness in these communities is relevant in the context of migrant language users. As part of this, I have overviewed how the acquisition of language variants by non-L1 speakers has been approached in traditional second language acquisition studies, and how we may expand our knowledge of advanced acquirers' variation with sociolinguistic approaches. I have also explored the framework of acculturation and the sociocultural relevance of the participant demographics chosen. Following this, I have discussed literature relating to morphosyntactic variation, outlining the morphosyntactic variants of interest in the present work, and why they have been selected as the subjects of analysis. Given all of these aspects that I have discussed so far, the present study is concerned with answering the following research questions:

2.7.1 Research questions

1. How do participants pattern in their acceptability judgement ratings of British-English morphosyntactic features, according to the type of morphosyntactic construction - argument movement; optional discourse-based movement; nonstandard agreement? How do Polish-born participants' acceptability judgements compare to those of English-born participants?
2. To what extent can factors relating to formal linguistic accounts (eg. degree of prior linguistic input; linguistic proficiency) capture the patterns of variation found in Polish-born participants' acceptability judgement ratings?
3. What more can we learn about participants' acceptance of morphosyntactic variants by considering macro-social factors (eg. age; gender; socioeconomic class)?
4. Do participants' acceptability judgement ratings pattern differently according to participants' British or Polish acculturation levels?
5. How does participants' acceptance of morphosyntactic variation vary according to participants' membership to, and level of embeddedness within, the LGBTQ+ community, compared with the overall Polish-born and English-born populations?

3. Methodology

3.1 Introduction

As discussed in Chapter 2, various factors have been reported to influence and constrain participants' acceptance of socially meaningful variation in a non-L1 language (ie. not their native language). The aim of the present study is to explore factors on participants' acceptance of morphosyntactic variation in a non-L1 with regards to three types of morphosyntactic construction, and to do so by considering the effects of macro-social (ie. demographic) factors; practical L2 acquisition constraints relating to degree of linguistic input and proficiency; and meso-social factors (relating to participants' degrees of embeddedness in communities). To recap from the previous chapter, the present study is concerned with answering the following research questions:

1. How do participants pattern in their acceptability judgement ratings of British-English morphosyntactic features, according to the type of morphosyntactic construction - argument movement; optional discourse-based movement; nonstandard agreement? How do Polish-born participants' acceptability judgements compare to those of English-born participants?
2. To what extent can factors relating to formal linguistic accounts (eg. age of onset; degree of prior linguistic input; linguistic proficiency) capture the patterns of variation found in Polish-born participants' acceptability judgement ratings?
3. What more can we learn about participants' acceptance of morphosyntactic variants by considering macro-social factors (eg. age; gender; socioeconomic class)?
4. Do participants' acceptability judgement ratings pattern differently according to participants' British or Polish acculturation levels?
5. How does participants' acceptance of morphosyntactic variation vary according to participants' membership to, and level of embeddedness within, the LGBTQ+ community, compared with the overall Polish-born and English-born populations?

In this chapter, I will discuss the three methodological elements that have been implemented to answer these research questions: (i) an Acceptability Judgement Task (Section 3.3); (ii) the Versant English Language Speaking Test (Section 3.4); and (iii) several sociological surveys (Section 3.5). I will also outline the participant sample characteristics and explain how participant recruitment was undertaken. Then, I explore the study design and justification for use of the given methods. I reflect on the pilot study process and changes made following this. Following this, I will explore how data was collected, cleaned and analysed, and how the statistical modelling procedure was conducted. Before I explore each of these elements, I first turn to an overview of the dependent (outcome) and independent (predictor) variables within this study.

3.2 Variables

The outcome (dependent) variable in this study is the ordinal value of acceptability rating, represented to participants through the proxy of *naturalness*. This variable is measured on a 7-point Likert-type scale ranging from -3

(*completely unnatural*) to 3 (*completely natural*), with the middling band of 0 equating to *neither natural nor unnatural*. Because this variable is ordinal in nature (discussed in Section 3.9), the raw numbers which form the bands of the scale are not a direct reflection of acceptability (for instance, because different participants may pattern differently with regards to the range of the scale used, or may have different interpretations of *naturalness*). Rather, we are interested in extracting the underlying variable (which we may describe as 'relative likelihood of acceptability'). This affects the type of statistical models that can be used for inferential analysis, which will be discussed in Section 3.9.1.

The focal predictor and main linguistic effect of interest in the present study is that of linguistic condition (ie. the different linguistic features being tested for), of which there is one control condition: Raising-to-subject with the verb *seem*; and 6 test conditions: Topicalization; Left dislocation; Right dislocation; Nonstandard *were*; Nonstandard *was*; and Nonstandard *weren't*. The two binary variables of birth country (*Polish-born* versus *English-born*) and LGBTQ+ status (*yes* versus *no*) are also very important to this study as, together, they capture the four participant groups that were controlled for in the participant sample (see Section 3.6), across the two types of community that are of interest. The meso-level (ie. community embeddedness) variables under investigation are those of acculturation level, ie. level of embeddedness within the respective culture (self-reported for Polish and English culture separately, measured using Acculturation surveys - see Section 3.5.2) and LGBTQ+ Community Involvement (also self-reported, measured using an LGBTQ+ Community Involvement survey - see Section 3.5.3).

Since this research focuses on non-L1 speakers of British English, in order to substantiate claims regarding any effects on AJT response, it is imperative to take into account possible effects resulting from English Language proficiency (measured using the Versant English Speaking Test - see Section 3.4) and years since English input onset. However, information relating to other factors such as Age of migration; Length of residence in England; Other languages spoken; and Type of prior English language instruction was also collected (all measured using the Demographic Information Form - see Section 3.5.4).

Macro-social demographic factors are also of key importance (see Section 2.3 for discussion of this). These have been measured using the Demographic Information Form (see Section 3.5.4), and the most important factors to the present study are: Age band; Gender; Region of residence; Socioeconomic status, measured using postcode and occupation; and Highest Education level.

3.3 Acceptability Judgement Task (AJT)

In the previous chapter, I argued the importance of considering which variants become socially meaningful across the speaker populations under study, and how this might happen. This has led to the investigation of features across three types of morphosyntactic construction: argument movement (raising-to-subject with the verb *seem*); optional discourse-based movement (topicalization, left dislocation, right dislocation); and nonstandard agreement (past-tense BE: nonstandard *were*, nonstandard *was*, nonstandard *weren't*). See Section 2.6 for a justification of the selection of these morphosyntactic features for this investigation.

Morphosyntactic features are notoriously hard to reliably capture in spontaneous speech (eg. Jamieson et al., in press: 3), even those which are considered acceptable in the standard language, and so in order to investigate participants'

attitudes towards the variants of interest, an Acceptability Judgement Task (AJT) was implemented. As it is already established that variability and variation characterise the features of interest in this study, the focus here is on how levels of community embeddedness pattern with respondents' attitudes and perceptions of nonstandard variation. The AJT methodology is especially beneficial for a study aiming to test a range of morphosyntactic variants, as it is fast-paced and user-friendly in a digital setting, allows for the testing of acceptance and *lack* of acceptance (cf Schütze and Sprouse, 2013), and not laboursome for participants. Bender (2001: 263) notes that i-languages are formed based on input from and interpretation of the e-language, therefore, i-languages are intimately linked to individuals' lived experiences and the social contexts they inhabit. I argue here that, when aggregated, information from individual acceptability judgements may inform us about general trends in community e-languages. In fact, Bross (2019: 7) stresses the importance of aggregating acceptability judgments to avoid relying on data skewed by individual effects.

The use of AJT methodologies has been problematised, particularly in testing respondents' intuitions around nonstandard morphosyntactic constructions. A caveat of this approach is that the introspective judgements provided by respondents, especially towards nonstandard constructions, may appear overtly negative, and yet the speakers may actually employ these forms in their own speech or, conversely, judgements towards morphosyntactic constructions may indicate acceptance, but these constructions might not necessarily appear in speakers' usage (Labov, 1996; as discussed in Jamieson et al., in press). This means that such measures of acceptance do not necessarily provide a direct indication of actual usage. Other factors considered by Jamieson et al. (in press) are the frequency and salience of the morphosyntactic constructions under investigation, in that nonstandard constructions which occur more frequently, or for which there is a higher degree of cultural awareness, and therefore, stigma, are likely to face harsher judgements. Taking into account these aspects of AJT methodologies, the aim of using an AJT in the present study is not to make direct claims about participants' perception of the grammaticality of the constructions under investigation, nor is the aim to determine patterns of respondents' actual usage of these constructions in their everyday speech. Instead, results from the AJT aim to provide an insight into respondents' attitudinal reactions and acceptance, or lack thereof, of these constructions. In the present study, AJT response results are used to compare patterns in acceptance according to respondents' membership to, and levels of embeddedness within, the communities under investigation. Therefore, even though it might be expected for the comparatively more nonstandard morphosyntactic constructions investigated in this study (nonstandard past-tense BE agreement) to be subject to harsher judgements than the comparatively more standard constructions (optional discourse-based movement), much of the focus of the present investigation will be on the *comparative* acceptance of these constructions between members of the communities under investigation, and according to the macro-social categories they belong to, rather than solely on the comparison between acceptance of the different types of constructions. Given the beneficial aspects of AJTs, in terms of their usefulness in eliciting respondents evaluations of morphosyntactic constructions which are otherwise very difficult to access (cf. Jamieson et al., in press), this type of methodology was chosen as most suitable for the purposes of the present research.

3.3.1 Design

This study was executed using Gorilla Behavioural Science Software (Anwyl-Irvine, 2019), as will be discussed in Section 3.7. In the present section, general details of the study design will be discussed. In the AJT, respondents were presented with a series of stimuli in a randomised order (following Greenbaum, 1977: 7), and asked to rate them in terms of

their *naturalness* (a proxy for acceptability). A 7-point AJT response scale provides more nuance than a binary yes/no measurement of acceptability and allows for a neutral option, but is more manageable than a wider 10-point scale (Bross 2019: 15).

Following studies such as Martin & Haroldson (1992: 523), the definition of *naturalness* was left up to participants' own interpretation. *Naturalness* was chosen over *acceptability* so as to be more intuitively interpretable to non-linguist participants. The Likert-type scale labels themselves were numerical, however, a key was provided on screen to clarify the 'naturalness' ratings equivalent to the scale increments. This was partially for formatting purposes (so the scale would be more compact and not run on to the next line given that participants' devices had screens of varying dimensions), but it also allowed for the omission of ambiguous intermediate scale labels such as *slightly more unnatural than natural*. The most extreme rating equivalents (*completely unnatural* and *completely natural*) were displayed on the leftmost and rightmost peripheries of the rating scale, respectively (as illustrated in Figure 3.1).

Adger & Trousdale (2007: 264-5) note that AJT designs may tend towards eliciting judgements informed by proscribed norms, resulting in reduced accuracy of respondents real-world language attitudes. To mitigate this, the AJT design in the present study was deliberately crafted in such a way so as to alleviate respondent bias towards prescriptive responses. The study was self-paced, but participants were instructed (though not forced) to spend no longer than five seconds making each judgement. This was in an effort to avoid the second-guessing of responses, and to obtain as true as possible a measure of participants 'gut feelings' towards these test items. Further instructions also explicitly directed participants to not judge the sentences based on their idea of prescriptively 'good' or 'bad' grammar. After selecting a rating for an item, participants were able to change their original answer before proceeding, in the event of a misclick; however, changing one's mind or spending more than five seconds on each rating was discouraged in the task instructions. After progressing to the next item, participants were unable to go back to change their ratings of previous items. This was in an effort to avoid the second-guessing of responses, and to avoid participant bias towards prescriptive norms.

A bar was displayed at the top of the test screen that enabled participants to track their progress throughout the task. The lack of enforced pacing was deliberate, in order to ensure that participants responded to every test item, and did not miss any items due to time constraints. This was especially important as the task was not being carried out in a controlled, observed lab setting, and quite often on phones and smaller devices with limited visibility or uncertain connectivity, so it was more likely that participants might need the sentence to be visible to them for longer. Individual response times were recorded as part of the metadata, making it possible to ensure that responses were generally provided within the suggested time frame of 5 seconds (see Section 3.8.3.1 for more detail).

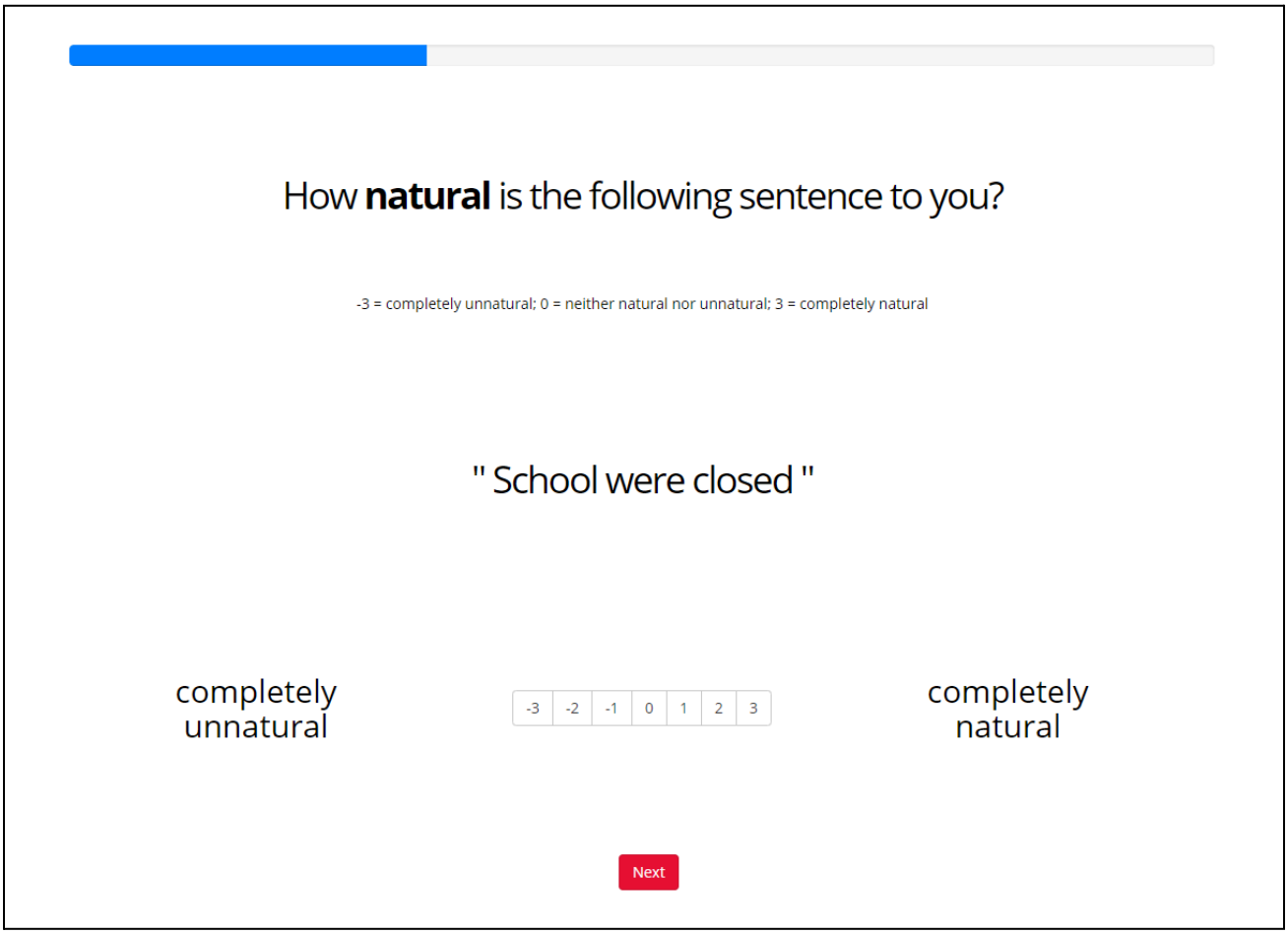


Figure 3.1. *Design of the Acceptability Judgement Task (AJT)*

Because many of the variants under investigation occur mostly in spoken as opposed to written language, participants were instructed to respond according to how they would feel had they heard the test items spoken aloud. Test items were presented in quotation marks to reflect this. A more conventional (Jamieson et al., in press) audio stimuli approach was considered in order to avoid any potential inherent unnaturalness in reacting to nonstandard variants (which are typically heard spoken aloud) presented in a written form, however, because some, but not all of the morphosyntactic constructions under investigation are regionally restricted, textual stimuli were opted for in order to avoid accent effects on ratings. Additionally, respondents were asked not to rate the test items on how natural they would be if *they* themselves were to say the utterances. This is because the focus is on respondents' *attitudes* towards the acceptability of the forms, and not necessarily their actual *usage* (especially given that self-report of usage is likely to be unreliable). Since the perceived acceptability of AJT test items has been shown to be affected by the contextual surroundings of the stimuli (Bever, 1970: 346-348) as well as the semantic content of the test sentence (Bross, 2019: 9), and these effects were undesirable in this study, test items were presented in isolation (ie. without prior context) and were formulated in such as way as to be plausible in their imagined contexts of use.

Typically, AJTs contain distractor/filler items to redirect test-takers' attention, so they are not aware of what the AJT is really testing for, and to provide controls to ensure the test is being completed properly. Filler items were chosen not to be used in the present study. This was decided for practical reasons, due to the additional length they would add to the

study which was already asking for a large amount of respondents' time, and because of the disparity and variety already inherent amongst the large number of morphosyntactic constructions under investigation. Inclusion of test items across the optional discourse-based movement constructions of left dislocation, right dislocation, and topicalisation, together with the positive control construction of argument movement (raising-to-subject with *seem*) meant that the nonstandard past-tense BE agreement constructions were counter-balanced amongst several other types of construction. The inclusion of negative control items meant that, for participants who did not consider nonstandard *was* and *were* to be acceptable, these would not be the only test items they found unacceptable. Therefore, the positive and negative control items also functioned to aid in distracting respondents, as well as to fulfil the function of providing control measures. This, and the fact that AJT test items were presented to participants in a randomised order was relied upon to distract participants from focusing on any one type of morphosyntactic construction under investigation. The suitability of this approach was confirmed during the piloting process, and it was ultimately decided that participants were sufficiently distracted from knowing what the AJT was testing for (see Section 3.7.2 for more on this).

A 'break' screen was incorporated at the half-way mark of the task which encouraged participants to rest for a moment, if needed, and also served to remind participants of key task instructions. Prior to the main task was a training round which had an identical premise and design to the main task, but consisted of a set of 6 training items in a randomised order. The purpose of this was to familiarise participants with the task, and to maximise the likelihood that, by the time participants began evaluating the test items, they were confident in how strict/lenient they were to be with their responses. These training items were designed to be of a similar nature to the items included in the main task, without containing any of the linguistic conditions tested in the actual AJT; The 6 training items consisted of: two monoclausal sentences considered grammatical across all varieties of English; two monoclausal sentences considered ungrammatical across all varieties of English (containing agreement errors and missing prepositions), and two monoclausal sentences containing nonstandard morphosyntactic features - one containing a negative concord structure, and the other containing non-standard plural *there's*.

3.3.2 AJT Test Items

The task consisted of a total of 64 test items (sentences), divided evenly across 7 testing conditions and 1 negative control condition. Each condition originally consisted of 8 items, though a number of these were excluded from the ultimate analysis due to high degrees of rating incongruity (see Appendix 9.1.1 for details).

Turning to the test items included within the three types of morphosyntactic construction investigated in the present study, the argument movement category, which was used as the positive control condition in the AJT, included raising structures with the verb *seem*. Because this morphosyntactic construction is widely considered acceptable across British-English varieties, it is a suitable proxy for benchmarking acceptability and acting as a reference point for participants' acceptance of optional discourse-based movement and nonstandard agreement. For consistency, only raising-to-subject sentences were included, and the verb (*seem*) was controlled for. Raising sentences with an experiencer - eg. *John seems* ***to Mary*** *to be happy* (Choe & Deen, 2015: 113) - were avoided, as were expletive subjects (only proper noun and 3psg subjects were included). A mixture of present and past tense forms of *seem* (ie. *seems*/*seemed*) were included, and a mixture of both

nonfinite embedded clauses (eg. *John seems* ***to be working***) as well as adjectival predicates (eg. *She seems* ***happy***) were included.

The optional discourse-based movement category contained three test conditions: Topicalisation, left dislocation, and right dislocation. The topicalisation condition contained declarative structures with a mixture of topicalised prepositional phrases (eg. ***For school***, *I read a lot of books*) and topicalised noun phrases (eg. ***Those rules***, *I don't agree with*). The left dislocation condition contained declarative structures with a mixture of dislocated pronouns (eg. *Him*), simple noun phrases (eg. *Coffee*) including some modified by adjectives (eg. *An honest politician*), and complex noun phrases containing preposition phrases (eg. *Those shoes with the red laces*). The right dislocation condition also contained declarative structures, with a variety of right dislocated noun phrases, including 1psg, 2psg, 3psg, and demonstrative pronouns, regular nouns (eg. *dogs*), proper nouns (eg. *Mary*), as well as a tense phrase (*he is*).

The nonstandard past-tense BE agreement category also contained three test conditions: Nonstandard *were*, nonstandard *was*, and nonstandard *weren't*. All three of these linguistic conditions contained a mixture of declarative and interrogative test items. The nonstandard *was* condition contained a range of subject types: simple singular noun phrases (eg. *schools*), 1ppl pronouns, and existential *there*. The nonstandard *were* and *weren't* conditions contained a range of subject types: 3psg pronouns, expletive *it* and existential *there*, and simple singular noun phrases (eg. *school*).

Additionally, in order to assess whether participants were completing the task accurately, the test stimuli also included a negative control condition (ie. as anchor values - Bross, 2019: 33) of universally unacceptable test items containing errors in agreement, case marking, or determiner placement. Items from the negative control condition were excluded from analysis, and a full list of these can be found in Appendix 9.1.1.

3.4 Versant English Speaking Test

English proficiency was formally measured using the commercially available Versant English Speaking Test (Pearson, 2011). This element of the study was only completed by Polish-born participants, who, following their completion of the main study, were contacted with details instructing them how to complete the test (see Versant Test instructions in Appendix 9.2.1).

This is a 15-20 minute-long test completed individually over the phone with an automated speech-recognition system. The test focuses mainly on oral proficiency, and involves a variety of exercises testing different linguistic skills, such as reading aloud, sentence repetition, comprehension exercises, sentence parsing and rearranging, and freely spoken responses to open questions. Proficiency, as measured by this test, is sub-categorised according to four metrics: Sentence Mastery, ie. a measure of syntactic processing ability and use of appropriate sentence and clause structures; Vocabulary, ie. familiarity with and ability to understand and use everyday forms and meanings; Fluency, ie. appropriate flow (ie. rhythm, timing and phrasing) of speech; and Pronunciation, ie. appropriate grasp of phonological processing within a sentence context. Upon successful completion of the test, a score report is generated, assigning the candidate a test score between 20 and 80. The score report breaks down the overall Versant test score and its four subscores, providing a verbal description of the candidate's overall proficiency and linguistic capabilities, including their ability to engage in native-paced conversation, the intelligibility of their pronunciation, and their ability to express information of different complexities. Also included is a

more detailed description of linguistic capabilities informed by candidates' responses to open questions, which explains the levels of English communication a typical candidate with the corresponding score should be able to engage in, divided into sections: listening, speech production, spoken interaction, language quality, and strategies & skills. The report also provides tips for candidates to improve their English proficiency, and also cross-references the overall Versant test score with other common scales of measurement of English proficiency including CEFR, TOEFL, and GCE (see sample score report in Appendix 9.2.2). The CEFR scale is relevant to the present study as it is used instead of raw Versant scores to report participants' English proficiency levels. The widely-used CEFR scale is a grouped 6-way categorical measure, categorising participants into *Basic*, *Intermediate*, and *Proficient* users of English, which, for the purposes of the statistical model implemented in the present study, can be better used to provide broad indication as to whether participants pattern differently across these proficiency levels. Additional detail about how the CEFR scale has been derived from Versant test measurements and used in the analysis is provided in Section 3.8.2. Copies of score reports were forwarded to participants following their participation in the study, for their own reference.

The Versant test has been implemented in previous non-L1 acquisition research (eg. Orfitelli & Grüter, 2013 - discussed in Section 2.4) and has many benefits. It has been argued (eg. Hulstijn, 2011: 245) that the core linguistic and processing speed skills that the test detects can be accurately and reliably measured using discrete-point methods (such as the Versant test). The test does not discriminate between what is considered 'good' pronunciation based on accent as the model is trained on many different world Englishes, though Van Moere (2012: 339) does suggest that accents which are less clear in terms of the targets the model is matching to will score lower on pronunciation. The results of the test are directly comparable among participants and account for a range of linguistic skills which reflect those used in everyday conversation (Van Moere, 2012: 339).

3.5 Sociological Surveys

The third mode of data collection utilised within this study was the use of sociological surveys. Participants completed these surveys directly after completing the AJT. To recap, several surveys were included within the study - an English Acculturation survey; a Polish Acculturation survey; and an LGBTQ+ Community Involvement survey. All participants completed an English Acculturation survey, and Polish-born participants additionally completed the (identical/equivalent) Polish Acculturation survey. These were used to calculate measures for English and Polish acculturation (ie. level of embeddedness within the respective culture). Additionally, any participants who had indicated that they belonged to the LGBTQ+ community also completed an LGBTQ+ Community Involvement survey, after the acculturation survey(s). This was used to calculate a measure of level of involvement with the LGBTQ+ community. Each individual survey will be explained in detail in this section but, first, a brief overview will be given of the general survey design implemented within this study.

3.5.1 General survey design

The Acculturation and LGBTQ+ Community Involvement surveys were similar in their technical execution. The use of self-reported Likert scale responses is a very frequently employed method in measuring acculturation (see Zane & Mak,

2003: Chapter 2 for an overview). As such, each survey question required a single-choice response on a 7-point Likert-type scale. Within each survey, respondents were asked a set of questions, each of which constituted an indicator towards its respective measure.

All survey questions were formulated in as concrete a way as possible to minimise the possibility of misinterpretation or ambiguity. Examples were provided where needed, especially to exemplify aspects of behaviours and practices that respondents might interpret differently than intended (eg. '*typical British traditions/customs*' being '*Christmas day, Boxing day, Bonfire night, etc.*', and '*LGBTQ+ digital spaces*' being LGBTQ+ '*forums, dating apps, online groups*'). Questions that required participants' subjective assessments were clarified as such; For instance, the question '*To what extent do you think your personal values align with typical British values?*' was followed up by the following note: *'(according to your personal interpretation of what 'British values' are)'*. Although literature on questionnaire construction (eg. Schleef, 2013: 48) argues the importance of keeping question items "short, simple, and natural-sounding", Schleef also notes the importance of avoiding biassing respondents' answers towards particular interpretations. In this case, it was necessary to test respondents' affective ties to (their interpretation of) the concept of 'British values', therefore, this exact wording was used for transparency, and examples were deliberately chosen not to be provided. As recommended by Schleef (2014: 48), questions that aimed to gauge participants' extent of engaging in particular practices were framed temporally rather than relying on participants' subjective interpretations; For example, the survey question '*Living in England, how often do you typically interact with [someone who was born in Britain], either in person or remotely?*' was assigned a response scale ranging from '*Never/ Less than once a year*' (1), to '*Several times a day*' (7), with intermediate values being '*Several times a month*' (3), '*Once a week*' (4), etc, rather than more vague increments such as '*sometimes*', or '*always*'. Because this research was carried out during the Covid-19 pandemic, prior to the surveys, a note was displayed that instructed participants to respond to questions as if under regular circumstances, so as to reflect participants' typical engagement in activities in ordinary life, when unrestricted by quarantine/lockdown. The specific indicators included within the Acculturation surveys are discussed in Section 3.5.2.

Upon completion of the surveys, participants were shown their results for the respective survey (see Appendices 9.3.3.3 and 9.4.3 for examples of these), divided into three bands - *High*, *Medium*, or *Low* acculturation/LGBTQ+ Community Involvement[4]. Polish-born participants completed *both* the British acculturation and Polish acculturation surveys before being shown their results for both measures alongside each other. The results pages for the acculturation surveys included a brief explanation that acculturation is a term used to describe the extent to which people adapt to the culture(s) they are exposed to, and that this study explores the types of acculturation undergone by Polish-born adults living in England. The results pages for the LGBTQ+ Community Involvement survey included a brief explanation that this research explores the types of acculturation that LGBTQ+ people undergo (and how this interacts with their language use), hence why an LGBTQ+ Community Involvement score has been calculated from their survey answers. These explanations were included in order to show participants the wider relevance of them having completed the various sociological surveys, and

[4] For the original purposes of providing participants with feedback, the three results bands were derived by evenly splitting the 7-point scale, whereby, a mean score of 2.5 or lower was deemed *Low*; between 2.5 and 5.5 was deemed *Medium*, and 5.5 or more was deemed *High*. For the purposes of analysis, however, these result bands were re-coded to ensure a more even distribution of participants across categories. See Section 3.8.1.1 for details on this procedure

what their results might be used for[5]. Below these overviews, participants' result band for that survey was displayed, as well as a list of several potential factors which could have contributed to them being placed in this particular band. For instance, a *high* English or Polish acculturation result yielded the following explanation: *This suggests that you very strongly affiliate yourself with social, psychological, and cultural markers associated with* [*English/Polish*] *culture.* Equivalent explanations were provided for the *medium* and *low* acculturation bands, amending the wording as appropriate (eg. Low acculturation yielded the following explanation: *This suggests that you do not affiliate very much with social, psychological, and cultural markers associated with* [*English/Polish*] *culture*). Several possible reasons for acculturation results were explored, including the personal significance participants likely place on their national identity, their level of interaction with multicultural (or non-English/Polish) social networks, and their level of association with sociopolitical aspects of English/Polish culture. Additionally, information was provided to the Polish-born participants about four different modes of acculturation: Integration - ie. high acculturation to both cultures (bicultural identity); Marginalisation - ie. Low acculturation to both cultures; Separation - ie. higher Polish acculturation than English; and Assimilation - ie. Lower Polish than English acculturation. This was included in order to provide participants with more information about the framework of acculturation, and the relevance it may have to their lived experiences. For LGBTQ+ participants, similar results pages were also provided based on scores from the LGBTQ+ Community Involvement survey. These were more brief than those for the acculturation survey, giving only a short explanation for the given result. For instance, a *high* LGBTQ+ Community Involvement result yielded the following: *This suggests that you strongly involve yourself with the LGBTQ+ community, and this is reflected in your behaviours, and also in your openness about your LGBTQ+ identity* whereas a *low* LGBTQ+ Community Involvement result yielded: *This suggests that you only slightly (or not at all) involve yourself with the LGBTQ+ community, and this is reflected in your behaviours. You might not be open about your LGBTQ+ identity with everyone in your life, or just prefer to involve yourself with other communities/groups and focus on other aspects of your life.*

3.5.2 Acculturation survey design

Participants' acculturation (ie. their level of embeddedness within the respective culture) was calculated using two acculturation surveys - one about English culture and one about Polish culture. To recap, all participant groups completed the English acculturation survey, and Polish-born participants additionally completed the Polish acculturation survey. The survey questions were identical (or as closely equivalent as possible) for both cultures, in order to ensure that scores would be directly comparable with each other, and both were conducted in English. In order to ascertain the extent to which individual domains of acculturation play into the relationship between acculturation and the acceptance (ie. perception) of variation, both acculturation surveys were broken down into three sections, based on three key acculturation indicators: (a) Language use; (b) Mindset (ie. a person's attitudes and sense of identity); and, (c) Lifestyle (ie. a person's behaviours and practices).

Acculturation has long been measured using self-reported surveys consisting of rating scales encompassing many indicators across several dimensions deemed significant to the phenomenon (eg. Berry, 1980). Scholars have designed specific

[5] After being shown this information about their English acculturation scores and, if applicable, Polish acculturation scores and LGBTQ+ community involvement scores, participants had the opportunity to rate how accurate they considered each of these results to be (also on a 7-point scale). In addition, an open text field was provided in which participants were invited to share any other feedback they may have about their ratings. However, due to practical constraints, this element of the study was not ultimately used in analysis

measures for acculturation among specific cultural groups (eg. Cuellar, Arnold, & Maldonado, 1995 for Mexican Americans; Cortes, Rogler & Malgady, 1994 for Puerto Rican Americans) as some indicators, such as orientation towards certain cultural values, are highly specific to the cultural context under investigation. Early studies theorised acculturation as unilinear, with changes occurring on a continuum between the culture of origin on one end and the target culture on the other (Miller, 2007: 120), however, given that people's orientations may be to both cultures, or neither, use of a single continuum has since been disputed. Instead, the bilinear interpretation of the framework of acculturation (eg. Berry, 1979) posits that acculturation should be thought of as occurring across two separate scales of measurement, thereby resulting in two separate factors to compare. Participants might have cultural ties with other cultures than those under investigation, therefore, this method allows for the acknowledgment of the duality (or multiplicity) of respondents' cultural backgrounds. Furthermore, Szapocznik et al. (1978) suggest a further split within each of these measures, by key sub-dimensions of acculturation such as behaviours and values. Providing equivalent indicators are used, results from acculturation measures can be comparable between the cultures under investigation, regardless of language of administration (Schwartz et al., 2014). Therefore, during the survey design stage of the present study, particular attention was given to the framing of the two cultures involved (English and Polish), resulting in the decision to create one survey per culture, rather than incorporating both cultures into the same survey. The surveys were set up such that the two cultural backgrounds were presented on separate continuums, generating two separate results that could be analysed in tandem, rather than framing the two cultures on separate poles of the same continuum, and treating them as inherently oppositional.

Certain indicators regularly arise as factoring in to the acculturation process, and different studies conceptualise these indicators as part of different domains of acculturation; Schwartz et al. (2010) distinguish between the domains of practices (behaviours), values (beliefs), and identifications (cultural solidarity); Domino & Acosta (1987) focus on cultural values; Cuellar & Gonzalez (1995) focus on beliefs and attitudes; Berry et al. (2006) focus on both values and attitudes; Szapocznik et al. (1978) and Birman & Tyler (1994) measure behavioural acculturation, and Birman & Trickett (2001) additionally incorporate the domains of language and identity, as well as behaviour; Phinney & Ong (2007) measure ethnic identity; Schwartz et al. (2012) additionally consider national identity; Kang (2006) and Szapocznik et al. (1980) note the significance of language use and cultural behaviours. This list is by no means exhaustive, but it serves to illustrate that the domains to consider when calculating acculturation, and even the criteria by which to group indicators into different domains, depends on the rationale of individual research interests.

In the present study, effort was made to tailor the survey methodology to the participant sample. For instance, the fact that Poland is largely ethnically homogenous means that many ethnicity indicators from methods such as the Ethnic Identity Scale (Umana-Taylor et al., 2004) were not appropriate here. After consulting and consolidating prior accounts of the measurement of acculturation, and assessing the numerous contributing domains and overlapping factors that have been observed to influence this process, three major domains were conceptualised to be most relevant to measuring the acculturation levels of the target participants in the present study. To reiterate from earlier, these are as follows: (a) Language use; (b) Mindset (ie. a person's attitudes and sense of identity); and, (c) Lifestyle (ie. a person's behaviours and practices). Consequently, the Polish and English acculturation surveys were each split into three sections, according to these three domains[6]. Participants viewed each section on a separate page of the survey, and could cycle between these. Each

[6] In both surveys, the three sections were entitled *Language use*, *Attitudes & Identity*, and *Lifestyle*, respectively.

domain consisted of several relevant indicators, as represented in Table 3.1. Each of these indicators constituted one item within the respective survey category. Each survey item consisted of a question, followed by a 7-point Likert scale answer field. The Likert scale labels varied according to what was most suitable for each survey item, though effort was made to retain as much consistency as possible. The polarity of the Likert scales remained consistent throughout - ie. the left periphery always denoted the most negative or minimal answer, and the right periphery always denoted the most positive or maximal answer.

Domain of acculturation	Sub-categories	Indicators
Language Use	Communicative function	*Workplace/daytime study*
		Household
		Interest group (hobby/regular activity)
		Peer group (main group of friends)
	Cognitive function	*Calculating/counting*
		Note-taking (for personal use/synthesising information/learning)
		Writing out a to-do list/personal action plan
		Writing a shopping list or informal reminder
Mindset (Attitudes & Identity)	Motivations	*Importance of being perceived as a native speaker*
		Importance of contact with native speakers
		Intended length of stay/potential future return
	Affirmation & belonging	*Sense of national identity*
		Sense of local community identity
		Alignment of values
		Sense of contentedness and 'fitting in' culturally
Lifestyle (Behaviours & Practices)	Social	*Frequency of contact with native speakers*
		Frequency of contact with local community
		Proportion of close personal friends that are native speakers
		Memberships to local groups/networks of native speakers which interact often
	Cultural	*Participation in national traditions/customs*
		Participation in typical regular cultural activities
		Consumption of cultural foods
		Consumption of cultural media (eg. TV, films, books, radio, etc.)

Table 3.1. *Acculturation survey indicators that were factored into measurement of acculturation level*

3.5.2.1 Language use

Linguistic factors play a significant role in an individuals' rate of acculturation (eg. Luna, et al., 2008), and language use can be considered a distinct domain of acculturation (Szapocznik et al., 1980; Kang, 2006; Guo et al., 2009). Although Schwartz et al. (2010) do define the use of language as a practice in itself, and therefore, this could perhaps have been included within the 'Behaviours and Practices' section of the acculturation survey, language use was deliberately

sectioned off in this survey for several reasons; Examining linguistic factors is crucial to answering the research questions of this study, and therefore, it was pertinent to consider a range of linguistic indicators to build up an overall picture of an individuals' language use. Secondly, the linguistic aspects considered within this section all pertained to cognitive or communicative functions of language (explained in detail below). These are more psychological, relating to inner mental processes, unlike the indicators within the Behaviours and Practices section, which mostly pertain to practical matters of individuals' lifestyles.

As such, the first section of the acculturation survey structure pertained to participants' language use. All survey items in this section were framed in terms of likelihood (ie. the Likert scale response options ranged from 1, *entirely unlikely*, to 7, *entirely likely*). This was chosen as the aim of this section was to ascertain the chance that participants feel they would use the particular language across various linguistic functions and contexts. The language use sections were identical between the English acculturation survey and the Polish acculturation survey, differentiated only by the language to which the questions pertained. For additional clarity, and because many participants had lived in countries other than England, it was clarified that the questions pertain to participants' use of language (either English or Polish) while living in England.

The language use section consisted of eight survey items, split into two sub-groups of four items each, and was divided according to two key areas of linguistic function, based on Hammer (2017). The first of these sub-groups pertained to indicators of communicative linguistic function, and probed participants' likelihood of using the respective language across the following settings: in a workplace or educational setting; in the household; with an interest group (such as a club or hobby activity); and with their peer group (main group of friends). The second of the two sub-groups pertained to indicators of cognitive linguistic function; The distinction between communicative and cognitive function has been made as measuring cognitive functions is crucial to understanding the extent of the internalisation of a language (eg. Pavlenko, 2014), and, hence, measuring linguistic acculturation on a deeper level. Previous research has identified several key linguistically-dependent cognitive domains, including mental calculation, information management, problem solving, and planning (Hammer, 2017). Given this, questions in this sub-group probed participants' likelihood of using the respective language when performing the following tasks: calculating/counting; note-taking (for personal use, synthesising information, or learning); making to-do lists or personal action plans; and writing shopping lists or informal reminders.

These indicators of linguistic cognitive function relate primarily to written language, rather than spoken. This is because, typically, spoken language more overtly fulfils communicative functions, indicators of which are already covered by the first of the two sub-groups of this section of the survey. Also, by their very nature, linguistic tasks that heavily engage a speaker's cognitive functions rather than communicative functions, for "internal, cognitive, and contemplative purposes" (Hammer, 2017: 73) which can be tangibly measured, tend to involve written language. Inner speech (ie. self-talk) was also considered for inclusion within the language use section, however, as Hammer (2017) puts it, this type of language is elusive in its nature. Furthermore, within this study, it was more pertinent to measure language use for social (communicative) purposes, and for information processing (cognitive) purposes.

3.5.2.2 Mindset (Attitudes & Identity)

The second section of the acculturation surveys grouped indicators relating to psychological factors - or, an individual's cultural mindset; ie. their attitudes, and sense of identity within that culture. The indicators within this section

consisted of factors relating to individuals' cultural motivations (ie. what they *want* from their external context), as well as factors relating to individuals' internal sense of affirmation and belonging within the culture. To account for cultural motivations, survey items measured the importance participants placed on interacting with, and being perceived as, members of that culture; as well as participants' intended length of residence in, or intended return to, the country relating to that cultural background. To account for affirmation and belonging, some survey items measured participants' sense of cultural identity (consideration of oneself as a member of that culture, ie. being 'English', rather than just someone who lives in England); sense of belonging within one's local community (as opposed to the country/culture as a whole); sense of alignment with that culture's typical values (according to individual interpretation of what these values are); and sense of contentedness and 'fitting in' within that culture. All of the indicators within this section aimed to probe participants' own feelings and interpretations about these aspects, rather than obtain any objective truth about their level of 'Englishness', etc.

3.5.2.3 Lifestyle (Behaviours & practices)

The third and final section of the acculturation surveys measured tangible aspects of participants' lifestyles in relation to their cultural orientation, ie. the behaviours and practices they carry out during their lived experiences. This section measured social aspects of participants' behaviours and practices. This included their frequency of contact with members of the cultural background being assessed (specifically, with native speakers of the respective language), as well as, more specifically, with members of the local community the participant currently lives within or has lived in for the longest extent of time (ie. when they lived in Poland). These aspects included digital interaction as well as physical. Participants also provided details about the proportion of their close personal friends who are native speakers of the respective language, and about their memberships to local groups or networks which meet frequently, such as sports teams, hobby groups, etc. Cultural behaviours and practices were also measured, including participation in typical traditions or customs associated with the respective culture, cultural behaviours or regular activities that are popular in the respective country (eg. British pub culture). Finally, consumption of traditional cultural foods was measured in this section, along with consumption of media from the respective country (eg. books, TV shows, films, etc.)

3.5.3 LGBTQ+ Community Involvement Survey design

The final sociological survey implemented within the present study was the LGBTQ+ Community Involvement survey. This survey consisted of a similar format to the acculturation surveys, with a series of 7-point Likert scale responses ranging from *entirely unlikely* to *entirely likely*. The aim of this survey was to ascertain participants' level of embeddedness within the LGBTQ+ community, and this was measured according to participants' self-reported responses regarding a range of aspects of LGBTQ+ community life. Information was sought about the likelihood of participants interacting with other LGBTQ+ people; the likelihood of them attending national or local LGBTQ+ pride events or meetups; their likelihood of accessing digital LGBTQ+ spaces such as forums, apps or social media groups; the frequency at which they visit LGBTQ+-associated venues, such as nightclubs or bars; and likelihood of seeking LGBTQ+-related support, through support groups or therapy. Participants were also asked how likely they are to consume media aimed at an LGBTQ+ audience (including podcasts, music, films, books, etc.).

The main element of the LGBTQ+ Community Involvement survey was identical for Polish-born and English-born participants. However, some additional elements were included in this survey which were not factored into the final measure of LGBTQ+ Community Involvement, but instead used to provide additional context to the main Community Involvement results. These additional elements varied for English and Polish respondents. Their purpose was to probe participants' attitudes towards LGBTQ+ people within the networks they frequent, and to ascertain whether they believe cultural background plays into this. Hence, these additional questions asked about the likelihood that participants would feel comfortable explicitly disclosing their LGBTQ+ identity to friends and to strangers or acquaintances. For Polish-born participants, these questions were asked both with regards to fellow Poles as well as British/English people, while, for English-born participants, these questions were asked with regards to fellow English/British people as well as non-English people.

In order to collect additional detail specifically about Polish-born participants' linguistic orientation towards LGBTQ+ identity, two extra questions were included in the version of the LGBTQ+ Community Involvement survey for Polish-born participants about their likelihood of using the English and the Polish languages, respectively, when interacting with other LGBTQ+ people. These were measured similarly, on a 7-point Likert scale, ranging from *entirely unlikely* to *entirely likely*. It was specified that this could refer to any form of interaction, either in person or remote, and participants had the option to leave this section unanswered if they do not interact with other LGBTQ+ people.

3.5.4 Demographic Information Form design

Lastly, all participants completed a Demographic Information form to provide background context to their other data. This was a traditional, information-seeking questionnaire that probed participants' socio-economic background (using their postcode, highest education level and occupation as a proxy), as well as their linguistic background, if applicable (eg. age at English input onset; length of exposure to English input; language instruction; other languages known). Other metadata collected consisted of respondents' age band, gender, ethnicity, religion (or lack thereof); length of residency in England (and in other countries, if applicable); and age of migration to England (where applicable). To conclude the form, participants were given the option to express consent to be contacted for the purposes of future research, however, this was stressed as entirely optional and non-committal.

The information gathered here was required in order to consider the extent to which macro-social and L2 acquisition factors alone are able to capture participants' perceptions of variation, and in order to consider participants' motivations for variation. Furthermore, this information has made it possible to build up a richer picture of who the participants are, since it was not possible to directly interact with them. It serves to reinforce or nuance the findings from the sociological surveys about Acculturation and LGBTQ+ Community Involvement by directly probing into additional background factors which are likely to have an effect on these things; For instance, participants' biggest motivation(s) for moving to England, and their stances towards cultural phenomena such as Brexit and orientation towards religion that, based on Szulc (2019), further complexify Polish-born participants' motivations concerning moving to, and staying in, England.

To ascertain a general sense of participants' linguistic backgrounds, all participants were asked if they know any languages outside of English (and Polish if applicable), and were asked to list these. Additionally, Polish-born and

English-born participants received slightly different versions of the Demographic Information Form, with the Polish-born version of the form containing a couple of additional questions probing into factors concerning migration and acculturation. The first was their age of migration (ie. arrival to England), selected from several age bands, ranging from *6 or younger* to *81 or older*. These increments were not entirely evenly distributed, but were arranged with cut-off points at key life stages of the critical period (around age 12), adolescence (around age 18), and early adulthood (around age 24). Beyond age 30, increments were evenly distributed with a 10-year range per band. Methods of formal English language tuition undergone were also measured, with participants ticking all that apply. These were as follows: English language lessons as part of the school curriculum; English language courses at a language school; Self-motivated formal English language study (eg. through textbooks/online courses; I have had no formal English instruction. Length of exposure to the English language, length of residence in England (in years), and length of residence in current English region (also in years) was measured in increments ranging from *Less than 1 year* to *41 or more years*. Bands on the lower end of the scale had smaller ranges (ie. *1-2 years*, *3-5 years*), middling bands were in four-year ranges (ie. *6-10 years*), and bands on the higher end of the scale had 9-year ranges (ie. *21-30 years*). Additionally, in order to assess participants' mobility within England, an open text field was provided for participants to list any other English regions they had lived in for periods of more than a year (if applicable). In order to understand in more detail why participants migrated to England, they listed their main motivation(s) for doing so (with up to three options choices allowed). Based on Szulc (2019), these were: *For study/education*; *For work/employment*; *To join family/friends/partner already living in England*; *Moved with parent(s) as a child*; *Due to cultural differences between Poland and England*; *Other (please specify)*. Two questions were included to probe participants on key sociocultural aspects of relevance to Polish-born participants - Brexit approval and religious affiliation (see Section 1.3 for a discussion of the relevance of these). The survey was conducted four years after the Brexit referendum, and Brexit approval was measured on a 7-point likert scale ranging from *completely disapprove* to *completely approve* with the middle of the scale corresponding to *neutral/no opinion*. Similarly, level of religiousness was also measured on a 7-point likert scale ranging from *Not at all religious* to *Extremely religious*.

Labov (2011: 118) argues that the most accurate measure of socioeconomic status combines multiple factors such as occupation, education and house value. In the present study, socioeconomic background was measured using a combination of current UK postcode and current official job title. Unemployed participants were asked to provide the latest job role they had held, and students were asked to state their student status. Following, for instance, Snell (2018: 8), postcode information was used to measure participants' socioeconomic status using the UK government's Indices of Multiple Deprivation (GOV.UK, 2019). This index can be used to calculate a score on a 10-point scale based on information from seven domains of deprivation associated with the provided postcode: income, employment, health deprivation and disability, education skills and training, barriers to housing and services, crime, and the living environment. The measure takes into account 32,482 small areas across England, ranked relative to each other based on how deprived they are. Additionally, postcode and occupation were used to calculate a rank for participants' occupation, based on the National Statistics Socio-economic Classification (NS-SEC) metric of Analytic Class. Lastly, highest education qualification was also measured, with the following prompts: *Entry-level qualification*; *Secondary/high school (GCSE or equivalent)*; *Sixth form/college (AS/A-level or equivalent)*; *Vocational qualification (NVQ/apprenticeship or equivalent)*; *Foundation degree or equivalent*; *Bachelor's degree or equivalent*; Master's degree or equivalent; *Doctorate/PhD*; *Other (please specify)*.

Age band was measured, ranging from *18-24* to *81+*. Anticipating a greater number of younger participants, the below-30 age bands were split into two smaller bands (*18-24* and *25-30*), while the above-30 age bands had a wider range (ie. *31-40*). Gender was measured using the options: *Female*; *Male*; *Non-binary*; *Prefer not to say*; *Other (please specify)*. Response options for sexuality included: *Homosexual (gay or lesbian)*; *Heterosexual (straight)*; *Bisexual*; *Pansexual*; *Asexual/Aromantic*; *Prefer not to say*; *Other (please specify)*. Ethnicity was measured using the options provided by the ONS guidelines for measurement of ethnicity in questionnaires (ONS, 2021).

A couple of other factors were taken into account that may influence participants' AJT responses. Participants were asked in an open text field whether they had formally studied Linguistics as a science (as opposed to simply learning a language). Additionally, participants were able to self-report whether they consider themself to have a Specific Learning Difficulty or other disability that might affect their language skills (eg. dyslexia, ADHD, dysgraphia, etc.)

3.6 Participants and Recruitment

The present study focuses primarily on English- and Polish-born adults living in England, half of whom are members of the LGBTQ+ community. The participant sample was sorted into four groups, according to participants' country of birth (and native language) - Poland (and Polish) or England (and British English) - as well as their LGBTQ+ status. The following section outlines the recruitment process undertaken in this study, details about the sample characteristics sought, and the numbers of participants recruited across the four groups.

3.6.1 Recruitment

Participants were self-recruited, each opting to take part in the online study if they considered themselves to belong to one of the four participant groups (which will be outlined in Section 3.6.2). The aim was to recruit as many participants in each group as possible, up to a maximum target of 40 per group. Calls for participants were disseminated through various avenues, many of which were made more easily accessible due to the insider methodology employed in this study, such as having existing memberships to LGBTQ+ and Polish networks (more on insider methodology in Section 3.7.1.1). Different avenues were used to reach different participant groups, and recruitment adverts were tailored to appeal to different target groups, according to where the advert was posted, and its anticipated audience. In general, the interactive and quiz-like nature of the study was emphasised, as well as the potential to compare results with friends, in the hopes that this would appeal to potential participants. In all adverts, key information about the study was provided in English, given that the study itself was to be in English, however, when the study was advertised through primarily Polish-speaking channels, a short Polish introduction was also included.

3.6.1.1 Procedure

Starting from initial leads, 'snowball' (social network) sampling methods were implemented to gain a broader reach within the targeted communities. The demographic information form was very thorough, in order to be able to acknowledge the extent of sample biases at the analysis stage (more on the design of the Demographic Information form in Section 3.5.4).

Online spaces and, in particular, social media platforms, were heavily utilised during participant recruitment; This was the most appropriate method of recruiting participants, given that the study itself was to take place online. Also, internet-based recruitment was especially crucial due to the fact that data collection took place during Covid-19 lockdown, so there were no physical, face-to-face avenues available through which to recruit participants. Due to the pandemic, certain avenues that could have been utilised to reach potential participants were not available. Many of the Polish communities in England do not have a strong online presence, and it was therefore much harder to reach parents of pupils at the Polish Saturday school and members of the Polish community centre in Sheffield, as these networks did not have mailing lists I could access. However, being myself embedded in Polish networks, and having access to their resources, proved invaluable in boosting the dissemination of the study, as numerous contacts spread the call for participants to their own Polish-born contacts.

Online networks were utilised on Facebook, Twitter, LinkedIn, Discord, Reddit, and other websites. By far the most wide-reaching method for disseminating calls for participants was Twitter; Together, the two tweets posted about the study amassed around 30,000 impressions and around 1,500 engagements, and were shared and forwarded onwards. Mailing lists were also a key tool; General lists that were used to spread word of the study, and encourage recipients to forward the study on to relevant contacts, included: the University of Sheffield student research volunteers list (*myAnnounce*); White Rose College of Arts & Humanities' networks; and *JISCmail* lists whose members might be interested in the study (eg. *ESOL-RESEARCH*, *VAR-L*, *LING-ETHNOG*). Several University of Sheffield departmental resources were also utilised for the same purpose, including: the School of English's networks; the *LingLunch* network; and *The Centre for Linguistic Research*. Personal contacts, including family and friend networks, were also directly called upon, either to participate themselves (if they matched the target criteria), or to forward the study on to relevant people. Additionally, many, more targeted, avenues were utilised to search specifically for LGBTQ+, or Polish-born participants, as outlined below.

LGBTQ+ participants were directly sought using local and national LGBTQ+ networks, including: LGBTQ+ student societies at English universities; online profiles of popular LGBTQ+ venues; LGBTQ+ organisations (eg. *SAYiT*, *The Proud Trust*, *Birmingham LGBT*, and *Mosaic LGBT+ Young Persons' Trust*); LGBTQ+-oriented social media pages/groups; and mailing lists (eg. the *LGBT* JISCmail list).

Polish-born participants were targeted through extensive outreach to various Polish networks in England, with recruitment channels including: Polish student societies at English universities; Polish academic networks (eg. *Polonium Network*); social media groups/pages for Poles in the UK, both national and region/city-specific; Sheffield's Polish Saturday School; contacts at the University of Sheffield's Modern Language Teaching Centre, and School of Languages and Cultures; as well as migrant networks, and academic networks focusing on migration research, including: charities (eg. *Migration Yorkshire*); organisations (eg. *the3million*, and *Polish Migrants Organising for Change (POMOC)*); migration research networks (eg. *UCL BASEES Polish Migration network*, and *Sheffield Migration Research Group*); Polish advertisements websites of both an academic nature (eg. *Pol-Int: Polish-Studies*) and, also, of a non-academic nature (eg. *mojawyspa.co.uk* and *strefa.co.uk*); and migration-related JISCmail lists (eg. *LMRG*, and *MIGRATIONRESEARCHNETWORK*).

Some Polish-born LGBTQ+ participants were recruited using the resources described above. However, several, more targeted avenues were also utilised to find the most specific participant category in this study, including: adverts, group posts, and direct messages to users on Polish LGBTQ+ websites (eg. *queer.pl*); academic contacts whose work involves LGBTQ+

Poles; as well as social media groups/pages for LGBTQ+ Poles, either based specifically in the UK, or more broadly. Since social media has been deemed a highly relevant resource utilised by Polish-born LGBTQ migrants in the UK (Szulc, 2019), accessing social media channels was crucial to the recruitment of this participant group. As in Szulc's (2019) methodology, the present study found that recruiting LGBTQ+ Poles by directly contacting members of *queer.pl* was very effective. The website is the most established online network for LGBTQ+ Poles, and incorporates many useful functions, such as the ability to post advertisements, search its members, narrow search criteria according to members' country or area of residence, and sort the members list according to their last login to the website.

Finally, the study itself finished with an ending screen encouraging participants (who were already invested so much as to have taken part) to share the study with any contacts that may be interested, with embedded links provided allowing participants to share via email, Twitter, Facebook, or direct URL. Here, like in the recruitment materials, it was suggested that participants could share the study with their friends and compare their results with each other.

3.6.1.2 Ethics & Informed Consent

This study has undergone approval via The School of English's ethics review procedure according to guidelines from The University of Sheffield's Research Ethics Committee. The original study proposed included a mixture of in-person focus-group data collection as well as online questionnaires, however the study's methodology was adapted during the course of the project due to Covid-19 (see Section 3.7.1.2 for more on this), however, the original ethics confirmation was used although only the second element of the study was carried out. The ethics application and committee feedback can be viewed in full in Appendix 9.7.1.

The landing page of the study outlined key contact information for the researcher and supervisors, and a disclaimer that the project had been ethically approved. Prior to consenting to the study, participants were instructed to carefully consult an information sheet (see Appendix 9.7.2), which was linked to within the landing page. The sheet could be downloaded and kept for participants' own records if desired.

In the information sheet, prospective participants were made aware that the research concerns language variation and society. The project's specific aims were communicated, with an explanation that results from this research will contribute towards the understanding of the use of English language variation by speakers whose first language is not English and, more broadly, towards a better understanding of the complexities of language use. So as to minimise priming effects on participants' answers, specific linguistic elements of interest to the study were not disclosed to participants, and participants were not even told that the study is about non-standard language. It was outlined that participation is self-selected, and that up to several hundred participants may be involved. The length of time commitment required for the study was also explained - 5 minutes for a language exercise; 10-15 minutes for interactive social and cultural background quizzes; and, only applicable to Polish-born participants, 15 minutes (at a later time) for an online English Speaking Test. Desired participant demographics were outlined and a note was included describing that 'LGBTQ+' stands for Lesbian, Gay, Bisexual, Transgender, Queer, but is also an umbrella term inclusive of many other minority sexualities and gender identities. Due to the importance of reflexivity and positionality in research concerning minority communities (see Section 3.7.1.1 for more on this), a disclaimer was given about my own status as an English-born person with a bilingual and LGBTQ+ background. So that participants were aware of what information they would need to provide, details of required demographic

details were listed: postcode, job type, and education level for the estimation of socioeconomic background, and email addresses of Polish-born participants to enable their completion of the Versant English Speaking Test. Additionally, optional email contact regarding follow-up research was mentioned, and presented as entirely non-obligatory and non-committal.

Potential disadvantages to taking part were listed - namely, that participation would necessitate the volunteering of time. Participants were reassured that nothing disclosed would be negatively scrutinised and broad assumptions and stereotypes about the groups involved will not be made. Additionally, to minimise discomfort, sensitive questions (eg. regarding how participants label their sexuality) are optional. As well as benefiting the research presence of minority groups in the UK, the benefits to individuals participating were also addressed; The interactive nature, and topics relating to personal identity, make taking part in this study potentially interesting and enjoyable. Additionally, Polish-born participants also benefit through receipt of an official Score Report for the widely accredited Versant English Speaking Test, whose scores can be cross-referenced with other common language testing scales (eg. GSE and CEFR). Access to the test is normally $34 but free to participants of this study, and could be a useful qualification in employment purposes.

Participants were reassured of the pseudonymity of their data (enforced through use of alphanumeric participant IDs), and that no identifiable personal data will be included in the analysis and presentation of data. Participants were made aware that their pseudonymised data may be used in future research, the results of which may appear in publications, reports, web pages, etc, and may be shared with other authorised data processors. A detailed outline was provided of the legal and ethical elements of the study, such as the legal basis for processing of personal data, along with contact details for the lead researcher, project supervisors, and the Head of Department, in the event of any issues or complaints arising about the research.

This study utilised an inferred method of consent, with the consent information being outlined after participants had consulted the information sheet, but prior to them beginning the study. Participants were informed that, by submitting the survey, they were consenting to the following outlined aspects of the study, divided into 3 key areas:

- Consenting to take part in the project: Participants agreed to dedicate the outlined amount of time to the study, and the elements of the study were reiterated. Additionally, participants confirmed they had read the information sheet, had a chance to consult with the researcher if needed, and knew they could withdraw their data from use in the study before 01/12/2020 without giving any reason and without there being any negative consequences.
- Consenting to the ways in which information will be used during and after the project: Participants agreed to provide background information about their identity and language use. They acknowledged that identifiable personal details would be stored securely without disclosure to external parties, and destroyed within 5 years of the study's completion, and that their involvement will be strictly anonymous. They agreed for their pseudonymised data to be used in future research, publications, reports, web pages, etc., potentially by other authorised researchers.
- Consenting to assign copyright over any generated materials to the university: Participants agreed to revoke copyright over their data, so the information can legally be used in research.

As an additional measure to encourage participants to read the consent details, participants were required to input the current date at the bottom of the consent page, to confirm they had seen this page of the survey and not just clicked through without reading.

3.6.2 Participant Sample

As discussed earlier in Section 3.6, the present study focuses primarily on English- and Polish-born adults living in England, half of whom are members of the LGBTQ+ community. During the online survey, participants were asked to sort themselves into one of four groups by selecting an option from a drop-down list. Participants were grouped according to their country of birth (and native language) - Poland (and Polish) or England (and British English) - as well as their LGBTQ+ status.

The aim of this study was to recruit a total of 160 participants, divided into four equal groups of 40. This number was chosen as appropriate based on prior literature discussing the recruitment of participants for Acceptability Judgement Tasks (eg. Sprouse & Almeida, 2012: 26), which has identified a desirable sample size of approximately 37 participants to capture a medium sized effect. Given that this study focuses on distinguishing between quite broadly different morphosyntactic effects, such a sample size is sufficient to capture the effects of this distinction.

Ultimately, due to restrictions of time, and the need to exclude several participants, the ideal participant sample was not quite met. 40 participants were indeed recruited for three of the four groups (all but Polish-born non-LGBTQ+). However, the data from two of the Polish-born native speakers of Polish who identify as LGBTQ+ had to be discarded as they were not residents of England, and two others failed the negative control of the AJT (also discussed in Section 3.8.3). Therefore, the overall participant sample for this study was 155, divided across the four groups. Although not quite symmetrical, these participant groups are still reasonably well balanced, and are sufficiently large to meet the desirable sample size criteria outlined in prior literature. The four participant groups recruited for this study are provided in Table 3.2, together with a description of each group and the number of participants recruited.

Participant group	Description of group	Number of participants
English-born LGBTQ+	English-born native speakers of British English, living in England, who identify as LGBTQ+	40
English-born non-LGBTQ+	English-born native speakers of British English, living in England, who do not identify as LGBTQ+	40
Polish-born LGBTQ+	Polish-born native speakers of Polish, living in England, who identify as LGBTQ+	36
Polish-born non-LGBTQ+	Polish-born native speakers of Polish, living in England, who do not identify as LGBTQ+	39

Table 3.2. *Participant groups recruited in this study, and the number of participants in each*

3.6.3 Sample Characteristics

While recruiting participants for this study, a number of characteristics were controlled for, which are outlined in the following sections.

3.6.3.1 Country of birth

This study recruited a total of 77 participants who were native (L1) speakers of Polish. All of these participants had been born in Poland, and were living in England at the time of their participation in this research. 38 of these participants identified as LGBTQ+, and 39 did *not* identify as LGBTQ+. Participants were recruited from all over England, with no set quotas for different areas of the country. However, responses were clustered around certain areas of England - namely Greater London, and Yorkshire & the Humber (more on how this clustering was navigated in Section 3.8.1.1).

Participants were not excluded from the study on the basis of any other factors, such as their level of English language proficiency, age at which English was acquired, length of residence in England, or degree of exposure to English (though the study did account for these factors in the Demographic Information form). Potential participants were made aware that the study was about their English language use, and that the study would be entirely in English. Therefore, it was left to potential participants' own discretion to self-evaluate whether they felt comfortable enough in their English proficiency to participate. Participants' English language proficiency was later formally assessed in the second element of this study, using Versant English Language testing software (see Section 3.4).

Mirroring the two Polish-born participant groups, this study also recruited a total of 80 participants who were native (L1) speakers of British English. 40 of these participants identified as LGBTQ+, and 40 did *not* identify as LGBTQ+. These participants had been born in England, had been living there for the majority of their lives, and were living in England at the time of their participation in this study. It was made clear to English-born participants that they did not need to have any connection with Polish culture, though English-born people with Polish heritage were not excluded from the study. As such, generally, these two groups of participants had no connection to Polish culture, though several English-born participants reported having Polish heritage in the Demographic Information form.

3.6.3.2 LGBTQ+ status

Participants in two of the groups were members of the LGBTQ+ community. They self-identified as being LGBT, queer, or an otherwise non-normative gender or sexuality. The following message was given to participants when they were selecting whether or not they identified as LGBTQ+:

Note: *'LGBTQ+' is an acronym used to describe sexual orientations and gender identities. It stands for:* Lesbian, Gay, Bisexual, Transgender, Queer, *but is also an umbrella term inclusive of many other minority sexualities and gender identities. Please select from the above options depending on whether you identify within this category.*

Therefore, belonging to the LGBTQ+ community was a self-selected category, and the 'legitimacy' of participants' inclusion within this category is not something that was tested or imposed within the study. It was not compulsory for participants to disclose the specific criteria by which they are members of this community (though the Demographic

Information form did contain optional questions about participants' sexualities and gender identities, see Section 3.5.4). Given the 'non-Polish', 'ideological', and other controversial connotations that are often associated with the LGBTQ+ community in Polish culture, it is likely that some Poles who do have non-normative genders or sexualities nonetheless reject identifying as part of the LGBTQ+ community. Although it is possible to get around this by using less sociopolitically weighted labels to recruit participants, such as 'men who date men', this would not befit the aims of the current research. People who do not label themselves as LGBTQ+ most likely do not feel much sense of belonging within the LGBTQ+ community, and are therefore unlikely to share similar practices, attitudes, and identities associated with the LGBTQ+ community. The goal of this study is not to 'compare' the language use of people with non-normative genders or sexualities against people with normative ones (or any other group); It is not that LGBTQ+ Polish-born L2 speakers of English utilise morphosyntax in a way that is inherently unique to their status as LGBTQ+. Rather, it is their agency in identifying with this community and, by extension, the shared behaviours and attitudes associated with this community, that can, in turn, impact the ways in which members of the LGBTQ+ community (much like any other community), utilise language.

3.6.3.3 Age

All participants recruited in this study were adults (ie. over the age of 18). This is because adults are beyond the critical period (eg. Lenneberg, 1967), so their language acquisition is constrained by "age-related maturational factors", but still open to the effects of "post-maturational factors" such as acculturation (Jiang et al., 2009: 481).

Recruiting only adult participants was also preferable for ethical reasons. Given the prevalence of discrimination against LGBTQ+ people in society (particularly so in Poland), LGBTQ+ identity is considered a sensitive research topic. Although the present study does not directly invoke sensitive topics (eg. relating to discrimination or sexual activity), it does relate to participants' affiliation with a protected group, and, therefore, the judgement was made that adult participants would be better able to make an informed decision regarding their consent to take part in the research. Beyond the minimum age criterion of 18 years, this study did not control for age. This was partially due to the specificity of participant criteria required, which already greatly limited the pool of potential participants.

3.7 General study design

To recap, several research components were required for this study. This consisted of: (i) an Acceptability Judgement Task, to measure participants' acceptance of British-English morphosyntactic features; (ii) the Versant English Language Speaking Test, to provide a formal measure of Polish-born participants' English language proficiency; (iii) Sociological surveys, to measure participants' Polish and English Acculturation levels, and their LGBTQ+ Community Involvement levels, and to collect demographic information. With the exception of the Versant English Speaking Test, this study was executed using Gorilla Behavioural Science Software (Anwyl-Irvine, 2019). This software was chosen as it is intuitive to use, highly customizable, and supports a wide range of features. It is specifically designed for experiment-based methods and can incorporate many different elements within the same system using its in-built 'Experiment Builder', for instance, it can combine experimental tasks and surveys. Participant recruitment can be facilitated through the sharing of a

single link to the study, and the study can be implemented on multiple types of devices, such as computers and tablets[7]. Following data collection, Gorilla allows for the exporting of experimental data into csv format (either in long form, with one row per experimental trial, or short form, with one row per participant), generating one spreadsheet per survey or task.

The software consists of a flowchart-like design tool (see Appendix 9.6 for the full design of the present study), which can be used to control technical elements relating to how participants navigate the experiment, including branching nodes redirecting to different surveys or pages based on previously indicated responses. For instance, prior to the surveys, participants were required to select their participant group from a drop-down list. This meant that only participants who had indicated they were Polish-born were able to access the Polish Acculturation survey, and only participants who had indicated they were LGBTQ+ were able to answer the LGBTQ+ Community Involvement survey.

For each of the sociological surveys, there were separate results pages for each of the three results tiers (*High*, *Medium*, and *Low*) so that, upon survey completion, participants were redirected to a results page corresponding to their mean survey score. The aim of including results pages within the study was to give participants some feedback about their score and provide them with some insight into the study's content and analysis. Gamifying the sociological surveys in this way aimed to instil in participants a sense of reward for their participation, and make taking part in the study more enjoyable. These results pages, along with the full surveys employed in this study, can be seen in Appendices 9.3 - 9.4. Additional features implemented in the study design included quota nodes, which did not allow participants to complete the study if the participant group they indicated as belonging to had already reached the maximum capacity of participants.

The software saves participants' embedded data in real time, and allows for the creation of custom scripts to process this data during the course of participants' completion of the study. Using Gorilla's scripting widgets, overall scores were generated for British Acculturation level and, if applicable, Polish Acculturation level and LGBTQ+ Community Involvement level. The scripts were adapted from Gorilla's 'Big 5 Personality test' scripting walkthrough (www.gorilla.sc). For each of the factors measured, the implemented scripts calculated and stored the mean value across all responses. Based on these mean values, which had a possible range between 1 and 7 (matching the Likert scale response increments), participants were sorted into one of three ordinal categories for each measure - *High*, *Medium* or *Low*. Mean scores were stratified according to the following boundaries: Less than 2.5 was deemed '*low*'; between 2.5 and 5.5 was deemed '*medium*', and more than 5.5 was deemed '*high*'. It was according to these values that participants were redirected to the appropriate results pages following their completion of the surveys.

Gorilla was particularly useful for the AJT design (and an example of the participant-facing setup of the AJT can be seen in Figure 3.1). The task's interface bears similarity to that of a presentation slide, with Gorilla allowing for the implementation of features such as buttons, progress bars, a Likert-scale response bar, and social media sharing links. Stimuli for the AJT could be imported from spreadsheet format, and were coded for the section of the task they pertained to (ie. training items versus testing items) so they appeared in the correct portion of the task. Testing items were coded with the linguistic condition the item belonged to. Additionally, within the training and testing sets of items, randomisation by trial was implemented. This means that all the AJT testing items across all conditions were randomised together, and all participants responded to the same set of items, though in different orders.

[7] The software itself also allows access from mobile phones but this was blocked in the present study due to issues with scaling and readability of the AJT

3.7.1 Justification for methods

3.7.1.1 Insider methodology

There has been much discussion regarding whether a researcher can ever take up an entirely neutral position, regardless of how hard they strive to be objective, without being at least partially influenced by their views, attitudes, and experiences (eg. Erickson, 1973: 10). Although arguably all researchers approach data analysis with their own preconceptions and biases, it is important to briefly note how researcher identity is relevant to the present study, as an insider methodology was a core reason the relevant participant demographics were able to be accessed. I am a member of the LGBTQ+ community and I am also a second generation immigrant and heritage speaker of Polish who was raised within a Polish cultural setting. Growing up, I experienced othering based on the social categories 'Polish' and 'immigrant', as well as labels associated with non-normative sexualities and genders. Throughout participant recruitment, I was able to utilise my status and perspectives as an inside member of the communities under study. My personal background was highlighted during participant recruitment in order to make potential participants aware that the research was being conducted from a position that was embedded in their communit(ies). Additionally, when the study was advertised through avenues which were primarily Polish-speaking, a short introduction to the study was included, written in Polish. Conversely, it is also important to note that, although I share many identities and backgrounds with my participants, I was not born in Poland and am not a migrant. Therefore, although I am invested emotionally in the communities involved within this research, I do not align entirely with my participants' experiences.

3.7.1.2 Evolution of methodology

The methodology employed in this study evolved significantly throughout the course of the project, from the initial plan to carry out a series of focus group interviews, to creating an experimental online study. This was largely due to disruption caused by the Covid-19 pandemic; The planning and execution of the data collection for this research was carried out during periods of intermittent lockdown and quarantine, where face-to-face interactions with participant groups and individuals (which had originally been planned), were no longer possible. Access to campus was also restricted, and it was not possible to utilise the faculty's physical resources, such as the linguistics lab equipment. What was originally intended as a secondary complement to the ethnographically-informed focus-group element of the study, was expanded to constitute an online study in its own right.

Furthermore, the trajectory of the research itself evolved such that an experimental online methodology was deemed most appropriate to meet the aims of this research. In a prior iteration of this study's methodology, semi-ethnographic interview data collection was considered, and participants' Acculturation levels and LGBTQ+ Community Involvement levels were to be assessed during several focus group interactions. However, based on prior research measures of acculturation, structured questionnaires were ultimately deemed the most appropriate tools with which to measure these factors. Not only would ethnographic data collection not have been feasible within the time frame of this project, but it would have been significantly complicated by Covid-19. Also, more importantly, as this project unfolded and developed, it became clear that the research aims did not warrant hugely in-depth detail about how participants spend their time and what networks and

Communities of Practice (CofPs) they belong to. Some information about participants' networks and CofPs is touched upon within the acculturation surveys, and this is enough to broadly distinguish participants into 'high', 'medium', and 'low' categories for their acculturation level(s) and LGBTQ+ Community Involvement levels.

3.7.2 Pilot study

Prior to the main study, a pilot study was conducted with 13 English-born participants, and one Polish-born participant. The main aim of this was to assess whether test items within each condition patterned similarly to each other, as well as to assess whether practical elements of the study design functioned as intended. Subsequently, following the pilot study, several changes were made to the AJT.

The negative control test items (ie. universally unacceptable items containing errors in agreement, case marking, or determiner placement, listed in Appendix 9.1.1) patterned exactly as expected, with a mean rating of -2.86 (very near the *completely unnatural* end of the scale) and was therefore not changed. However, the argument movement (ie. raising) positive control condition, consisting of items expected to be accepted, received unexpectedly low acceptability, with a mean response of 1.76 of a maximum of 3. The raising condition had originally included both subject-to-subject raising (eg. *She seems happy*) and object-to-subject raising (eg. *Mary seems to John to be happy*), however, the latter construction was found to be skewing the mean rating so only subject-to-subject raising items were kept in this test condition.

The remaining 6 test conditions were plotted in terms of their mean scores per item in order to assess the spread of each condition (see Figure 3.2). Each coloured 'cube' in Figure 3.2 represents the mean AJT Response of one of the test items belonging to the linguistic condition of the corresponding colour. Comparing mean scores per test item in this way showed relatively good clustering for all linguistic conditions except topicalisation, which has a far wider distribution than other conditions. As such, several of the test items within the topicalisation condition were simplified to remove potential distracting elements following the pilot.

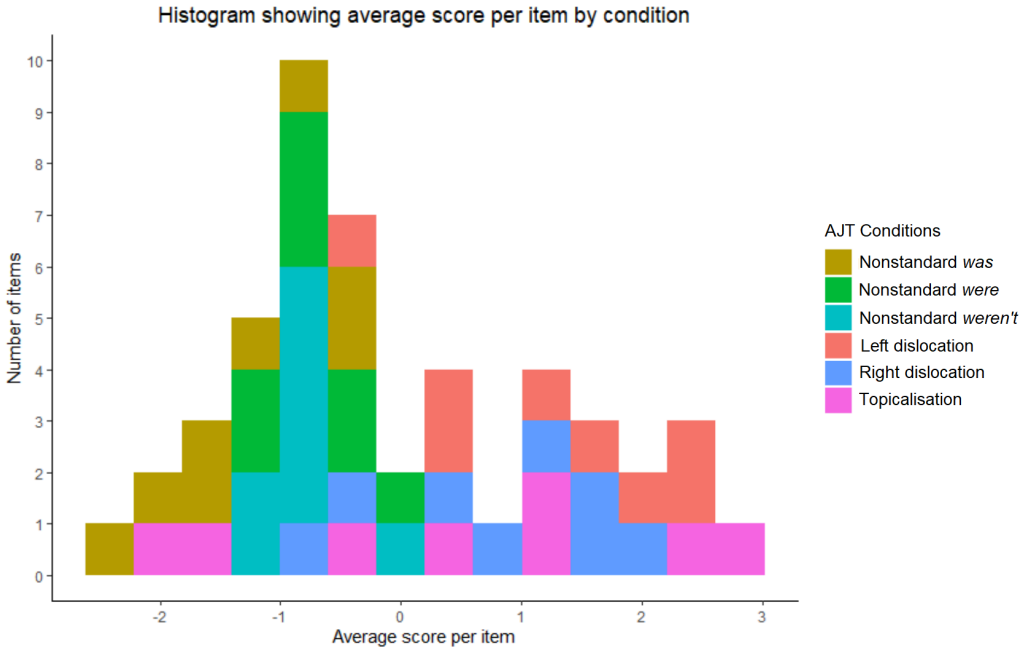


Figure 3.2*. Histogram showing distribution of mean scores per AJT test item across the 6 non-control test conditions*

Other changes to the AJT following the pilot include the addition of a training segment incorporated prior to the main set of testing items, to give participants a chance to acclimatise to the exercise and reduce any negative influence on the first few trials of the main task. Originally, the Raising condition included raising constructions with both the verbs *seem* and *believe*, however, for the sake of consistency, this was simplified to only contain *seem*.

The presentation of the AJT response scale was modified as a result of the piloting process: Originally, each increment of the scale had been labelled according to its corresponding degree of 'naturalness' (eg. *completely unnatural*, *very unnatural*, *slightly more unnatural than natural*, etc). However, based on pilot participant feedback, this resulted in the interface being very text-heavy and caused formatting issues for some participants whose devices had screens of varying dimensions. Furthermore, originally, the scale had included some intermediate labels which introduced potential ambiguity in their interpretation, such as *slightly more unnatural than natural*. Following piloting, the number of options on the response scale remained the same (7) but, for the sake of clarity, the labels of the scale were replaced with 7 numerical increment labels (ranging from -3 to 3), and 'naturalness' labels were only displayed for the most extreme ratings at either periphery of the scale (*completely unnatural* and *completely natural*). A key was additionally included above the rating scale, reiterating the naturalness labels that these peripheral scale increments corresponded to. This key additionally included a label for the middle increment of the scale - *neither natural nor unnatural*. The resulting scale resembled a hybrid between a Likert scale and a semantic differential scale - the latter of which is a type of scale where only the peripheries are marked by bipolar descriptors (in this case, *completely unnatural* and *completely natural*). Semantic differential scales have had utility in linguistic research, particularly in the measurement of participants' subjective assessments (eg. Garrett, 2010), and Dörnyei & Taguchi (2010: 32) suggest that such hybrids between numerical Likert scales and semantic differential scales can be beneficial, if the resulting interface allows participants to cognitively process and provide their responses with minimal distraction from potentially ambiguous intermediate labels (as was the case with some of the original labels like *slightly more unnatural than natural*).

The scale itself was changed from a sliding scale which could be clicked and dragged to the appropriate increment, to a numerical scale consisting of individual buttons and requiring an additional button to be clicked before the next test item would show. Finally, for additional clarity, quotation marks were added around test items to illustrate to participants that the sentences are expected to be interpreted as being spoken aloud. See Figure 3.1 in Section 3.3.1 for an example of how the scale appeared to participants in the final AJT design.

Finally, since the AJT employed no filler/distractor items, and the large variety of linguistic test conditions were instead used as counter-distractor items (see Section 3.3.1 for justification for this), pilot study participants were asked to provide feedback about what elements of language they believed the AJT to be probing, in order to ascertain whether participants were sufficiently distracted from the true purpose of the task. A couple of fellow linguists, who were familiar with the research, picked up on the inclusion of nonstandard *were* and *was*, but it was ultimately decided that, in general, participants were sufficiently distracted by the range of test items included in the AJT.

3.8 Data Processing & Analysis

The present study generated a large amount of data, resulting in a database with 11,131 rows and 163 columns. Through the process of data cleaning and the removal of unwanted data, such as that from participants who failed the negative control condition, a filtered form of the original database was created, with 8216 rows, and this was used for statistical analysis. The process of data cleaning and analysis is detailed below.

3.8.1 Database creation & cleaning

The database for this study was created using the data generated and downloaded using Gorilla Behavioural Science Software. Each element of the study (ie. each survey and the AJT) generated a separate spreadsheet in csv format. The AJT downloaded data formed the basis of the final database, and this was downloaded from Gorilla in long form (ie. with each trial of the task forming one row of the spreadsheet). R Studio's merge function (R Studio, 2020) was used to combine the long form data from the remaining study elements (ie. data regarding participants' Acculturation levels, LGBTQ+ Community Involvement level, and demographic information) into a single database. Unnecessary columns as well as rows containing negative control test items were removed, and additional columns were added to code participants according to key sample characteristics (LGBTQ+ status and birth country).

3.8.1.1 Re-coding variables

In order to be more easily visualised and interpreted, as well as to be processed by statistical models, several variables required re-coding. For instance, socioeconomic data required cross-referencing with government databases to calculate metrics based on participants' postcodes. Where warranted, several categories were collapsed with adjacent or similar categories for better clustering of data and statistical model performance. Participants who selected *Prefer not to say* for any of the demographic factors were coded for these factors as *Unspecified*. All factors under consideration within this study whose categories required further coding or alteration after the data collection process are outlined below, along with justifications for this re-coding process.

The gender factor was re-coded such that any participant whose gender was outside the binary of male and female was collapsed into a single category (*Not binary*). This is distinct from the specific label of *Nonbinary* (which most participants in this group identify with), as some participants had specified other labels in the *Other (please specify)* field, such as *Genderqueer*. It is important to acknowledge that these identities beyond the binary are distinct from each other and it is problematic to lump these together into an 'Other' category, but because there were very few participants with these *Not binary* labels, the decision was made to collapse these all into a single category in order to be able to take these participants into account in the statistical analysis. Furthermore, two transgender men specified their transgender status in the open text field instead of selecting *Male*. It is possible these participants were trying to provide additional information to aid in the research, rather than to entirely distinguish themselves from the *Male* category. These were collapsed into the *Male* category following consultation with transgender academics. In terms of the factor of sexuality, *Queer* was the most commonly added label within the *Other (please specify)* field, therefore, this was added to the list of sexuality categories. Regarding ethnicity,

one Polish-born participant specified *White Polish* within the *Other (please specify)* field, and was merged with the other Polish-born participants into the *White (other)* category. For formal English instruction, the *Other* forms of formal English instruction provided by participants were collapsed into an *Other* category. Other factors which had open text responses required more involved coding. Other languages known, for instance, were broadly divided according to level of knowledge of language (ie. beginner, intermediate, expert, or unspecified) as well as the number of other languages known (one other or multiple others). Formal study of linguistics was coded into a binary of *yes* versus *no*. Participants were sorted into one of four ordinal categories according to the highest level of education they have completed, adapted from the Census education classification levels (ONS, 2011). These were: *Higher Education postgraduate qualification*: Master's degree or equivalent, Doctorate/PhD; *Other Higher Education qualification*: eg. Bachelor's degree or equivalent, Foundation degree or equivalent, Certificate of Higher Education; *Further education qualification*: Sixth form/college (AS/A-level or equivalent), Vocational qualification (NVQ/apprenticeship or equivalent); *Secondary school qualification*: eg. GCSE or equivalent, Entry-level qualification. These were also later simplified further to create a binary factor for Higher-educated status (*yes* or *no*), which was ultimately used in statistical analysis.

Postcode information was cross-referenced with the Office for National Statistics' Open Geography Portal (ONS, 2022) data on broader geographical classifications such as county, and region. Based on this, participants' postcodes were coded according to the 9 region groups of England (*South West*; *South East*; *London*; *West Midlands*; *East Midlands*; *East*; *North West*; *Yorkshire & The Humber*; and *North East*). However, participant numbers were not evenly spread across these regions; 45 participants were in Yorkshire & the Humber, compared to only 1 in the North East. Therefore, these regions were reduced down further into only 3 categories: *North*, *Midlands*, and *South*. *South West*, *South East*, *London*, as well as *East* were grouped into the *South*. *East Midlands* and *West Midlands* were grouped into *Midlands*. Finally, *North East* and *North West* were grouped together along with *Yorkshire & the Humber* into *North*. This same 3-way region grouping was used to code for whether participants had lived outside of their current 3-way split region of residence. Postcode was also used to calculate Indices of Multiple Deprivation (IMD) Decile, a UK government socioeconomic predictor. IMD deciles were calculated using the Postcode Lookup tool (Swirrl IT Ltd., 2019). A list of postcodes contained within an imported csv can be automatically looked up and extracted back into csv format with the corresponding Lower-layer Super Output Area (a small geographical area) for each postcode, and the IMD data for that area.

3.8.1.2 Collapsing variables

This study controlled for the participant sample characteristics of LGBTQ+ status (LGBTQ+ and non-LGBTQ+) and country of birth (Polish-born and English-born), meaning participants were roughly evenly distributed across these two binary criteria. Participants' distribution across the remaining factors under consideration, however, had not been deliberately controlled and, therefore, had to be manually examined to ensure participants were distributed roughly evenly across the categories of each of these factors (eg. for age, each age band should contain enough participants that a statistical model would be able to capture the effect of this factor on the dependent variable of AJT Response). Given that these participant distributions were not something that was controlled for, for many of the factors, participants were rather unevenly distributed: For instance, in the case of age, participants who chose to take part in the study tended to be younger, which resulted in a skewed distribution of participants across age band, with comparatively fewer participants in the older

age bands. This type of skew can be an issue for inferential statistical analysis, when adding multiple predictors into a model, as some cross-sections of data will be 'missing'. For example, if we wanted to include the factors of birth-country, LGBTQ+ status, and IMD Decile (the socioeconomic status measure) in a statistical model, the model would not be able to work as there happen to be no data points for Polish-born LGBTQ+ participants in IMD Decile 6. A solution to this is to collapse down the categories of factors with many categories (such as IMD Decile, which originally had 10 categories) into fewer categories. Doing this might result in categories with uneven ranges; For instance, when a 10-category factor such as IMD Decile is collapsed into a 3-category factor, one of the collapsed categories must contain 4 of the original categories, while the others contain only 3. However, this is not an issue for ordinal logistic regression models, as these function on distinguishing between categories according to their order, and do not require the categories to have the same 'width'. In order to proceed with statistical modelling, each of the categorical factors relevant to the study were considered in turn and, where appropriate, their categories collapsed. While doing so, both the distribution of data as well as the underlying substance of the data being represented was taken into consideration in a balanced way. The factors that were chosen to be collapsed are discussed below.

The factor of IMD Decile (the socioeconomic status measure) originally contained 10 categories (numbered 1-10), but this was collapsed down into 3 categories - *Low, Medium* and *High*. Deciles 1, 2, and 3 were collapsed and labelled the *Low* IMD band; Deciles 4, 5, and 6 were collapsed and labelled the *Medium* IMD band, and Deciles 7, 8, 9, and 10 were collapsed and labelled the *High* IMD band. It was decided that the *High* band would be the one to contain four of the original IMD Decile categories because this resulted in three roughly evenly distributed IMD categories of 52, 50, and 50 participants, respectively. Similarly, the three highest participant age bands (*41-50*, *51-60*, and *61-80*) were collapsed together to form a single fourth band '*41+*' which helped to level out the distribution of participants such that *18-24* and *25-30* each contained 39 participants, *31-40* contained 48 participants, and *41+* contained 29 participants.

Polish participants' migration age bands were collapsed from five categories into three, retaining the key thresholds between adolescence, early adulthood, and adulthood, as discussed in Section 2.3.1: *7-18*, with 13 participants; *19-24* with 33 participants, and *25+* with 29 participants. Finally, English Language Exposure (in years) had very few participants in the *3-5 years* and *41+ years* categories, therefore, *3-5 years* was grouped with *6-10 years* into *3-10 years*, and *31-40 years* was grouped with *41+ years* into *31+ years*, resulting in a 5-way categorical split for this factor.

English Lifestyle, one of the three sub-categories of English acculturation which was ultimately decided to be used as a proxy for overall acculturation (see Section 3.9.2.3 for details), was manually re-categorised from the original categorisation that had been implemented during the survey itself. Originally, in the Polish or English acculturation score to result band conversion, a mean score ≤ 2.5 was deemed *Low*; > 2.5 and ≤ 5.5 was deemed *Medium*, and > 5.5 was deemed *High*. However, very few participants had English acculturation scores of 4 or lower meaning that the three categories were not well balanced. It was important to match the amended categories across both English and Polish Lifestyle (acculturation) to ensure the two remain comparable. To a lesser extent, it was also preferable to match the LGBTQ+ Community Involvement measure to the same parameters. Polish Lifestyle and LGBTQ+ Community Involvement also had very few participants scoring below 4, therefore, the following split was made across all three measures: Scores of ≤ 4 were coded as *Low*, scores > 4 and ≤ 5.5 were coded as *Medium*, and scores > 5.5 were coded as *High*.

Finally, I turn to the treatment of the formal measure of English language proficiency, attained using the Versant English Speaking Test. This test grades participants across four sub-categories of English proficiency and calculates an overall score based on these. From the Versant Test Description and Validation Summary (Pearson, 2011): *Sentence Mastery* measures sentence construction and comprehension, *Vocabulary* measures passive and active vocabulary use, *Fluency* measures phonological fluency, and *Pronunciation* measures the pronunciation of rhythmic and segmental units, both lexical and phrasal. Possible Versant English Speaking test scores can range between 20 and 80, and the scores found in the present sample range between 44 and 80, meaning a reasonably high range of possible proficiency levels is represented in the participant sample. Nonetheless, on evaluation of the distribution of scores, these were found to be very skewed towards the higher end of the scale, with a mean Versant score across participants of 69.2 out of 80. Mean scores for the *Sentence Mastery* and *Fluency* sub-categories (the former of which is most closely related to morphosyntactic proficiency) are very close to the mean total Versant score at 70.9 and 70.5, respectively. Overall mean scores for *Vocabulary* and *Pronunciation* are, however, lower, at 66.2 and 65.8, respectively. *Sentence Mastery* is the most common sub-category in which participants scored full marks, with 35.3% (18 of 51) participants scoring 80 here, suggesting high levels of morphosyntactic proficiency within the participant sample. The use of the results for the *Sentence Mastery* category (as opposed to total Versant score) as an overall proxy for English proficiency was considered, as this category is most closely relevant to the morphosyntactic processes under investigation in this study. This was, however, rejected in favour of using the total score in the interests of preserving the full range of data collected by the relatively brief 15-minute test, as well as to ensure results are more directly comparable to prior linguistic literature that has made use of the Versant testing system in their methodologies (eg. Orfitelli & Grüter, 2013). In order to more easily compare across participants of different English proficiency levels, raw Versant scores were transformed into a categorical variable. Using the cross-referencing information provided on the Versant test score reports, participants' numerical Versant scores were converted into their equivalent scores on the *Common European Framework of Reference* (*CEFR*) classification - a widely recognised 6-way categorical measure of English proficiency. This scale splits speakers into *Basic* (levels A1 and A2), *Independent* (levels B1 and B2), and *Proficient* (levels C1 and C2) users of English. As the participant sample is, overall, made up of highly proficient users of English, only 3.9% (2 participants) fall under *Basic user* (A2), 39.2% (20 participants) fall under *Independent user* (B1 and B2), and 56.9% (29 participants) come under *Proficient user* (C1 and C2). Of the *Proficient users*, 9 participants achieved the maximum possible score of 80, and no participants placed in the lowest *Basic user* category 'A1'. For the purposes of statistical model convergence, it was necessary to further collapse this English proficiency measure down, grouping the lower *CEFR* ratings *A2*, *B1* and *B2* into *B2 or lower*, and the higher ratings of *C1* and *C2* together into *C1 or C2*. Although not ideal, this resulted in a more even participant distribution of 22 participants in the former category and 29 in the latter.

3.8.2 Controls and evaluation of respondents

Prior to analysis, several checks were carried out to ensure the quality of the collected data. Firstly, responses to the negative control condition of the AJT were assessed. This condition consisted of 6 test items considered unacceptable across all varieties of English spoken by participants in this study: These contained errors in agreement, case marking, or determiner placement (see Appendix 9.1.1 for a list of negative control items used). Two participants (both Polish-born LGBTQ+) were excluded based on this as the mean of their responses to these items were of a positive polarity (as opposed

to the expected negative). A further two respondents (also Polish-born LGBTQ+) were eliminated due to not residing in England.

Remaining participants were also assessed for several other factors which might influence their AJT results, namely their AJT Response times, potential language impairments, and degree of formal study of linguistics. Also considered was the degree of linguistic and ethnic diversity across the participant sample.

3.8.2.1 AJT Response times

In order to capture participants' unconscious assessments towards the linguistic items as accurately as possible, it is common for AJT methodologies to encourage or enforce a time limit for responding to each item. This is in order to discourage overthinking, which would open the door for other factors to influence participants' decisions. Therefore, response time is an important consideration in the assessment of the overall credibility of AJT responses. Although a time limit per item was not overtly enforced in order to not lose data points (as the task was being done by participants outside a controlled environment), participants were instructed to spend only around 5 seconds on each question. There was a break mid-way through the task reminding participants of this time allowance. The mean AJT reaction time across all participants was indeed 5.0 seconds, with 65.8% of response times within this margin. Breaking this down by country of birth, for English-born participants, the mean reaction time across all AJT test items was 4.3 seconds while, for Polish-born participants, it was slightly longer, at 5.8 seconds. Response time itself is not a direct focus of investigation in the present study. These response times were considered satisfactory for the purposes of this study and this was decided in relation to previous AJT methodologies, for instance, Christensen et al. (2013: 57) found mean response times across their test conditions to range between around 3.6-5.2 seconds, placing the response times across participants in the present study within a similar ballpark.

3.8.2.2 Potential language impairments

In an effort to account for any potential impacts on performance in the Acceptability Judgement Task caused by language processing difficulties, information was collected regarding whether participants consider themselves to potentially have any language-related difficulties or impairments such as dyslexia, ADHD or other learning difficulties. The vast majority - 89% (138/155) participants - do *not* self-identify as having a potential language-related impairment. The 17 participants who *do* do not seem to have been adversely affected by this in their AJT performance; None failed the negative control measure (ie. none had a mean rating for the ungrammatical sentences of the AJT which indicated that they found them unacceptable). Therefore, all of these participants' data was deemed fit to remain within the final sample.

3.8.2.3 Formal linguistic study

Because of the snowball sampling methods employed in this research, which utilised many linguistics-oriented networks, it was important to assess participants' degree of exposure to formal study of linguistics, as a proxy for determining the likelihood that they are familiar with the language features and methodological process of the Acceptability Judgement Task. Participants' responses were categorised into *No formal linguistic study*, *Undergraduate level study*, and *Postgraduate level study*, and this information is presented in Table 3.3, below.

Group	No formal linguistic study	Undergraduate	Postgraduate	Total participants
English-born LGBTQ+	70.00%	5.00%	25.00%	40
English-born non-LGBTQ+	77.50%	15.00%	7.50%	40
Polish-born LGBTQ+	75.00%	5.56%	19.44%	36
Polish-born non-LGBTQ+	69.23%	5.13%	17.95%	39
% of Total	76.77%	8.39%	14.84%	**155**

Table 3.3*. Distribution (%) of participants according to degree of formal linguistic study, by participant group*

The vast majority of participants - 76.8% (119/155) - have never had any exposure to formal linguistic study. 14.8% (23/155) participants have undergraduate-level exposure to linguistics, however, only 9 of these have actually done or are doing a linguistics degree, and the other 14 have only studied elements of linguistics as part of other degrees such as modern languages, philology (similar to an 'English Language' degree in the UK), Russian studies, or psychology.

13 postgraduate linguists took part in the study, comprising 8.4% of participants, and most are English-born. 8 of these are doing or have completed a PhD in linguistics or related fields, and the others are/were MA or MSc students in linguistics or related fields such as Speech and Language Therapy. Even despite this, however, the overall amount of exposure to formal linguistics in this participant sample is relatively low and deemed not to be a problem for the Acceptability Judgement Task methodology.

3.8.2.4 Linguistic diversity

In order to provide some general insight into the linguistic diversity of the participant sample, participants were asked to list any languages that they know, other than English (and also Polish in the case of Polish-born participants), and provide detail as to their rough degree of proficiency. This was collected using an open text field and, as such, the level of detail given by participants varied greatly. Therefore, the responses have been coded in such a way as to provide a very general overview of participants' linguistic backgrounds, which is detailed in Table 3.4, below.

Group	No other languages known	Beginner or intermediate knowledge of multiple other languages	Beginner or intermediate knowledge of one other language	Expert knowledge of one other language	Expert knowledge of one other language and intermediate knowledge of other languages	Total participants
English-born	28.75%	45.00%	23.75%	-	2.50%	80
Polish-born	42.67%	28.00%	28.00%	1.33%	-	75
% of Total	**35.48%**	**36.77%**	**25.81%**	**0.65%**	**1.29%**	**155**

Table 3.4*. Distribution (%) of participants' exposure to languages not considered in the present study, by birth country*

35.5% (55) of the 155 participants have no familiarity with any other languages. 36.8% (57/155) have some knowledge of multiple other languages, and 25.8% (40/155) have some knowledge of one other language. In both of these latter two categories, the level of knowledge is typically to a relatively basic level such as GCSE or equivalent. Only 3 participants have expert knowledge (ie. fluency) in another language not under focus in the present research. In terms of specific languages known, the vast majority are predictably European languages (especially those commonly taught in school), with 56 participants having some level of familiarity with French, 45 with German, 30 with Spanish, and 10 with Russian (which is taught more commonly in Poland than in England). Overall, however, most participants in this sample who

do have experience with other languages have a relatively limited degree of this, for example, through school-level instruction rather than more substantial depth of exposure such as having lived in other countries.

3.8.2.5 Ethnic diversity

Table 3.5 shows the distribution of participants by self-selected ethnicity labels in this study, subdivided by participant group. The official government census ethnicity criteria were used for this (ONS, 2021). One participant declined to provide an ethnicity label and thus their ethnicity cannot be specified.

Group	White (other)	White (British)	White (Irish)	Mixed (White and Asian)	Mixed (other)	Unspecified	Total participants
English-born	6.25%	83.75%	1.25%	3.75%	3.75%	1.25%	80
Polish-born	47.10%	0.65%	-	-	0.65%	-	75
% of Total	50.32%	43.87%	0.65%	1.94%	2.58%	0.65%	**155**

***Table 3.5**. Distribution (%) of participants across ethnicity categories, by participant group*

Poland is a very ethnically homogenous country. With 96.7% of Poland's population composed of ethnic Poles, who are 'White (other background)', the participant sample is, as might be expected, not very ethnically diverse. In line with the overall ethnic makeup of the Polish-born population, 97.3% (73) of the 75 Polish-born participants in this study identify ethnically as 'White (other)'. One Polish-born participant identifies as 'Mixed (other)' and one as 'White (British)'. Of the 80 English-born participants, around 84% (67 participants) are 'White (British)'. This exceeds the 74.4% of the English population which is White British (Race Disparity Unit, 2021). Although this is higher than the overall British population, it is more comparable with the Polish sample used in this study.

3.9 Methods of Analysis

I now provide an overview of the methods of data analysis used in the present study. The analysis is broadly composed of two components; (1) an initial descriptive statistical analysis, considering the individual influence of each factor measured within this study on the dependent variable of AJT Response, and (2) an inferential statistical analysis using a mixed effects ordinal logistic regression model, through which I compare the effects of a narrowed-down selection of factors on participants' AJT Response.

A general descriptive analysis was conducted to assess the patterning of the dependent variable (AJT Response) according to each of the AJT Conditions (ie. the linguistic features of interest in this study). This included discussion of AJT Response results according to the three types of morphosyntactic construction into which the individual linguistic conditions can be grouped - ie. the positive control of argument movement; optional discourse-based movement; and nonstandard agreement (see Section 2.6 for a discussion of these categories) - as well as a breakdown of how these pattern across the four participant groups (English-born LGBTQ+; English-born non-LGBTQ+; Polish-born LGBTQ+; Polish-born non-LGBTQ+). Following this, univariate correlations were investigated between each independent variable of interest and the dependent variable (AJT Response). For each independent variable, the distribution of participants across the categories

of the variable was considered. Most participant distributions were explored according to the four participant groups but, where relevant, comparisons were made more broadly according to birth country (ie. distributions of Polish-born versus English-born participants), and LGBTQ+ status (ie. distributions of LGBTQ+ versus non-LGBTQ+ participants). Following distribution analysis, each category of each independent variable of interest within the present study (for instance *Low*, *Medium*, and *High* in the case of socioeconomic status) is compared using stacked bar charts, which show the percentage distributions of responses across each increment of the AJT scale. These are presented according to the three types of morphosyntactic construction under consideration in the present study.

Certain aspects of the descriptive analysis benefitted from the reporting of average responses across certain breakdowns of the participant sample or categories of the variables under consideration. Because the data derive from Likert scale measures which are ordinal in nature, the information being measured is the underlying relative level of acceptability via the patterning of responses to the relative ordered levels of the scale, and not directly the numbers of the scale - the numerical labels here are simply placeholders representing levels in an ordered hierarchy. This means that typical measures of assessing patterns of distribution of continuous data, such as the calculation of mean values, were not the most appropriate in this case (Bross, 2019: 19). For instance, a mean average across AJT Responses does not correspond to any meaningful value in the data as there are no values 'between' the independent increments of the scale. As such, where the reporting of averages benefitted the analysis, median values were instead calculated for AJT and survey results. These types of averages capture the increment of the scale which occurs at the midpoint of the entire distribution of responses collected. The median value constitutes a more accurate measure of the average value across ordinal data as it has equal probability of garnering a response higher or lower than it. Importantly, medians are more informative in cases where data is subject to floor-and-ceiling effects, where mean averages may unfairly indicate middling responses (Sullivan & Artino, 2013: 541).

The main aim of the descriptive analysis (explored in Chapter 4) is to provide an initial exploration of potential effects on the dependent variable (participants' AJT Response) by investigating univariate (ie. one-to-one) correlations between each independent variable under analysis and the dependent variable. Furthermore, understanding how individual factors pattern with AJT Response allowed for the investigation of potential correlation and multicollinearity that might need to be mitigated prior to the main modelling process. Effects of the independent variables on AJT Response were visualised according to the three types of morphosyntactic construction investigated in this thesis, using graphical methods to provide a general overview.

Following this, inferential statistical methods were used for more advanced hypothesis testing. The purpose of this was to test the strength of association, or statistical significance, of the observed effects and relationships between different independent variables and the dependent variable, and, hence, the likelihood of the participant groups rating the linguistic test conditions in different ways from each other. The following section outlines the inferential statistical methods used in the present study, and the decisions undertaken along the way will also be explored here.

3.9.1 Statistical modelling procedure

To recap, the dependent variable under analysis in this study is participants' 7-point Likert-type response to the Acceptability Judgement Task (AJT) test items (henceforth, AJT Response). In the AJT, 'acceptability' was treated in terms of 'naturalness' along an equidistant scale ranging from -3 (*completely unnatural*) to 3 (*completely natural*), with 0, at

the centre of the scale, equivalent to *neither natural nor unnatural*. Because each point within the Likert rating scale constitutes a sequentially ordered discrete level of the response variable (the test item's perceived degree of 'naturalness') - ie. -2 is always 'less natural' than -1; 1 is always 'less natural' than 2, etc - the dependent variable is of an ordinal nature. An ordinal variable is made up of discrete categories that, although numbered, only represent the underlying ordered categories and not the numbers themselves (Liu & Zhang, 2018). Although the categories of an ordinal variable are known to be ordered, it is generally unknown (and not strictly relevant) whether each category represents the same 'distance' along the scale. This distinguishes ordinal data from interval data (which is made up of equally spaced intervals), the latter of which is considered a minimum requirement for parametric approaches such as linear regression.

The attribution of Likert scale numerical values to judgement ratings has been problematised as the association between a particular sentiment and its corresponding integer value is not "mathematically demonstrable" (Hodge & Gillespie, 2005: 49). Participants are known to use rating scales differently, with some skewing towards particular ends of the scale, or restricting the range of the scale they interact with (Schütze & Sprouse, 2013: 39). Ordinal data, especially that pertaining to non-standard acceptability judgements, is likely to polarise respondents into one of the extreme ends of the scale. As such, these types of ordinal scales may be subject to floor and ceiling effects, which describes a distribution where the responses skew towards the peripheral scale categories (Christensen, 2015: 3) - eg. towards -3 and 3, in the case of the present study. Anticipating such a floor and ceiling effect, the bimodality_coefficient test from the *modes* package (Ashman et al., 1994) was performed on the data, and this indicated a bimodality coefficient of 0.716. According to Pfister et al., (2013), data that surpasses a threshold of 0.555 is likely to be bimodal. Therefore, this test confirmed that the distribution of the AJT Response dependent variable is bimodal - ie. there is not one 'peak' in the distribution of responses, but two. To confirm this is a floor and ceiling effect (ie. to confirm that the two modes, or 'peaks', are indeed located at the peripheries of the scale, the *locmodes* function from the *multimode* package (Ameijeiras-Alonso et al., 2021) was used to estimate the modes for AJT Response. This test confirmed that the two modes are indeed located at the peripheries of the scale (at -2.802 and 2.877, respectively), with an antimode (the lowest point between the two 'peaks') near to the middling value of the scale (-0.116). In short, these tests confirmed that more extreme responses (toward either polarity) were relatively more common compared to more middling responses.

Additionally, responses on an ordinal scale are inherently constrained by the parameters of the scale itself (eg. *completely unnatural* and *completely natural*, in this case) - responses cannot exist beyond the limitations of -3 and 3, respectively. This means that, unlike continuous data, for which the numerical extremes would be theoretically unlimited, the discrete data from ordinal scales cannot conform to a normal distribution (Christensen, 2015: 3). Normal distribution is one of the key assumptions underlying linear models, and it has been argued that the use of continuous-variable models (such as the linear regression models often employed in Linguistics) for ordinal dependent variables introduce a large degree of bias and error (eg. Winship & Mare, 1984: 512; Hodge & Gillespie, 2007: 2). Given these issues, much consideration was given as to the best method of modelling ordinal data.

Linguistic studies which measure a continuous dependent variable (which can take one of a theoretically infinite number of potential numerical values between two points - such as frequency or response time) often use linear regression to model the effects of the independent variables of interest on this dependent variable. If the dependent variable is categorical (consisting of a set of discrete 'options' - such as a binary 'yes' vs 'no', or one of a larger set of categories), however, logistic

regression is commonly used instead. In the present study, an ordinal logistic regression model[8], an extension of the logistic regression model, was used for inferential statistical analysis, as this type of model is used with categorical dependent variables whose categories inherently have some kind of relative ordering to them (such as the AJT Response dependent variable being considered here). This type of model works by predicting several threshold coefficients (ie. intercepts), one for each ordinal category (level) of the dependent variable, except for the first level which functions as a reference level relative to which the remaining coefficients are calculated. Logistic regression models differ from linear regression in that the coefficients are calculated in terms of *logits*; In mathematical terms, these are natural logs of the proportional odds ratios of the model predictors. In simple terms, logits are a transformed measure derived from probability that are easier for the model to work with. Essentially, each of the coefficients predicted by the model represents the odds of the AJT Response falling at that level of the response scale or higher.

Following model fitting, for ease of interpretation, these logit values first require conversion into their odds ratio form. ORs are ratios of two odds values, and the odds values themselves are ratios of two probabilities. Explained in more detail in Chapter 5, odds describe the ratio between the probability of an outcome and the probability of another outcome, and odds *ratios* (henceforth, ORs), in this context, are ratios comparing the odds of an outcome for a given category of the predictor variable relative to its reference category.

Much like for the categories (levels) of the dependent variable, it is the case that, for the categories of each model predictor, one category is assigned as the reference category. Examining the ORs of the model predictors, therefore, tells us the odds of an AJT Response being higher for any given category of a predictor variable relative to that predictor's assigned reference category. The ORs of each non-reference category of a predictor can then be compared with each other, to assess how much higher or lower the respective odds ratios are, and to show if particular outcomes are more or less likely than others.

Ordinal logistic regression models have other relevant advantages over linear regression, for instance, they are not negatively impacted by the floor and ceiling effects prevalent with ordinal data, which is especially advantageous when a high degree of skew is present in the dependent variable (Winship & Mare, 1984: 512-4), as is the case in the present study. Ordinal logistic regression models also allow for the calculation of interaction effects between model predictors, which is something that is important to take into account among extralinguistic factors (eg. Sankoff, 1988: 17).

3.9.2 Modelling prerequisites

3.9.2.1 Proportional Odds assumption

Although model assumptions such as that of independence are not applicable to mixed effects models, the most crucial assumption underlying the use of ordinal logistic regression models is that of proportional odds (also called parallel slopes). In simple terms, this is the assumption that the effect of a given predictor is consistent across the entire range of the scale. So, if a given predictor makes someone more likely to respond with a higher value on the scale, this relative increase in likelihood should be the same between level 1 and 2 of the scale as it is between level 2 and 3 of the scale, for instance.

[8] Other names for this type of model include *Proportional Odds Model* and *Ordered Logistic Regression*

As such, it was initially necessary to assess whether the data met this assumption and were suitable for use in this type of modelling, which was done using the *pomcheck* function from the *pomcheckr* package (Wong, 2021). This involved running a series of binary logistic regressions on the response variable (AJT Response) against the predictor variables and assessing whether the slope coefficients for the levels of each predictor are equal. For this, a basic model was run without any random effects or interaction terms, and incorporating the fullest dataset (ie. all participant groups). Only one predictor is able to be assessed at a time, therefore, the process had to be repeated for each predictor. The outputs of the *pomcheck* function generated logit plots (eg. Figure 3.3) which could then be visually assessed. Taking the focal predictor of AJT Condition as an example, the R script was as follows:

```
plot(pomcheck(AJT_Response ~ AJT_Condition, data = db_master), legend.position = "right")
```

In the resulting logit plot, ideally, the distance between the set of points within each level of AJT Response should be similar. In other words, the different coloured points represented in Figure 3.3 (with vertical connecting lines manually added for additional clarity) should be roughly evenly spaced apart on each row of the plot (ie. for each category of the predictor variable being tested). Although this is ideal, it is not realistic with complex data, and it is not expected that these vertical lines will be exactly linear. What is most important is that these vertical lines do not overlap each other and are not very scattered. Furthermore, because the data in the present study is subject to floor and ceiling effects, meaning the majority of data points are clustered towards the peripheries of the scale, in verifying the proportional odds assumption, I focus mainly on the peripheral levels of AJT Response[9]. The logit plots only include 5 of the 7 levels of AJT Response because the lowest level (-3) is excluded by the ordinal logistic regression model. The next lowest level (-2) is used as the reference level by *pomcheckr*, and is centred on the logit value of 0 on the x axis (though these centred values for -2 are not shown on the plot). Assessing the logit plot for the focal predictor of AJT Condition (Figure 3.3), although the lines for levels 1, 2, and 3 are not perfectly vertical, they do not overlap one another. The sparseness of data for AJT Response levels -1 and 0 are a likely result of the relative lack of vertical linearity across these levels of the response variable. Based on this, the proportional odds assumption is reasonably met for the purposes of using the ordinal logistic regression model.

[9] These decisions were made in consultation with statisticians from the University of Sheffield's Statistical Services Unit

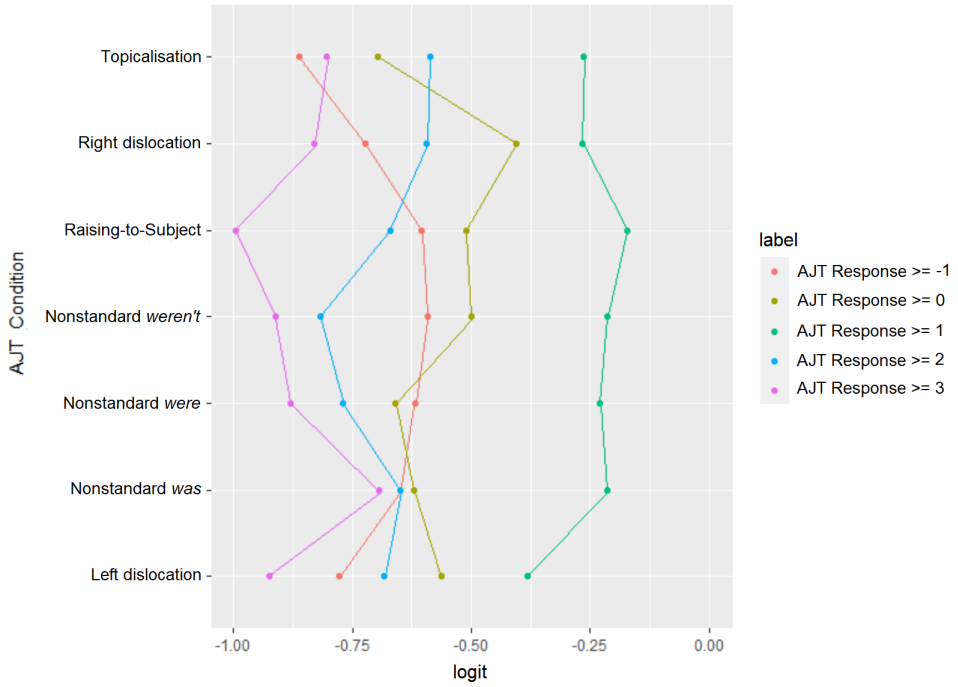


Figure 3.3. *Logit plot for testing the proportional odds assumption in the focal predictor of AJT Condition*

3.9.2.2 Internal consistency checks

Prior to statistical modelling, Cronbach's alpha measurements were calculated for each test condition. This a formal test of the degree to which items within each of the AJT Conditions are measuring the same underlying variable across participants; ie. the degree to which all the individual test items in each condition pattern similarly (eg. Sullivan & Artino, 2013: 542). A score of 0.7 is considered acceptable, 0.8 is considered good, and 0.9 is excellent. The results (Table 3.6) confirm that all 7 test conditions do indeed satisfy Cronbach's alpha calculation of internal consistency.

			Sample units: 155
	Cronbach's alpha	Bootstrap 95% CI based on 1000 samples	
		2.50%	**97.50%**
Raising-to-Subject - Items: 8	**0.775**	0.685	0.833
Left Dislocation - Items: 7	**0.708**	0.612	0.775
Right Dislocation - Items: 7	**0.848**	0.802	0.883
Topicalization - Items: 8	**0.744**	0.671	0.792
Nonstandard *was* - Items: 7	**0.882**	0.846	0.908
Nonstandard *were* - Items: 8	**0.895**	0.863	0.921
Nonstandard *weren't* - Items: 8	**0.936**	0.916	0.952

Table 3.6. *Cronbach's alpha results for the 7 test conditions*

3.9.2.3 Model fitting

The specific model implemented in the present study was a Cumulative Link Mixed-Effect Model (CLMM) using the *ordinal* package (Christensen, 2018) in R Studio. Several other implementations of ordinal logistic regression models were considered, but, ultimately, *ordinal's 'clmm'*[10] was chosen as it allows for the incorporation of multiple random effects, to account for individual effects such as those of participants and test items. As each participant gave multiple responses, and multiple test items formed each test condition, it is important to take into account any variation and, hence, loss of model accuracy, caused by these factors (Bross, 2019: 17). In particular, factoring in participant random effects can reduce Type I errors whereby predictors such as age or gender appear significant due to individual variation in the participants within those groups (eg. Baayen, 2008).

Given the structure of the participant sample of interest, a nested approach was taken to analyse results among different groups within the participant sample. This was particularly necessary as not all predictors related to all participant groups (eg. LGBTQ+ Community Involvement band only applies to LGBTQ+ participants). Two models were required to test the effect of English language proficiency in Polish-born speakers, with the broader model factoring in English language exposure in years, and the nested model instead factoring in a formal measure of English proficiency (represented by CEFR classification). This latter data was only available for a subset of Polish-born participants, requiring a separate model to account for this measure.

[10] The *clmm2* function was also considered as it allows for the relaxing of the proportional odds assumption, however, this approach was ultimately rejected as *clmm2* only allows for one random effect and, in any case, the proportional odds assumption was deemed to be satisfactorily met

The modelling procedure, therefore, required running four separate models:

1. The broadest model, which takes into account all four participant groups and all predictors that apply to all participants
2.
    a. A Polish-born participant model, which takes into account only Polish-born participants (both LGBTQ+ and non-LGBTQ+) and all predictors that apply to them (namely English language exposure)

    b. A Versant (English proficiency) model, which takes into account only Polish-born participants with a Versant English speaking test score and all predictors that apply to these participants (namely CEFR classification)
3. An LGBTQ+ model, which takes into account only LGBTQ+ participants (both English-born and Polish-born) and all predictors that apply to them (namely LGBTQ+ Community Involvement rating)

Many factors were measured in the present study, and it was important to first narrow down the set of predictors incorporated as fixed effects in statistical modelling. Some predictors were expected to co-vary, such as age of arrival to England, length of residency in England, and length of English language exposure. The predictor most relevant to answering the research questions was selected - in this case, the indicator of English language proficiency, length of English language exposure. For the acculturation measures, the strength of correlation between raw mean total scores (as opposed to categorical ratings) and the dependent variable of AJT Response was tested. For this, Spearman Rank correlation was used, which, unlike Pearson's correlation, does not require continuous data nor a normal distribution and is therefore appropriate for the ordinal AJT Response with floor and ceiling effects. First, overall English and Polish acculturation scores were compared against AJT Response, however, Polish acculturation did not significantly correlate with the dependent variable ($p = .269$). Although English acculturation did show a significant weak positive correlation ($rho = .102$, $p = < .001$), this was not used as it was more desirable to be able to compare amongst equivalent measures for both cultures. Therefore, the three sub-categories of acculturation were considered instead. The *Lifestyle* subcategory had the most statistically significant correlation with AJT Response for both English ($rho = .128$, $p = < .001$) and Polish ($rho = .048$, $p = .001$) cultures. For this reason, and because the *Language Use* category was decided to be swaying the overall acculturation scores (because people who are living and working in England are very likely to use the English language on a regular basis and are comparatively less likely to need to use the Polish language, even if they have a strong alignment with Polish culture), *Lifestyle* was selected to be used as a proxy for acculturation measurements.

Likelihood ratio tests were performed on the remaining variables, one by one, in order to individually assess the significance of the variable's inclusion within the model (henceforth, 'predictor' will be used instead of variable). This process compares the goodness of fit of two nested models, and was conducted as such: For each *predictor X*, a CLMM was run (using the *ordinal* package in R), containing the focal predictor of *AJT Condition* and *predictor X*, both as fixed main effects. Another CLMM was then run, instead containing an **interaction** between *AJT Condition* and *predictor X*. The goodness of fit of these two nested models were then compared. If a statistically significant difference was found between the model with the interaction effect versus that without, then *predictor X* and the interaction were both kept for inclusion in the model. If the Likelihood Ratio test did not produce a statistically significant result, then the model with fixed main effects (ie. without the interaction) was instead compared against a model containing **only** *AJT Condition*, in order to assess whether *predictor X* is

significant as a main effect outside of the interaction. If this was also insignificant, then the predictor was removed from the modelling process[11]. The significance of *AJT Condition* itself was also confirmed by comparison against a null model ($p = < .001$). P values from Likelihood Ratio testing can be seen in Table 3.7, below. Following Likelihood Ratio testing, the predictors of *Sexuality* and *Polish Lifestyle* (acculturation) were found to be insignificant both as main effects and interaction effects, and were therefore eliminated. *AJT Condition*, *Birth country*, *Region of residence* and *English Lifestyle* were all found to have significant main effects on AJT Response, and the latter three also have significant interaction effects with *AJT Condition*. The remaining predictors were only significant in interaction with the focal predictor.

Predictors	**Main effect** (*p value to 3dp*)	**Interaction with AJT Condition** (*p value to 3dp*)
AJT_Condition	< .001	N/A
Birth country	< .001	< .001
LGBTQ+ status	.373	< .001
Sexuality	.385	5.184
Gender	.310	< .001
Age band	.328	< .001
IMD (socioeconomic status)	.848	< .001
Higher-educated status	.346	< .001
Region of residence	.002	< .001
English Lifestyle	.002	< .001
Polish Lifestyle	.865	.863
English language exposure (years)	.821	.007
CEFR (English proficiency)	.644	< .001
LGBTQ+ Community Involvement	.393	< .001

Table 3.7*. Likelihood ratio test results showing statistical significance (p values) of each predictor considered for statistical modelling, both as a main effect and as an interaction effect with the focal predictor of AJT Condition. Predictors whose rows are shaded were ultimately rejected from statistical modelling due to low significance.*

[11] This process was undertaken based on advice from statisticians from the University of Sheffield's Statistical Services Unit

Prior to running the models, some setup was required[12]. In order for R to treat different types of data in the appropriate way, predictors were manually coded as either factors or ordered factors. Given the type of model used, the most crucial variable to re-code was the dependent variable (AJT Response) as, by default, R treats any numbers in the dataframe as simple numeric data, and the model requires the response variable to be an ordered factor. Therefore, R was instructed to treat the response variable and other ordinal predictors (eg. English lifestyle, age band, English language exposure, CEFR classification) as such. Other categorical, but not ordinal, predictors (eg. birth country, LGBTQ+ status, gender, higher-educated status) were coded as factors, as opposed to R's default of 'character'. Ordinal logistic regression models require each categorical predictor to be designated a reference category, against which the other categories are compared. These are alphabetical by default, so it was necessary to manually assign these reference categories. For AJT Condition, the argument movement condition (raising-to-subject with *seem*) was used, as it was most appropriate to use this positive control condition as the reference category. For other categorical predictors, reference categories were: Birth country - *English-born*; LGBTQ+ status - *No*; Region - *Midlands*; Gender - *Male*; Higher-educated status - *No*.

Using the list of predictors that had been narrowed down with Likelihood Ratio tests (as demonstrated in Table 3.7), the R package *buildmer* (Voeten, 2020) - specifically, *buildclmm*, the function for CLMMs - was used to create maximal feasible models containing as many as possible of the model predictors of interest, but still able to successfully converge (ie. able to reach a stable point where the model has completed all its calculations). This was an automated process, using a method called stepwise elimination, in which model predictors are consecutively added or removed to assess the relative effect that each predictor has on the model's performance[13]. For each of the four models, all potentially relevant predictors and their interactions with the focal variable of AJT Condition were included in the process. This was in order to include both predictors which had a significant main (direct) effect on participants' AJT Response, and which ones had an effect on participants' responses to the different AJT Conditions (ie. had an interaction effect with the AJT Condition predictor). Because it was important to include the random effects terms for participant and AJT test item in the final model terms, these were enforced as required components of the *buildmer* output, meaning that *buildmer* would not eliminate them during the process. The sets of fixed main and interaction terms included and rejected by *buildclmm* throughout this process can be seen in Table 3.8). Predictors marked ☒ were incorporated into the final model while those marked ╳ were originally selected for inclusion but did not pass the *buildmer* process. Predictors marked `N/A` are not applicable to that model (eg. *birth country* among only Polish-born participants). The model outputs from this process can be viewed in Appendix 9.8.

[12] There are some other technical aspects of the model setup that have not been mentioned in-text: In running the CLMM models, the default link function (logit link) was used, as well as the default optimiser (ucminf), a flexible threshold, and the default Laplace approximation. Other alternatives to these were considered but these default options resulted in the best model fit. Although random intercepts were included for participant and test item, random slopes were not included in the models used in this study, to aid with model convergence. Additionally, following Christensen (2023), Hessian values of all model outputs were assessed to ensure model optimisation reached a near-optimal level. None were higher than 25,000, which is within the upper threshold of 106 that Christiansen suggests.

[13] Specifically, both 'forwards' and 'backwards' stepwise elimination was conducted, and the elimination criterion was based on likelihood-ratio. See Voeten (2020) for more information about this.

	Predictor	Model			
		1. *All participants*	2a. *Polish-born participants*	2b. *Polish-born participants with CEFR rating*	3. *LGBTQ+ participants*
Main effects	AJT Condition	☑	☑	☑	☑
	Birth country	☑	N/A	N/A	☑
	LGBTQ+ status	☑	☑	☑	N/A
	Age band	☑	☑	☑	☑
	Gender	☑	☒	☑	☑
	Region of residence	☑	☑	☑	☑
	Socioeconomic status	☑	☒	☒	☑
	Higher-educated status	☒	☑	☑	☑
	English language exposure (years)	N/A	☑	N/A	N/A
	English proficiency (CEFR)	N/A	N/A	☑	N/A
	English lifestyle (acculturation)	☒	☑	☑	☑
	LGBTQ+ Community Involvement	N/A	N/A	N/A	☑
Interaction effects with AJT Condition	Birth country	☑	N/A	N/A	☑
	LGBTQ+ status	☑	☑	☑	N/A
	Age band	☑	☑	☑	☑
	Gender	☒	☒	☑	☑
	Region of residence	☑	☑	☑	☑
	Socioeconomic status	☑	☒	☒	☒
	Higher-educated status	☒	☑	☑	☑
	English language exposure (years)	N/A	☑	N/A	N/A
	English proficiency (CEFR)	N/A	N/A	☑	N/A
	English lifestyle (acculturation)	☒	☑	☑	☑
	LGBTQ+ Community Involvement	N/A	N/A	N/A	☑

Table 3.8. *Fixed main and interaction terms included within the nested model structure. Shaded predictors are specific to the model in question*

3.9.3 Reporting of results

Prior to analysis, it was important to account for the effect of multiple comparisons (Chen et al., 2017: 1725) - ie. the fact that each model incorporates a very large number of terms. As such, all terms across the four models underwent p value correction to ensure integrity of any results deemed statistically significant. Originally, Bonferroni correction and other methods of control of Family-Wise Error Rate (FWER) - ie. the probability of occurrence of a Type I error - were considered. However, these types of corrections risk being overly conservative when many hypotheses are simultaneously tested, or if multiple hypotheses are correlated (Chen et al., 2017: 1727), resulting in artificially low statistical power.

Instead, Benjamini-Hochberg (BH) adjustment was opted for which, instead, controls for False Discovery Rate (FDR), an estimate of the presence of Type II errors. This controls the expected proportion of false discoveries amongst the rejected hypotheses and, although an increased risk is present that this method may permit an occasional false positive, it is overall more advantageous as model terms which are actually significant will not be erroneously rejected. Model terms which retained $p < 0.05$ following BH adjustment were extracted and compiled into tables, along with their associated coefficient estimates (in terms of ordered log-transformed odds, or 'logits') and other elements of the model output (standard errors and z values of the coefficient estimates). Proportional odds values were calculated by exponentiating the raw logit values from the model outputs. Confidence Intervals (CIs) were calculated for the proportional odds values and assessed to ensure the range between the 2.5% and the 97.5% CI did not overlap the null value of 1. Model outputs can be found in Appendix 9.8.

Following the compilation of results from model outputs, key findings (namely, the patterning of statistically significant linguistic conditions in each of the models) were visualised using probability curves. This process was based on Ackerman's (2019) method which utilises similarly structured ordinal data. The plots demonstrate how participants across the key participant groups within the present study (ie. all participants, versus only Polish-born participants, versus only LGBTQ+ participants) utilise the AJT response scale similarly or differently according to the linguistic conditions of interest. Visualising these results in terms of probability is appropriate to the type of statistical model used as probabilities can be directly calculated from the logit/log odds values calculated by the proportional odds logistic regression models. Furthermore, because the ordinal dependent variable (AJT Response) is made up of independent, discrete, ordered levels it is more appropriate to compare the probabilities of one level of the scale being selected over another than to use methods intended for continuous data, such as assessing the average responses across the test conditions (Ackerman, 2019). The vertical distances between test conditions on these plots indicate the degree of similarity with which participants utilise the rating scale when assessing the different test conditions, meaning that probability curves provide a very clear visual indication of how test conditions pattern relative to each other.

3.10 Conclusion

To conclude, in this chapter I have outlined the methodological structure of the present study. In order to tailor the method to this study's research aims, several distinct methodological elements were required. I have discussed the design of these elements, which are as follows: (i) an Acceptability Judgement Task, to measure participants' perceptions of British-English morphosyntactic features; (ii) the Versant English Language Speaking Test, to provide a formal measure of Polish-born participants' English language proficiency; (iii) Sociological surveys, to measure participants' Polish and English acculturation levels, and their LGBTQ+ Community Involvement levels; and (iv) a Demographic Information form, to be able to account for key constraints on the acquisition of linguistic variation according to traditional sociolinguistic accounts (such as participants' age, gender, socioeconomic status, and the dialect region in which they live), as well as constraints typically considered in L2 acquisition research (age of onset, degree of input, prior language instruction). I have outlined the participant sample characteristics and explained how ethical review and participant recruitment was undertaken. Then, I explored the study design and justification for use of the given methods. I covered the pilot study process and the changes made following this. I also outlined how the data was collected, cleaned and analysed, and how the statistical modelling procedure was conducted. In the next chapter, I will explore the descriptive results of this study.

4. Descriptive Results

4.1 Introduction

The following chapter outlines the descriptive results of this study. Firstly, a general overview is provided of the patterning of the dependent variable (AJT Response) according to each of the AJT Conditions (ie. the morphosyntactic features of interest in this study). I then explore the univariate (ie. individual) effects of relevant factors on AJT Response, and these are divided across five broad types: (i) the key sample characteristics of birth country and LGBTQ+ status; (ii) the macro-social factors of interest, namely age, gender, region, and the socioeconomic factors of socioeconomic status and education status; (iii) the L2 (second language) factors influencing Polish-born participants, namely their age of arrival to England and their English language proficiency, and; (iv) the meso-social factors considered in this study (ie. those related to community embeddedness) - participants' levels of English and Polish acculturation, as well as their levels of LGBTQ+ Community Involvement. AJT Response results will be considered according to the three types of morphosyntactic construction investigated in this thesis: (i) argument movement, consisting of raising-to-subject with the verb *seem* (used here as a positive control condition); (ii) optional discourse-based movement, consisting of left dislocation, right dislocation, and topicalisation; and (iii) nonstandard agreement, consisting of past-tense BE agreement: nonstandard *was*, nonstandard *were*, and nonstandard *weren't*.

4.1.1 Interpreting descriptive results

For each independent variable (ie. factor under consideration), the distribution of participants across the categories of the variable will be considered. Distributions are provided in terms of the percentages of participants who fall into each category of the independent variable being compared, and this is provided per participant group as well as across all participants. Most participant distributions are explored according to the four participant groups but, where relevant, comparisons are made more broadly according to birth country (ie. distributions of Polish-born versus English-born participants), and LGBTQ+ status (ie. distributions of LGBTQ+ versus non-LGBTQ+ participants). This distribution analysis is useful for assessing whether the sample is reasonably equally distributed according to country of birth or LGBTQ+ status for each factor, and whether this is suitable for incorporation into statistical analysis.

Certain aspects of the descriptive analysis require the reporting of average responses across certain breakdowns of the participant sample or categories of the variables under consideration. Because the data derive from Likert scale measures which are ordinal in nature, where required, average AJT and survey responses across variables and/or sections of the participant sample will be reported in terms of median values as opposed to mean averages (see Section 3.9 for a justification of this and of the methods of descriptive analysis more broadly).

Following distribution analysis, each category of each independent variable of interest within the present study (for instance *Low*, *Medium*, and *High* in the case of socioeconomic status) is compared using stacked bar charts, which show the percentage distributions of responses across each increment of the AJT scale. These visualisations are presented according to the three types of morphosyntactic construction investigated in this thesis - ie. argument movement, optional discourse-based

movement, and nonstandard agreement. The total width of each stacked bar equates to 100% of AJT Responses within the indicated category, and the distribution of AJT Responses can be assessed by evaluating the relative width of each different coloured section of the bar, each corresponding to a different increment of the AJT Response scale. Furthermore, the overall range of AJT Responses (namely the proportion of positive compared to negative evaluations) can be assessed by comparing the relative horizontal positioning between each of the stacked bars. Results from these visualisations will primarily be summarised in terms of the overall percentages of positive versus negative responses (ie. responses between -3 and -1 versus those between 1 and 3), highlighting any categories with noticeably higher or lower proportions of responses at the extremities of the AJT scale (ie. of -3 or 3). Percentages will be reported to two decimal places.

4.2 Morphosyntactic conditions

Before I consider how AJT Response patterns according to each individual independent variable, I will consider broad patterns in the individual AJT Conditions (ie. the different morphosyntactic features) and patterns within the three types of morphosyntactic construction investigated in this thesis: argument movement, optional discourse-based movement, and nonstandard agreement. Table 4.1 shows the median responses to the individual AJT Conditions as well as the three types of morphosyntactic construction. These are compared across all participants and also by birth country.

Morphosyntactic construction	*AJT Condition*	Median AJT Response (*range: -3 to 3*)					
		All participants		English-born		Polish-born	
Argument movement	Raising-to-subject with *seem*	3	3	3	3	3	3
Optional discourse-based movement	Left dislocation	2	2	2	2	2	2
	Right dislocation	2		2		1	
	Topicalisation	2		2		2	
Nonstandard agreement	Nonstandard *was*	-1	-1	0	1	-2	-2
	Nonstandard *were*	-1		1		-2	
	Nonstandard *weren't*	0		1		-2	
Overall median AJT Response		1		2		1	

***Table 4.1**. Median AJT Responses by AJT Condition and type of morphosyntactic construction, across all participants and by birth country*

The overall median AJT response across all participants and test items is 1. According to country of birth, Polish-born participants, on average, provide AJT responses that tend towards more negative evaluations (ie. -1 or lower) compared to English-born participants, who provide a higher proportion of positive evaluations (ie. 1 or higher). This is reflected in the fact that the median response across all test items among English-born participants is 2, but only 1 across Polish-born participants. The median values of responses towards each of the three types of morphosyntactic construction follow a hierarchy of relative acceptability, with argument movement averaging the maximum AJT response of 3, optional discourse-based movement averaging 2, and nonstandard past-tense BE agreement averaging a negative median of -1. Comparing by country of birth, both argument movement and optional discourse-based movement are rated similarly by

Polish-born and English-born participants - both average the same overall medians of 3 and 2, respectively. Nonstandard past-tense BE agreement, on the other hand, is more negatively evaluated by Polish-born participants, with a median of -2, compared to English-born participants, who average a positive median of 1 for this type of construction. Breaking down optional discourse-based movement into its individual morphosyntactic features, left dislocation and topicalisation are evaluated similarly regardless of country of birth (each with a median AJT Response of 2), while right dislocation is evaluated less positively by Polish-born participants (median response of 1) than English-born (median of 2). As for nonstandard past-tense BE agreement, nonstandard *was*, *were*, and *weren't* appear to be equally negatively evaluated by Polish-born participants (all with medians of -2). English-born participants are comparatively more accepting of all three non-standard conditions, on average rating both nonstandard *were* and *weren't* positively (with a median response of 1) and rating non-standard *was* as less acceptable than the former two (with a median of 0).

To assess the distribution of AJT Responses in more detail, I now consider the positive versus negative evaluations of the morphosyntactic conditions considered in this study, grouped according to type of morphosyntactic construction. Predictably, argument movement test items containing raising-to-subject constructions with the verb *seem* (which functioned as a positive control condition in the AJT), have a 96.89% positive response rate, with 85.73% responses at the maximum level of acceptance (3). Optional discourse-based movement is, overall, rated positively in 72.58% of cases, and negatively in 21.76% of cases. Nonstandard past-tense BE agreement has the most negative evaluations overall, rated positively across all participants in 40.76% of cases and negatively in 53.86% of cases.

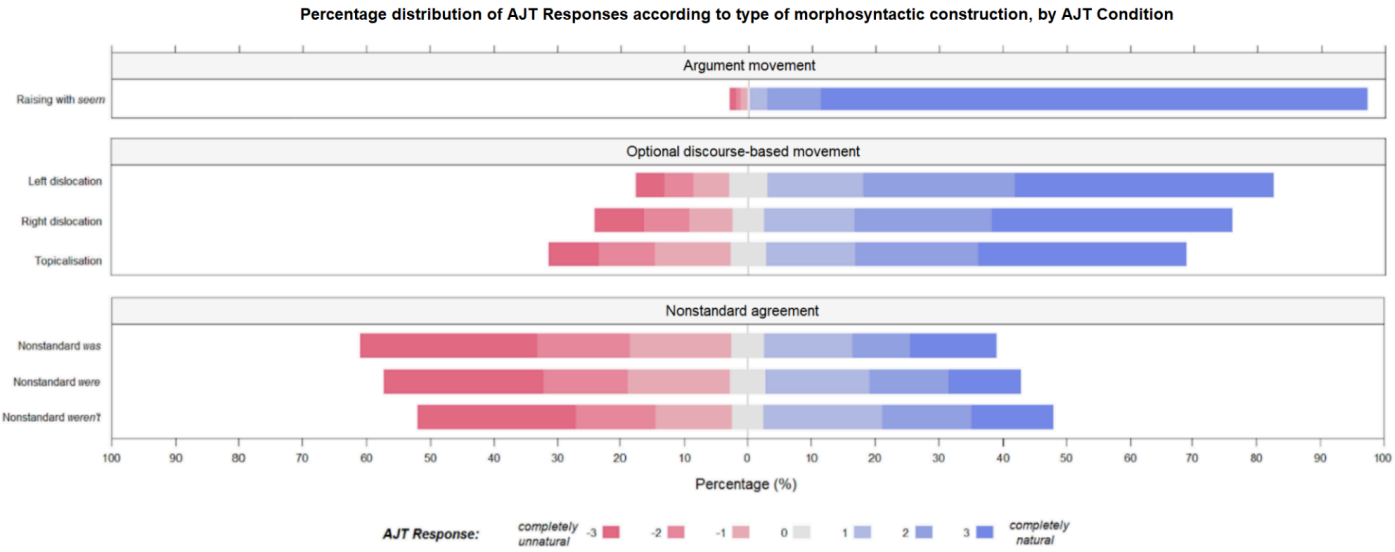


Figure 4.1. *Distribution (%) of AJT Responses according to type of morphosyntactic construction, by AJT Condition (morphosyntactic feature)*

Narrowing in on the individual AJT Conditions, Figure 4.1 shows that left dislocation is the non-control condition with the most positive range of AJT Responses (ie. the stacked bar has the right-most positioning), with 79.45% positive responses to this condition and only 14.47% negative. This is followed by right dislocation, which has 73.36% positive and 21.47% negative responses, and then topicalisation, with 65.89% positive and 28.39% negative responses. Finally, across

nonstandard past-tense BE agreement constructions, nonstandard *weren't* has the least negative range, with 45.40% positive and 49.44% negative responses, followed by nonstandard *were*, with 40% positive and 54.27% negative responses, and nonstandard *was*, with 36.31% positive and 58.43% negative responses.

4.2.1 Key sample characteristics

This study recruited 155 participants, with a roughly equal four-way split according to the two main characteristics of interest: country of birth (75 Polish-born and 80 English-born), and LGBTQ+ status (76 LGBTQ+ and 79 non-LGBTQ+). The groups are composed as follows:

- Polish-born native speakers of Polish, living in England, who identify as LGBTQ+ (henceforth *Polish-born LGBTQ+*); N = 36
- Polish-born native speakers of Polish, living in England, who do **not** identify as LGBTQ+ (henceforth, *Polish-born non-LGBTQ+*); N = 39
- English-born native speakers of British English, living in England, who identify as LGBTQ+ (henceforth, *English-born LGBTQ+*); N = 40
- English-born native speakers of British English, living in England, who do **not** identify as LGBTQ+
- (henceforth, *English-born non-LGBTQ+*); N = 40

4.2.1.1 Birth country

Figure 4.2 shows AJT responses by birth country, according to the three types of morphosyntactic construction.

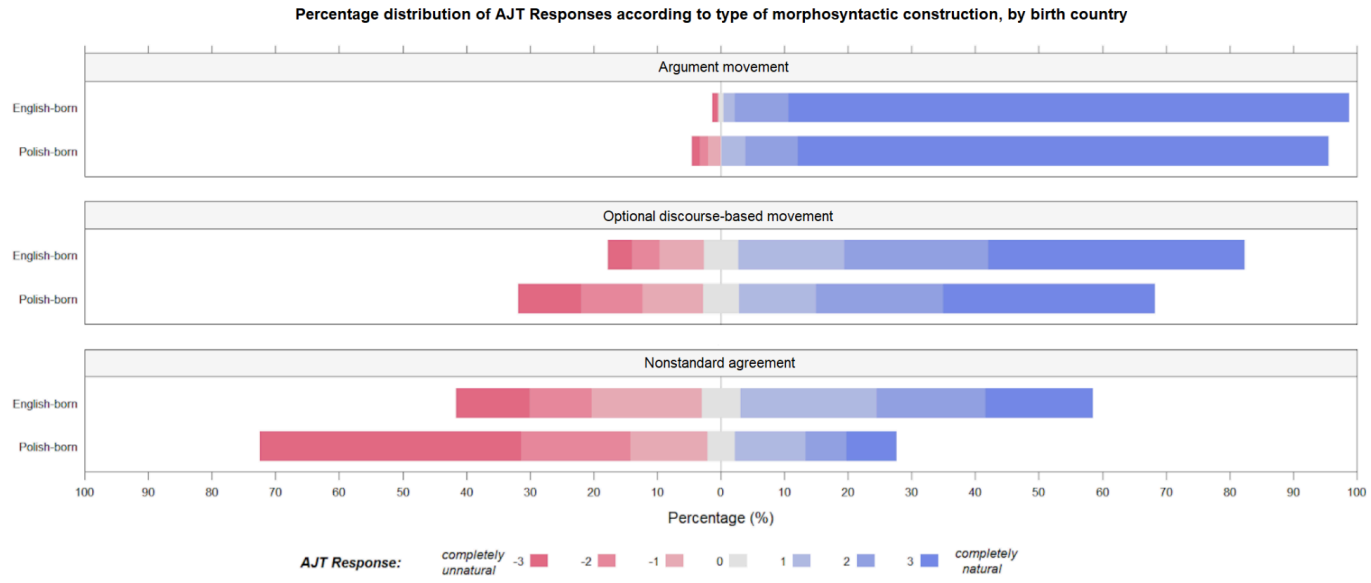


***Figure 4.2**. Distribution (%) of AJT Responses according to type of morphosyntactic construction, by birth country*

Polish-born participants' ratings of the positive control condition of argument movement (raising-to-subject with *seem*) are similarly positive compared to the whole participant sample, with 95.33% positive evaluations by Polish-born participants (compared to 98.28% across English-born participants), and 83.33% responses of '3', the maximum level of

acceptance (compared to 87.97% across English-born participants). Although Polish-born participants do have a slightly higher percentage of negative evaluations, the difference is minimal. Optional discourse-based movement is rated less positively among only Polish-born participants compared to English-born participants, with 65.21% positive ratings and 28.91% negative ratings across Polish-born participants, and 79.49% positive ratings and 15.01% negative ratings across English-born participants. Nonstandard past-tense BE agreement conditions are generally negatively evaluated among Polish-born participants, with only 25.23% positive ratings and 70.32% negative ratings. This is far more negative than English-born respondents' ratings, which are 55.27% positive and 38.42% negative. 40.99% of Polish-born participants' responses to nonstandard past-tense BE agreement test items are at the minimum level of acceptance (-3) compared to only 11.41% across English-born participants. The disparity in distributions of ratings between the two birth countries is far higher for nonstandard past-tense BE agreement conditions than it is for optional discourse-based movement ones.

Although not shown in Figure 4.2, I briefly touch upon the rating distributions of individual morphosyntactic conditions by birth country. Across Polish-born participants, the optional discourse-based movement condition of left dislocation has 74.67% positive responses and 18.86% negative, compared to 83.93% positive and 10.36% negative among English-born participants. Additionally, 41.14% of Polish-born responses to items containing left dislocation were at the maximum level of acceptance similar to the 40% across English-born respondents. Topicalisation has 63% positive and 31.5% negative responses among Polish-born participants, compared to 68.6% positive and 25.47% negative responses among English-born participants. The right dislocation patterns are even more different from the other two optional discourse-based movement conditions, according to birth country, with 58.29% positive and 36% negative responses among Polish-born participants, compared to 87.5% positive and only 7.86% negative responses among English-born participants. Finally, across the nonstandard past-tense BE agreement constructions, the disparity in patterns of responses by country of birth is far larger than for the other types of morphosyntactic construction. Nonstandard *weren't* has 24.67% positive and 71.17% negative responses across Polish-born participants compared to 64.84% positive and 29.06% negative across English-born. Nonstandard *were* has 28% positive and 54.27% negative responses among Polish-born participants compared to 51.25% positive and 42.03% negative responses among English-born participants. Nonstandard *was* has 22.86% positive and 72.76% negative responses among Polish-born participants, compared to 48.93% positive and 45% negative responses among English-born participants.

4.2.1.2 LGBTQ+ status

Before I discuss the patterning of AJT Responses by participant group, which hinges on the distinction between the LGBTQ+ and non-LGBTQ+ groups controlled for in this study, I first briefly touch on the sexuality labels employed across participants, and explain why the broader LGBTQ+ versus non-LGBTQ+ distinction is more appropriate than specific sexuality labels for the purposes of analysis. First, Table 4.2 shows the percentage distribution of participants across sexuality categories in this study, according to participant group. Participants provided the label they most identify with, but, in reality, participants may identify with more than one label. Two participants (both non-LGBTQ+) declined to specify a sexuality label and are hence *Unspecified*.

Group	Heterosexual (straight)		Homosexual (gay or lesbian)		Bisexual		Pansexual		Queer		Asexual or Aromantic		Unspecified	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
English-born LGBTQ+	-	-	14	35.00%	12	30.00%	5	12.50%	6	15.00%	3	7.50%	-	-
English-born non-LGBTQ+	38	95.00%	-	-	-	-	-	-	-	-	1	2.50%	1	2.50%
Polish-born LGBTQ+	-	-	16	44.44%	15	41.67%	2	5.56%	1	2.78%	2	5.56%	-	-
Polish-born non-LGBTQ+	37	94.87%	-	-	1	2.56%	-	-	-	-	-	-	1	2.56%
Total	75	48.39%	30	19.35%	28	18.06%	7	4.52%	7	4.52%	6	3.87%	2	1.29%

Table 4.2*. Distribution of participants across sexuality categories, by participant group*

By far the most frequent sexuality label is *Heterosexual*, with 48.39% (75) participants identifying this way. Given that roughly half of participants are non-LGBTQ, this is expected (though, one English-born non-LGBTQ+ participant does identify as *Asexual or Aromantic* and one Polish-born non-LGBTQ+ participant identifies as *Bisexual*). All *Heterosexual* participants are non-LGBTQ+, which means there are no participants who overlap between heterosexuality and the LGBTQ+ community (eg. straight transgender people). Across LGBTQ+ participants, several sexuality labels are employed; *Bisexual* is the third most common label after *Heterosexual* and *Homosexual*, followed by *Pansexual* and *Queer*, and *Asexual or Aromantic*. The number of participants in the top three most common labels (*Homosexual*, *Heterosexual*, and *Bisexual*) are split roughly equally by country of birth. This is not as much the case in the remaining labels where, for instance, English-born participants identifying primarily as *Queer* outnumber Polish-born 6 to 1.

As detailed in Section 4.1.2, LGBTQ+ status was chosen over the Sexuality variable for inclusion in further analysis. This is primarily because Sexuality was determined through Likelihood Ratio testing to not have a main effect on AJT Response ($p = .385$), nor a significant interaction with AJT Condition ($p = 5.184$), whereas LGBTQ+ status does significantly interact with AJT Condition ($p = <.001$). Furthermore, because LGBTQ+ participants are spread widely across a range of sexuality labels while non-LGBTQ+ participants are mostly *Heterosexual*, it is necessary to group the LGBTQ+ participants in order to compare their AJT Responses against non-LGBTQ+ participants anyway, meaning that a comparison between LGBTQ+ versus non-LGBTQ+ status is most fitting. Most importantly, because the primary focus of this investigation is on LGBTQ+ community membership, this comparison is most relevant to the study aims.

Figure 4.3 shows AJT responses by participant group, according to the three types of morphosyntactic construction.

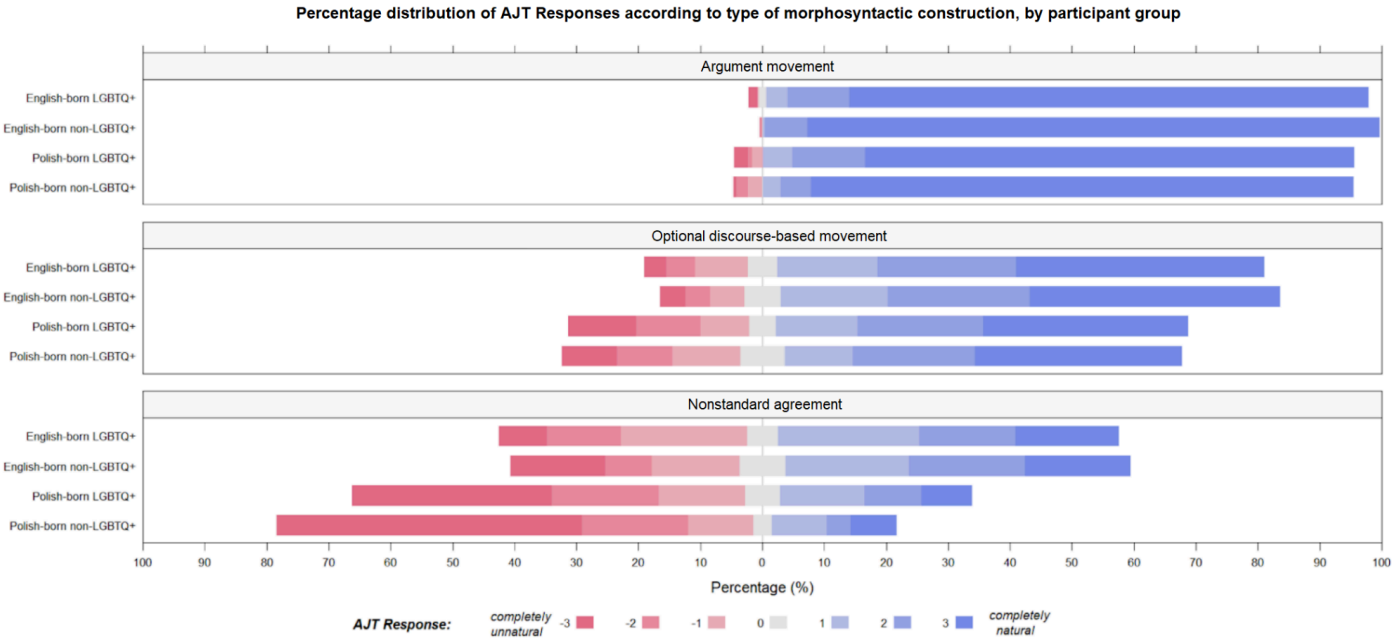


***Figure 4.3**. Distribution (%) of AJT Responses according to type of morphosyntactic construction, by participant group*

By participant group, we can see that the positive control condition of argument movement (raising-to-subject with *seem*) is rated very similarly across all four groups. Optional discourse-based movement is rated similarly regardless of LGBTQ+ status, the biggest distinction being by country of birth, as already discussed. The biggest difference by participant group is in the ratings of nonstandard past-tense BE agreement. Here, LGBTQ+ status has little impact on English-born participants' ratings, however, we see a difference by LGBTQ+ status among the Polish-born participants. Responses from Polish-born LGBTQ+ participants are 30.92% positive and 63.29% negative, while responses from Polish-born non-LGBTQ+ participants are only 20.07% positive and 76.81% negative.

4.3 Macro-social factors

Next, I turn to the macro-social factors of interest to the present study. These consist of age, gender, region, and the socioeconomic factors of socioeconomic status and education status. See Chapter 2 for an overview of the relevance of each of these factors to this study. The percentage distribution of AJT Responses across each of these independent variables will be discussed according to the three types of morphosyntactic construction - argument movement, optional discourse-based movement, and nonstandard agreement), however, as argument movement test items pattern very similarly across all of these variables (ie. they are almost unanimously accepted), discussion of this category will be omitted, unless particularly relevant.

4.3.1 Age

Table 4.3 shows the distribution of participants across age bands, according to the participant groups in this study.

Group	18-24		25-30		31-40		41 +	
	N	%	N	%	N	%	N	%
English-born LGBTQ+	13	32.50%	14	35.00%	9	22.50%	4	10.00%
English-born non-LGBTQ+	8	20.00%	10	25.00%	10	25.00%	12	30.00%
Polish-born LGBTQ+	13	36.11%	7	19.44%	11	30.56%	5	13.89%
Polish-born non-LGBTQ+	5	12.82%	8	20.51%	18	46.15%	8	20.51%
Total	39	25.16%	39	25.16%	48	30.97%	29	18.71%

***Table 4.3**. Distribution of participants across age bands, by participant group*

Other than requiring participants to be over the age of 18, there were no age-related constraints or controls on the participant sample. Accordingly, the resulting age distribution of participants is skewed towards the young adult end of the age range; 81.3% (126/155) participants are aged 40 or under, and only 18.71% (29/155) are 41 or older. Although 31-40 is the most common age band (with 49 participants), the two younger (and narrower) age bands 18-24 and 25-30 account for 40 and 39 participants, respectively. The four participant groups are roughly well represented within the youngest three age bands, although there is a larger cluster of 31-40 year old Polish-born non-LGBTQ+ participants compared to other age bands for this group. The distribution shows a tendency towards younger LGBTQ+ participants, with a total of 26 LGBTQ+ 18-24 year-olds and only 13 non-LGBTQ+, and fewer older LGBTQ+ participants compared to non-LGBTQ+ participants. Within each age band, participants are, for the most part, similarly split by country of birth (eg. within the 18-24 band, there are 21 English-born and 18 Polish-born participants). Some age bands are less balanced by country of birth, however; Within the 25-30 band, there are more English-born (24) than Polish-born (15) participants and, conversely, within the 31-40 band, there are more Polish-born (29) than English-born (19) participants.

Figure 4.4 shows AJT responses by age band, according to the three types of morphosyntactic construction.

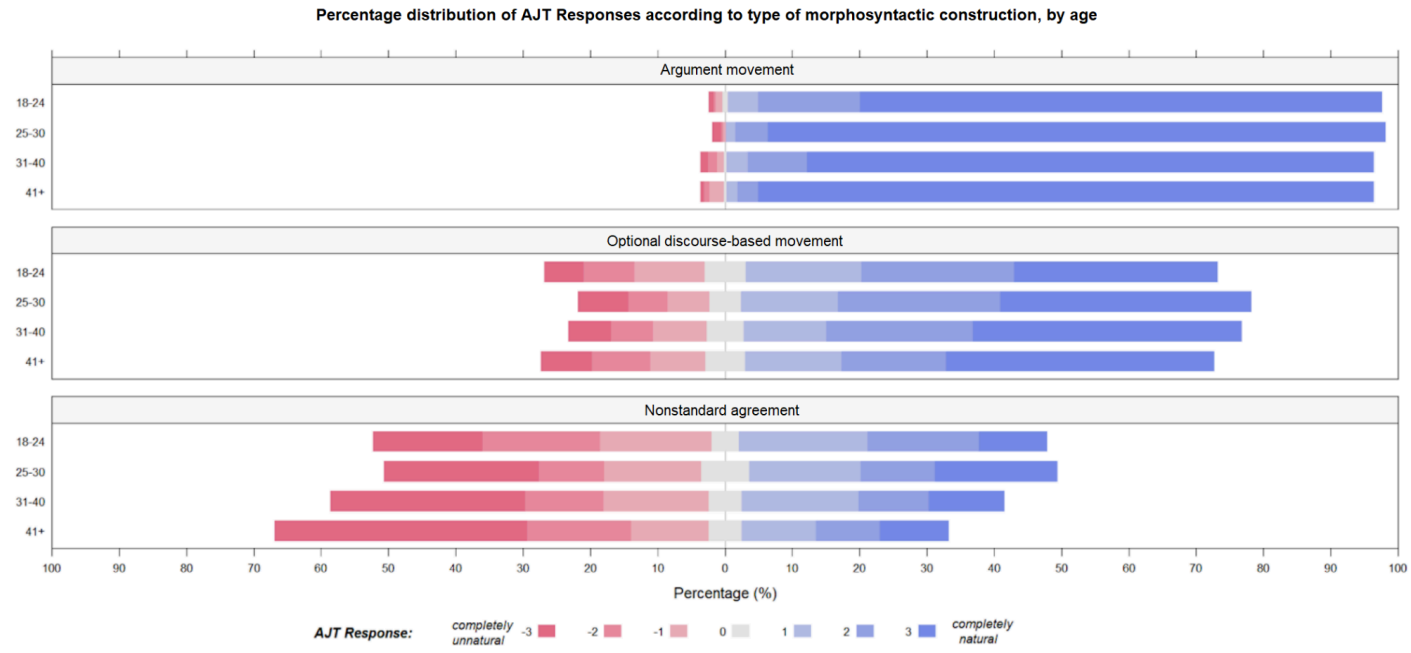


Figure 4.4*. Distribution (%) of AJT Responses according to type of morphosyntactic construction, by age band*

Assessing the patterning of AJT Responses according to participant age bands, we can see that age seems to have an effect on nonstandard past-tense BE agreement, with older participants generally rating these conditions less favourably than younger ones. The eldest age band (*41+*) contains the largest proportion of negative responses to nonstandard past-tense BE agreement - 64.32% of all responses, with 37.33% of all responses at the minimum increment of the scale (-3 on the AJT Response scale). Optional discourse-based movement does not seem to pattern consistently by age band, though the proportion of maximal responses (3 on the AJT Response scale) is highest for the eldest age band (*41+*), with 39.81% responses falling within this increment.

4.3.2 Gender

Table 4.4 shows the distribution of participants across gender categories in this study, according to participant group. One LGBTQ+ participant declined to specify their gender category and has hence been labelled as *Unspecified*.

Group	Female		Male		Not binary		Unspecified	
	N	%	N	%	N	%	N	%
English-born LGBTQ+	20	50.00%	13	32.50%	6	15.00%	1	2.50%
English-born non-LGBTQ+	25	62.50%	14	35.00%	1	2.50%	-	-
Polish-born LGBTQ+	14	38.89%	18	50.00%	4	11.11%	-	-
Polish-born non-LGBTQ+	27	69.23%	12	30.77%	-	-	-	-
Total	86	55.48%	57	36.77%	11	7.10%	1	0.65%

Table 4.4*. Distribution of participants across gender categories, by participant group*

The majority of participants - 86 (55.48%) participants - who responded to this study are female. The only group for which women do not outnumber men is Polish-born LGBTQ+ participants. Despite this, though, considering gender was not controlled for, the proportion of participants who are male - 57 (36.77%) - is surprisingly high. Although effort was made to advertise the study in gender-diverse spaces, only 11 participants (7.10%) fall outside the gender binary.[14] Unsurprisingly, almost all *Not binary* participants identify as LGBTQ+, except for one English-born participant.

Figure 4.5 shows AJT responses by gender, according to the three types of morphosyntactic construction.

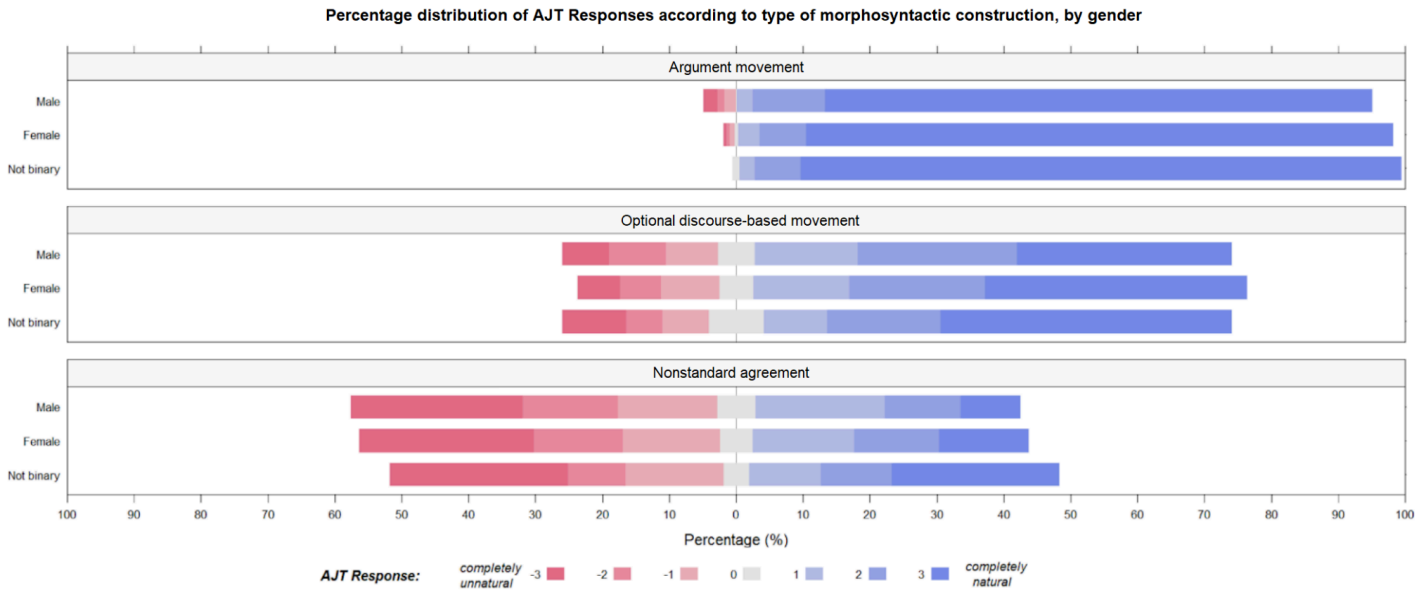


Figure 4.5*. Distribution (%) of AJT Responses according to type of morphosyntactic construction, by gender category*

Assessing the patterning of AJT Responses according to gender categories, we can see a potential effect of gender. The proportion of positive and negative responses across the three gender categories are very similar for optional discourse-based movement, however, participants who are *Not binary* have a higher proportion of maximal acceptance ratings (ie. 3 on the AJT Response scale) compared to the binary gender categories, with 43.39% of their responses falling within this increment of the scale. These participants also have a comparatively more positive distribution of responses to nonstandard past-tense BE agreement compared to male and female respondents (ie. the stacked bar is further towards the right of the graph). Specifically, *Not binary* participants provide 46.25% positive and 49.8% negative responses to nonstandard past-tense BE agreement, compared to 41.2% positive and 53.79% negative responses across female participants, and 39.44% positive and 54.69% negative responses across male participants. *Not binary* participants also have the largest proportion of maximal responses to these conditions, with 24.9% of responses of a 3 on the AJT Response scale. Finally, *Not binary* is also the only gender category for which no negative responses were provided for argument movement. There are, however, far fewer participants in the *Not binary* category, therefore, further statistical modelling is required to explore the response patterns of *Not binary* participants.

[14] Not all of the *Not binary* participants identify straightforwardly with the nonbinary label. See Section 3.8.1.1 for a discussion of the labelling choices for the *Not binary* gender category

4.3.3 Region of residence

Table 4.5 shows the distribution of participants across grouped region categories in this study, according to participant group. Based on postcode information, participants were binned into three deliberately very broad regions of current residence (see Section 3.8.1.1 for justification of this). Two participants (both non-LGBTQ+) did not specify their postcode and therefore, their region of residence is *Unspecified*.

Group	North		Midlands		South		Unspecified	
	N	%	N	%	N	%	N	%
English-born LGBTQ+	14	35.00%	6	15.00%	20	50.00%	-	-
English-born non-LGBTQ+	21	52.50%	8	20.00%	10	25.00%	1	2.50%
Polish-born LGBTQ+	11	30.56%	3	8.33%	22	61.11%	-	-
Polish-born non-LGBTQ+	13	33.33%	3	7.69%	22	56.41%	1	2.56%
Total	59	**38.06%**	20	**12.90%**	74	**47.74%**	2	**1.29%**

***Table 4.5**. Distribution of participants across grouped region categories, by participant group*

Of the three regions that England was divided into, the region within which the most participants (74 participants, or 47.74%) currently reside is the South. 30 of these participants are from Greater London, 17 are from the South East, 12 from the South West, and 15 from the East (which have been grouped in with the South as discussed in Section 3.8.1.1). 59 participants (38.06% reside in the North (44 in Yorkshire & the Humber, 14 in the North West, and 1 in the North East). Significantly fewer (only 20 participants, or 12.9%) reside in the Midlands, with 11 in the East Midlands and 9 in the West Midlands. Across all three broad regions, roughly equal numbers of participants are LGBTQ+ and non-LGBTQ+. In the North and South, representation across the two countries of birth is also adequately balanced, however, in the Midlands, Polish-born participants outnumber English-born roughly 2 to 1. This is not an issue for descriptive analysis, though it does mean that results concerning the distribution of AJT Responses in the Midlands must be assessed further according to country of birth. However, this disparity did not cause convergence errors in the statistical modelling process reported in Chapter 5.

I also briefly turn to the data on participants' intra-national mobility (shown in Table 4.6) in order to provide some potential explanation for these patterns. To assess how mobile the participants in this sample have been in the past, data was collected regarding participants' history of residence in England, and coded according to whether they have lived outside their current broad region of residence (North, Midlands, or South).

Group	No		Yes	
	N	%	N	%
English-born	23	28.75%	57	71.25%
Polish-born	55	73.33%	20	26.67%
Total	78	50.32%	77	49.68%

***Table 4.6**. Distribution of participants according to whether they have lived outside of their current region of residence, by participant group*

This reveals a clear divide between Polish-born and English-born participants; 73.3% (55/75) Polish-born participants have *not* lived outside of their current region of residence, while 71.25% (57/80) English-born participants *have*. 7 participants have lived in all three of the grouped regions, and only 2 of them are Polish-born. The English-born participants will naturally have had longer to become familiar with different regions of the country. For instance, they will likely have moved to different region(s) for higher education (which most participants have completed some form of - see Section 4.2.4.2).

Figure 4.6 shows AJT responses by region of residence, according to the three types of morphosyntactic construction.

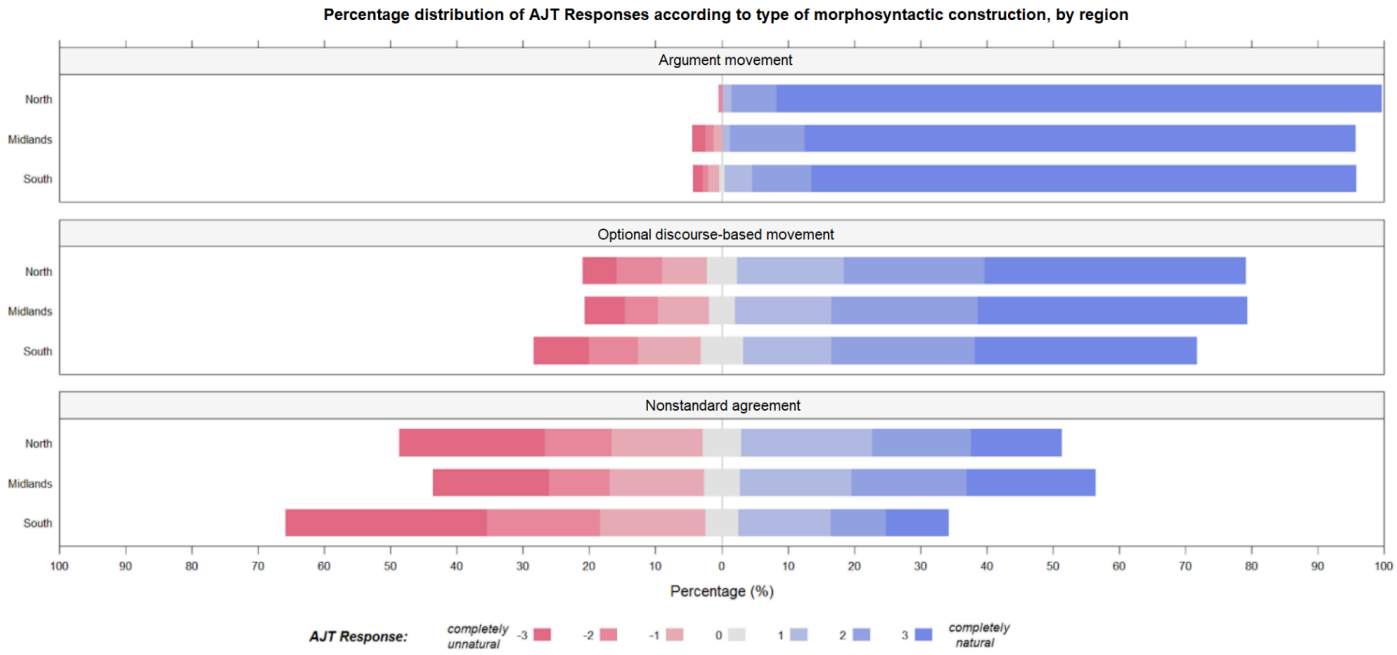


***Figure 4.6**. Distribution (%) of AJT Responses according to type of morphosyntactic construction, by current English region of residence*

Assessing the patterning of AJT responses according to grouped English region of current residence, participants in the *South* category have the most negative (ie. left-most) range of responses to both optional discourse-based movement and nonstandard past-tense BE agreement. Southern participants have 68.37% positive and 25.06% negative responses to optional discourse-based movement conditions, compared to participants in the North and Midlands, who have 76.66% and 77.27% positive, and 18.57% and 18.64% negative responses, respectively. For nonstandard past-tense BE agreement, participants in the Midlands have the most positive range of responses (ie. the right-most bar), with 53.7% positive and 40.87% negative responses, as well as having the lowest proportion of minimal responses (of -3) and highest proportion of maximal responses (of 3) compared to the other two regions. The South is by far the least accepting region of nonstandard past-tense BE agreement among the participant sample, having the fewest responses of 3 and the most of -3. Only 31.67% of Southerners' AJT Responses to nonstandard past-tense BE agreement are positive, and 63.28% are negative, compared to Northerners, who have 48.34% positive and only 45.69% negative responses. Narrowing this down according to country of birth (not visualised here), this pattern is consistent among Polish-born participants, who, in the Midlands, also evaluate nonstandard past-tense BE agreement more positively than in other regions, with 39.13% positive responses and 57.25%

negative responses, compared to 25.2% positive and 69.47% negative in the South, and 22.64% positive and 74.28% negative the North, respectively. Additionally, the Midlands region has the highest relative proportion of maximal responses of 3 (25.36%) and lowest proportion of minimal responses of -3 (23.91%) to nonstandard past-tense BE agreement across Polish-born participants. Poles in the North, on the other hand, have the lowest proportion of maximal responses (4.53%) and highest proportion of minimal responses (46.56%) to nonstandard past-tense BE agreement across all regions. Further inferential statistical modelling is necessary to assess the statistical power of these effects and to explore these patterns further.

4.3.4 Socioeconomic factors

This study measured several factors which are indicative of socioeconomic characteristics, namely Indices of Multiple Deprivation (IMD) Decile (measured from postcode information) and highest level of education. The significance of these to the present study is described in Section 2.3.2. The distributions of participants across categories of these factors, as well as the distributions of AJT Responses across these factors are described below.

4.3.4.1 Socioeconomic status

Socioeconomic status was measured using the UK Government's Indices of Multiple Deprivation (IMD) metric based on participant postcodes. This measure takes into account the postcode's deprivation with regards to: Income; Employment; Education, Skills and Training; Health and Disability; Crime; Barriers to Housing and Services; and Living Environment (see Section 3.8.1.1 for more detail about this). The IMD Decile is a 10-point ordinal measure, which has been binned here into 3 categories, with *Low* representing the 3 relatively most deprived IMD Deciles, *Medium* representing the next 3 most deprived IMD Deciles, and *High* representing the 4 relatively least deprived IMD Deciles. Table 4.7 shows the distribution of the participant sample by socioeconomic status, according to participant group. Three participants did not provide postcode information, so are *Unspecified*.

Group	Low		Medium		High		Unspecified	
	N	%	N	%	N	%	N	%
English-born LGBTQ+	12	30.00%	13	32.50%	15	37.50%	-	-
English-born non-LGBTQ+	10	25.00%	15	37.50%	13	32.50%	2	5.00%
Polish-born LGBTQ+	16	44.44%	8	22.22%	12	33.33%	-	-
Polish-born non-LGBTQ+	14	35.90%	14	35.90%	10	25.64%	1	2.56%
Total	**52**	**33.55%**	**50**	**32.26%**	**50**	**32.26%**	**3**	**1.94%**

Table 4.7*. Distribution of participants by socioeconomic status (derived from IMD Decile measures), by participant group*

Table 4.7 shows that participants are more or less evenly distributed by socioeconomic status, with 52, 50, and 50 participants in the *Low*, *Medium*, and *High* categories, respectively, as well as by participant group.

Figure 4.7 shows AJT responses by socioeconomic status, according to the three types of morphosyntactic construction.

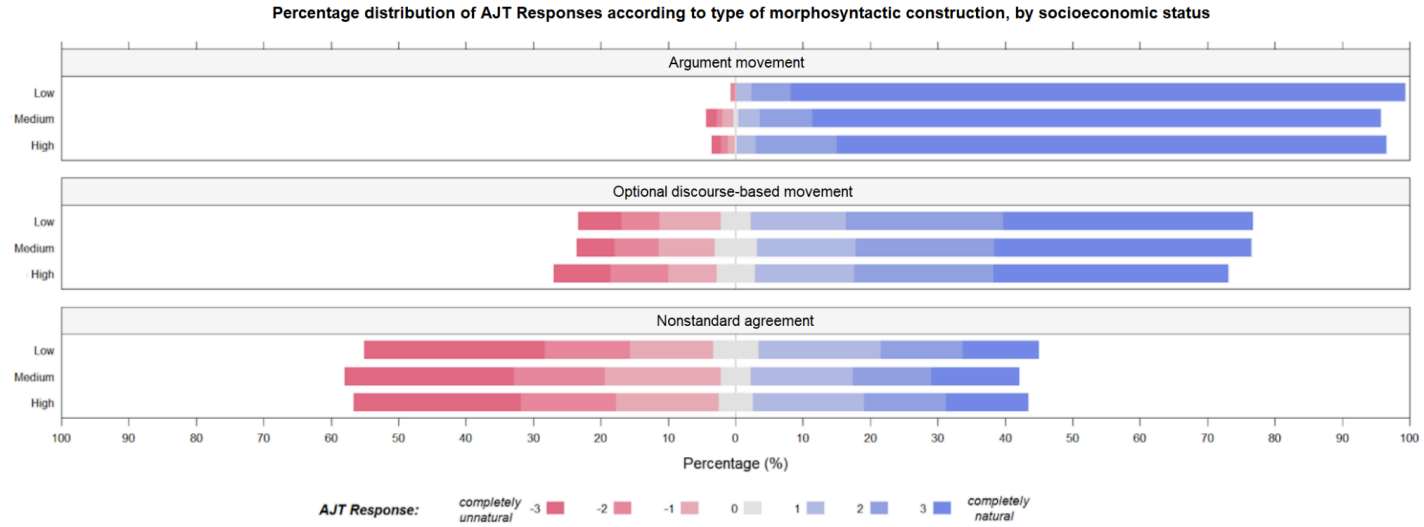


***Figure 4.7**. Distribution (%) of AJT Responses according to type of morphosyntactic construction, by socioeconomic status band*

Across the three socioeconomic status bands, the percentages of negative responses to optional discourse-based movement range between 20.36%-24% and positive responses range between 70.01%-74.39%, and, for socioindexical conditions, the percentages of negative responses range between 51.67%-55.74% and positive responses range between 39.74%-41.56%. Based on these results, in this participant sample, socioeconomic status does not seem to have a large effect on participants' acceptance of the different types of morphosyntactic construction.

4.3.4.2 Higher-educated status

As detailed in Section 3.8.1.1, participants were originally sorted into one of four ordinal categories according to the highest level of education they have completed: Higher Education postgraduate qualification (master's degree, PhD, etc.); Higher Education graduate qualification (eg. bachelor's degree, Certificate of HE); Further Education qualification (A-Level, NVQ, etc.); Secondary school qualification (GCSE, etc.), and this was then re-coded to compare participants by whether they had or had not completed some level of higher education. Table 4.8 shows the distribution of participants according to higher-educated status.

Group	Yes		No	
	N	%	N	%
English-born LGBTQ+	33	82.50%	7	17.50%
English-born non-LGBTQ+	34	85.00%	6	15.00%
Polish-born LGBTQ+	28	77.78%	8	22.22%
Polish-born non-LGBTQ+	31	79.49%	8	20.51%
Total	126	81.29%	29	18.71%

***Table 4.8**. Distribution of participants according to higher-educated status, by participant group*

Because the study was advertised through many academic and education-focused networks, and the nature of the research matter likely made it appealing to students and graduates, the distribution of participants according to education status is strongly skewed towards the higher levels of education. 81.29% (126) participants have exposure to higher education. 72 of these (ie. 46.5% of all participants) have completed some level of postgraduate study. 48 hold bachelor's degrees, 5 hold foundation degrees and one has a Certificate of Higher Education. 27 participants have attained further education qualifications, 24 of which have completed sixth form/college and hold AS/A-Levels or equivalent qualifications, and 3 have completed a vocational qualification such as an NVQ or apprenticeship. In contrast, only two participants ended their education at secondary/high school. Due to this skew in educational attainment across the participant sample, participants are instead categorised by higher-educated status, ie. whether or not they have completed some form of higher education qualification. Although this is still unbalanced in that far more participants are higher educated than not higher educated, participants are distributed roughly evenly within the higher-educated versus non higher-educated groups according to both country of birth and LGBTQ+ status. Furthermore, the disparity between the two groups did not cause convergence errors in the statistical modelling process.

Figure 4.8 shows AJT responses according to higher-educated status, according to the three types of morphosyntactic construction.

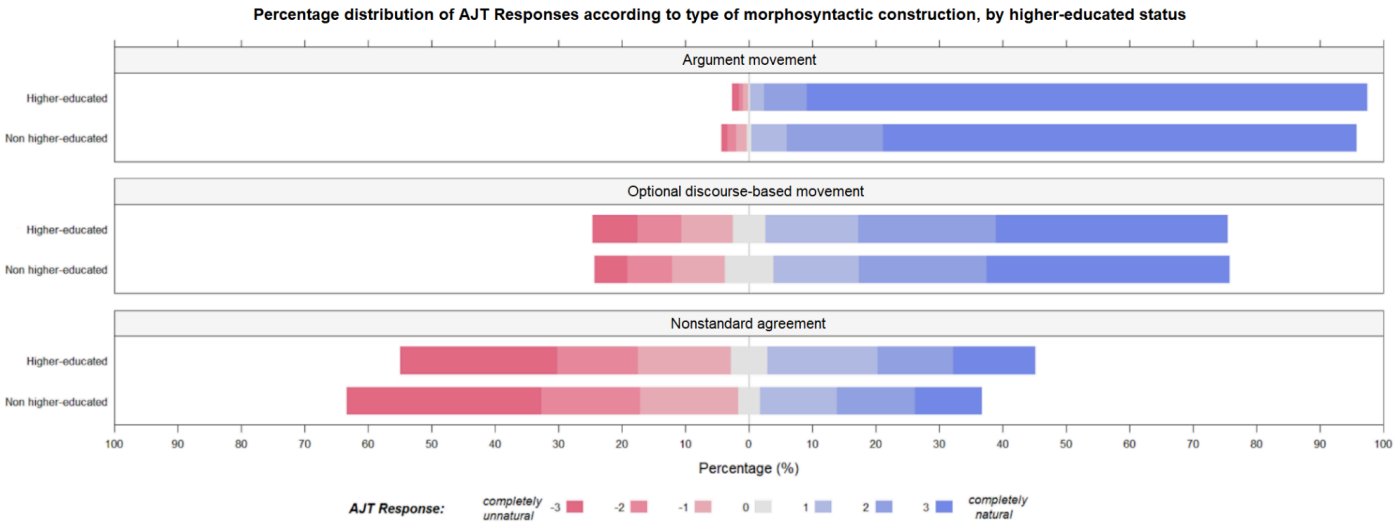


Figure 4.8*. Distribution (%) of AJT Responses according to type of morphosyntactic construction, by higher-educated status*

Assessing the percentage distribution of AJT Responses by higher-educated status in Figure 4.8, participants' involvement in higher education does not seem to impact their rating of argument movement or optional discourse-based movement, but seems to have an effect on ratings of nonstandard past-tense BE agreement. Higher-educated participants have a more positive range of AJT Responses to nonstandard past-tense BE agreement (ie. the stacked bar is further right) and have 42.13% positive and 52.07% negative responses compared to non higher-educated participants, who have 34.78% positive and 61.62% negative responses. However, inferential statistical methods are required to assess whether this effect is indeed significant.

4.4 L2 factors

In addition to the macro-social factors explored in the prior section, several key factors related to the use of English as a second language (L2) warrant consideration, namely Polish-born participants' age of arrival (ie. age of migration) to England, and their English language proficiency (including their English Language Exposure, and further contextualised by the types of formal English language instruction they have undergone). The following section provides an overview of the distributions of participants according to the categories of each of these factors and the patterning of AJT Responses across them, too, by the three types of morphosyntactic construction. This section only presents data from Polish participants.

4.4.1 Age of Arrival to England

Table 4.9 shows the distribution of Polish-born participants according to their age band when they began living in England (see Section 3.8.1.2 for justification of these age bands).

Group	7-18		19-24		25+	
	N	%	N	%	N	%
Polish-born LGBTQ+	9	25.00%	15	41.67%	17	47.22%
Polish-born non-LGBTQ+	4	10.26%	18	46.15%	17	43.59%
Total	**13**	**17.33%**	**33**	**44.00%**	**29**	**38.67%**

Table 4.9*. Distribution of Polish-born participants across migration age bands, by LGBTQ+ status*

Of the 75 Polish-born participants in this study, 44% (33) participants moved to England aged 19-24, this being the most common migration age band, followed by 38.67% (29) participants moving at age 25 or older. No participants moved to England after the age of 40. Also, comparatively very few moved in the early stages of life; Only 7 participants moved during adolescence (13-18 years of age), and only 6 participants moved during childhood years (ages 7-12), together totalling 13 participants (or 17.33% of all Polish-born participants) who moved during or prior to the critical stage of adolescence (ie. at age 18 or younger). For the purposes of ensuring the participant sample is balanced, I note here that the migration age bands are adequately balanced by Polish-born participants' LGBTQ+ status.

As an additional informative aside to the information presented in Table 4.9, of the Polish-born participants in this study, 32% (24) participants have been living in England for 11-15 years, this being the most common band, followed by 21.3% (16 participants) living in England for 6-10 years and 17.3% (13 participants) for 3-5 years. Only a few Polish-born participants are very recent migrants; 12% (9 participants) had been living in England for 2 years or fewer at the time of data collection. This distribution of Polish-born participants allows ample scope for exploration of the effects of long term migrants' acculturation patterns.

Figure 4.9 shows the patterning of AJT responses with regards to Polish-born participants' age of migration to England.

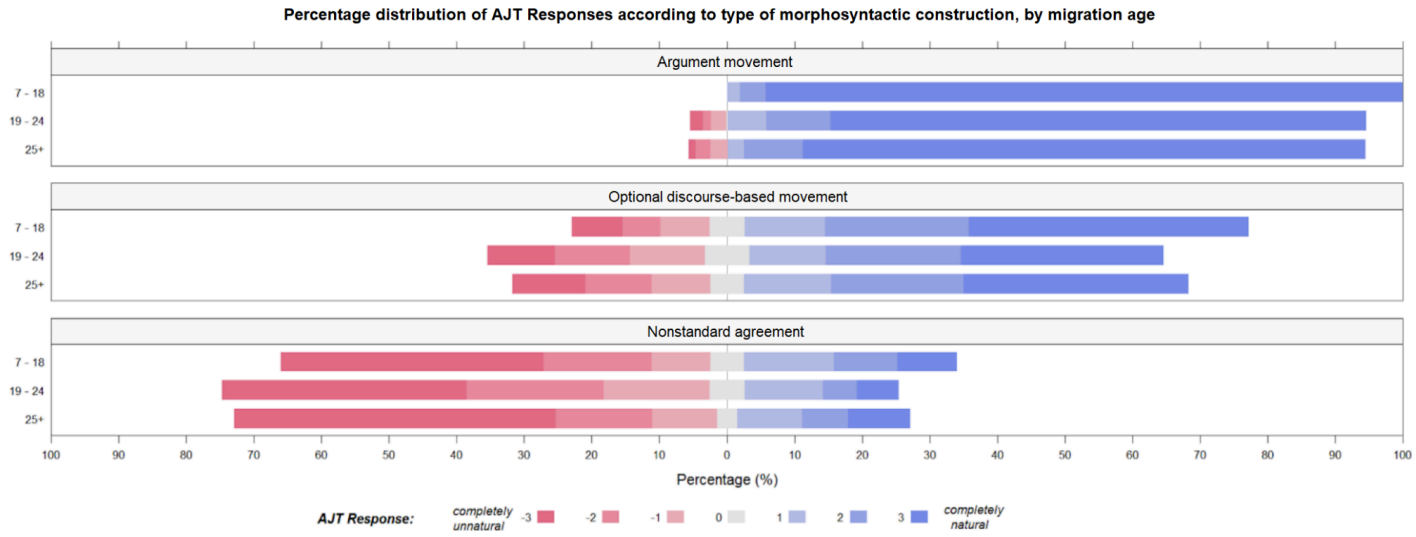


***Figure 4.9**. Distribution (%) of AJT Responses according to type of morphosyntactic construction, by migration age band*

It seems that participants who migrated during childhood and adolescence (ie. the *7-18* band) are more positive in their AJT evaluations across the board. Those who migrated between *7-18* are the most positive about optional discourse-based movement (at 74.45%), and the least negative (at only 20.28%). Similarly, they are the most positive about nonstandard past-tense BE agreement (at 31.44%) and the least negative (at 22.66%). Additionally, the *7-18* band is the only band whose ratings of argument movement are entirely positive.

However, it is important to note that the data in this band is based on a relatively small number of participants (only 17.33% of all Polish-born participants), which may be skewing results. This, combined with the fact that participants in the *25+* migration age band appear more accepting of optional discourse-based movement and nonstandard past-tense BE agreement than those in the *19-24* band (ie. the stacked bars are positioned relatively more rightwards), calls into question whether participants' age of migration indeed influences participants' acceptability ratings. In order to get a clearer picture of the degree to which L2 factors might be impacting participants' AJT performance, I instead turn to factors more directly indicative of participants' levels of English proficiency.

4.4.2 English Language Proficiency

The primary measure for English language proficiency in this study was the Versant English Speaking test (see Section 3.4 for justification of the use of this method). However, due to the limited availability of Versant data across Polish-born participants (only 51 of the 75 Polish-born participants completed this test), linguistic proficiency is also explored through the proxy of participants' English language exposure (in years), for which data is available across all Polish-born participants. The methods through which participants across the sample have acquired English is also considered below.

4.4.2.1 English Language Instruction

Here, a general overview is provided of the types of English language instruction undergone by Polish-born participants, giving some additional context about the participant sample. This is shown in Table 4.10. Participants could indicate multiple forms of instruction if applicable.

Type of English instruction undertaken	N	% of Polish-born participants
English language lessons as part of the school curriculum	64	85.33%
English language courses at a language school	38	50.67%
Self-motivated formal English language study (eg. through textbooks/online courses)	33	44.00%
No formal English instruction	6	8.00%
Other English language-focused course/training	5	6.67%
Private English Language lessons	4	5.33%

***Table 4.10**. Distribution of Polish-born participants by English language instruction undergone*

Participants in this sample have been quite well-roundedly exposed to formal English language instruction. Only 6 of the 75 Polish-born participants (8%) have had zero formal English instruction, meaning 92% have undergone formal English language instruction of some nature. This is in line with the overall skew towards high English proficiency levels, which will also be shown in Section 4.3.2.3. By far the most common source of exposure to formal English instruction is through English language lessons as part of the school curriculum, with 85.33% (64 of 75) Polish-born participants engaging with this to some extent. 50.67% (38 of 75) participants have undergone English language courses at a language school and 44% (33 of 75) participants have engaged in self-motivated formal English language study using resources such as textbooks or online courses. Finally, 5.33% (10 of 75) participants have engaged in private English language lessons or some other English language-focused training. Although information is not available about the quality or length of time participants had invested in these types of instruction, most (61.3%, or 46 of the 75) participants have had exposure to formal English instruction through more than one of the methods listed in Table 4.10, above: 25 participants listed three different methods, 19 listed two, and 2 participants listed four, while 29 participants (38.7%) listed only one source of English instruction. Not only does this information help contextualise participants' English language-learning backgrounds, it also confirms that the vast majority of participants will have had at least some level of exposure to prescriptive English grammar norms.

4.4.2.2 English Language Exposure (in years)

An expected co-linear variable with (and, hence, proxy measurement of) Polish-born participants' English proficiency is the amount of exposure they have had to English input. This data reports the number of years since Polish participants began learning English, and is binned into 5 bands. The percentage distribution of Polish-born participants across these bands is shown in Table 4.11.

Group	3-10 years		11-15 years		16-20 years		21-30 years		31+ years	
	N	%	N	%	N	%	N	%	N	%
Polish-born LGBTQ+	2	5.56%	10	27.78%	10	27.78%	11	30.56%	3	8.33%
Polish-born non-LGBTQ+	3	7.69%	6	15.38%	10	25.64%	14	35.90%	6	15.38%
Total	**5**	**6.67%**	**16**	**21.33%**	**20**	**26.67%**	**25**	**33.33%**	**9**	**12.00%**

Table 4.11*. Distribution of Polish-born participants by LGBTQ+ status, according English Language Exposure (in years)*

Since the participants in this study are expected to be relatively proficient speakers of English, it is not surprising that most Polish-born participants have had a substantial number of years of English input. Only 6.67% (ie. 5 of the 75 Polish-born participants) have been learning English for 10 or fewer years. The vast majority - 81.3%, or 61 of 75 participants - have between 11 and 30 years of English exposure, with the most common band being *21-30 years* (which accounts for 33.3%, or 25 of 75 participants). There are also relatively few participants with over 31 years of English language exposure (only 9 out of 75, or 12%). This is unsurprising given that the age distribution of the participant sample skews towards younger participants (see Table 4.3). Overall, however, the distribution of amounts of exposure to English language instruction amongst Polish-born participants is fairly broad and, within each band, participants are roughly evenly divided by LGBTQ+ status, meaning neither group is relatively over- nor under-represented for the purposes of statistical modelling.

I now turn to the patterning of AJT responses with regards to Polish-born participants' length of English language exposure (Figure 4.10).

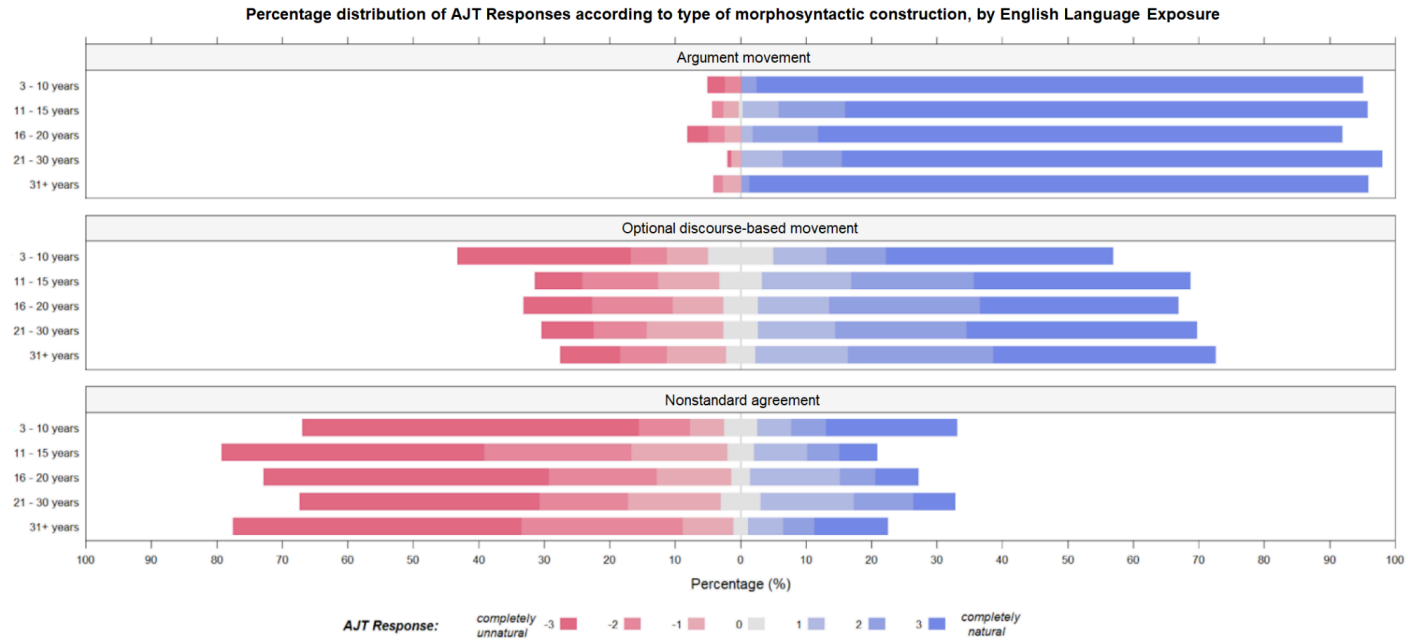


Figure 4.10. *Distribution (%) of AJT Responses according to type of morphosyntactic construction, by English Language Exposure (in years)*

It seems that the length of English language exposure correlates positively with acceptance of optional discourse-based movement - ie. For this condition, the stacked bars generally trend rightwards with increasing exposure. Those in the lowest band of *3-10 years* exposure have a noticeably lower proportion of positive responses to optional discourse-based movement compared to other bands, however, data in this band derives from only 5 participants which could be inflating these results.

The effect of language exposure on nonstandard past-tense BE agreement is not as clear as that on optional discourse-based movement. The middle three bands of English language exposure indicate a rightwards trend, towards higher acceptance with increasing exposure, however, the lowest and highest bands (*3-10 years* and *31+ years*, respectively) do not align with this. These two bands contain the fewest number of participants (5 and 9, respectively). These low numbers could be skewing the results, meaning further statistical modelling is required to assess this. To investigate the effect of English proficiency on nonstandard past-tense BE agreement, as well as on other conditions, further, I now turn to results from the Versant English Speaking Test.

4.4.2.3 English Proficiency (*CEFR*)

A more formal measure of English language proficiency was conducted using the automated Versant English Speaking Test (the method for which is outlined in Section 3.4), and this was completed by 51 of the 75 Polish-born participants (28 Polish-born LGBTQ+ and 23 Polish-born non-LGBTQ+ participants). To recap, the test generates an overall score incorporating four domains of English proficiency: *Sentence Mastery* measures sentence construction and

comprehension, *Vocabulary* measures passive and active vocabulary use, *Fluency* measures phonological fluency, and *Pronunciation* measures the pronunciation of rhythmic and segmental units, both lexical and phrasal (Pearson, 2011). Prior to analysis, participants' raw Versant scores have been converted into their equivalent scores on the *Common European Framework of Reference* (*CEFR*) classification - a widely recognised 6-way categorical scale measuring English proficiency, which splits speakers into *Basic* (levels A1 and A2), *Independent* (levels B1 and B2), and *Proficient* (levels C1 and C2) users of English (see Section 3.8.1.2 for more details on this). As explained in Section 3.8.1.2, although raw Versant scores attained by participants are reasonably wide-ranging, falling between 44 and 80 (of a total possible range of 20 to 80), this sample is, overall, skewed towards highly proficient users of English, with the mean Versant score being 69.2. On the *CEFR* scale, only 3.9% (2 participants) participants fall under *Basic user* (A2), 39.2% (20 participants) fall under *Independent user* (B1 and B2), and 56.9% (29 participants) come under *Proficient user* (C1 and C2). Of the *Proficient users*, 9 participants achieved the maximum possible score of 80 and no participants placed in the lowest *Basic user* category 'A1'. This distribution is not unexpected given the majority of Polish-born participants are not recent migrants (as outlined in Section 4.3.1) and, as discussed in Section 4.3.2.2, most have had a substantial length of exposure to English input.

As outlined in Section 3.8.1.2, this disparity in distribution of participants across *CEFR* proficiency bands necessitated the collapsing of these categories prior to analysis, to ensure more equal weighting and statistical model convergence. Thus, the *Basic* and *Independent* users (ie. *CEFR* categories *A2*, *B1* and *B2*) have been grouped together and compared against the *Proficient* users (ie. *CEFR* categories *C1* and *C2*). The distribution of participants across grouped CEFR classification is shown in Table 4.12.

Group	Basic and Independent Users (B2 or lower)		Proficient Users (C1 or C2)	
	N	%	N	%
Polish-born LGBTQ+	11	39.29%	17	60.71%
Polish-born non-LGBTQ+	11	47.83%	12	52.17%
Total	**22**	**43.14%**	**29**	**56.86%**

Table 4.12*. Distribution of Polish-born participants by English proficiency level (grouped CEFR classification), according to LGBTQ+ status*

43.14% (22 participants) are *Basic* or *Independent* users of English, and 56.86% (29 participants) are *Proficient*. Each category is sufficiently evenly split by LGBTQ+ status and this distribution is satisfactory for the purposes of statistical modelling.

Figure 4.11 shows the percentage distribution of AJT responses according to grouped *CEFR* classification.

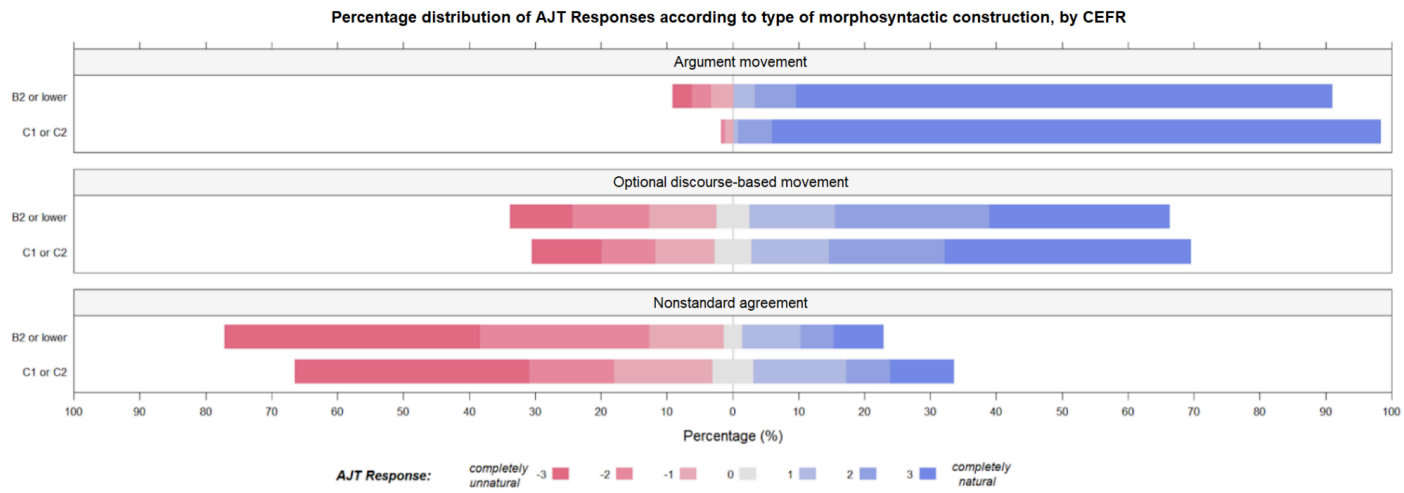


Figure 4.11. *Distribution (%) of AJT Responses according to type of morphosyntactic construction, by grouped CEFR classification (English language proficiency)*

Participants' English language proficiency positively impacts AJT acceptance across all three types of morphosyntactic construction - the stacked bars for the higher English proficiency levels are consistently further right than those for the lower proficiency levels. Even the positive control condition of argument movement is impacted by proficiency, with 90.91% positive and 9.1% negative responses among *Basic* and *Independent* users (*B2 or lower*), compared to *Proficient* users (*C1 or C2*), who have 98.28% positive and 1.72% negative responses. Optional discourse-based movement has 63.64% positive and 31.2% negative responses among *Basic* and *Independent* users, compared to 66.61% positive and 27.59% negative responses among *Proficient* users. Additionally, *Proficient* users have a higher percentage of maximal acceptability ratings for optional discourse-based movement, with 37.3% of responses being a 3 on the AJT scale, compared to 27.27% among *Basic* and *Independent* users. Nonstandard past-tense BE agreement has 21.34% positive and 75.69% negative responses among *Basic* and *Independent* users, and 30.28% positive and 63.27% negative responses among *Proficient* users. The difference in acceptance of nonstandard past-tense BE agreement between the two proficiency levels is larger than it is for optional discourse-based movement, suggesting that English language proficiency has a stronger effect on nonstandard past-tense BE agreement compared to optional discourse-based movement. However, inferential statistical methods are required to assess whether the effects found here remain significant when random individual and test item effects are negated.

4.5 Meso-social factors

Next, I turn to the meso-social factors of interest to the present study (ie. ones related to participants' degrees of community embeddedness). These consist of English and Polish acculturation levels, and the levels of LGBTQ+ Community Involvement among the LGBTQ+ participants. See Chapter 2 for an overview of the relevance of each of these factors to this study. As has been the case throughout this chapter, the percentage distribution of AJT Responses across each of these factors will be discussed according to the three types of morphosyntactic construction - argument movement, optional

discourse-based movement, and nonstandard agreement, however, as acceptability of argument movement (the raising construction with the verb *seem*) patterns very similarly across all of these variables (ie. they are almost unanimously accepted), discussion of this category will be omitted.

4.5.1 Acculturation

To recap from Section 3.5.2, English and Polish acculturation levels were measured using equivalent surveys. The acculturation survey items were devised so as to capture as many relevant aspects of life within each culture as possible, and were divided across three broad, equally-weighted sub-categories of acculturation: *Attitudes & Identity*; *Lifestyle*; and *Language use*. Overall acculturation scores were calculated by averaging scores across all the responses from the respective survey. To provide an overall snapshot of the acculturation results, this section will first explore the median scores for English and Polish acculturation, as well as for the three sub-categories for each of these, according to country of birth and LGBTQ+ status. This will be followed by an assessment of the percentage distributions of participants across the three bands of acculturation for each acculturation measure - *Low*, *Medium*, and *High*, using only the *Lifestyle* sub-category instead of the total acculturation score (see Section 3.8.1.2 for discussion of how these bands were derived, and see Section 3.9.2.3 for justification of the use of only the *Lifestyle* sub-category as a measure of acculturation in the analysis). Finally, the percentage distribution of AJT Responses across participants' acculturation bands will be explored for each of the two cultures' *Lifestyle* (acculturation) measures.

4.5.1.1 English Acculturation

English acculturation will first be examined in terms of median averages, incorporating the overall acculturation score, along with each of the three sub-categories - *Attitudes & Identity*, *Lifestyle*, and *Language use*. Because these median scores are derived from 7-point Likert data, the maximum possible median is 7. This data is displayed in Table 4.13 and, in addition to the results across all participants, median values are shown by participant group, as well as by birth country.

Group	English Attitudes & Identity	English Lifestyle	English Language Use	Overall English Acculturation
English-born LGBTQ+	4	7	7	7
English-born non-LGBTQ+	5	7	7	7
English-born overall	4	7	7	7
Polish-born LGBTQ+	4	7	7	6
Polish-born non-LGBTQ+	4	6	6	5
Polish-born overall	4	6	6	5
Total across all participants	**4**	**6**	**7**	6

Table 4.13*. Median values for overall English acculturation and its 3 sub-categories, by participant group*

The median for overall English acculturation score across all participants is 6. Comparing by birth-country, this varies between 7 for English-born participants and 5 for Polish-born participants, meaning Polish-born participants average a high (but not as high as English-born participants) level of English acculturation. English-born LGBTQ+ participants have

the same median English acculturation score as their non-LGBTQ+ counterparts (7), while Polish-born LGBTQ+ participants have a higher median English acculturation than their non-LGBTQ+ counterparts (6 compared to 5). Breaking English acculturation down into the three sub-categories, English *Attitudes & Identity* has the lowest total median across all participants (4), while English *Lifestyle* is higher (6), and English *Language Use* is highest (7). An obvious reason for this latter result is the fact that people who are living and working in England are very likely to use the English language on a regular basis. Country of birth does not impact the median for English *Attitudes & Identity* (the average is 4 across both), but Polish-born participants do have lower, though still relatively high, medians for English *Lifestyle* and English *Language Use* compared to English-born participants (6 compared to 7, in each case). Across English-born participants, LGBTQ+ status affects only the median English *Attitudes & Identity* score, which is lower for LGBTQ+ participants than for non-LGBTQ+ (4 compared to 5). Across Polish-born participants, LGBTQ+ status does not impact median scores across any of the three sub-categories (but, as mentioned previously, Polish-born LGBTQ+ participants do have a higher median overall English acculturation score than their non-LGBTQ+ counterparts).

Table 4.14 shows participants' percentage distribution across the three bands of English *Lifestyle* (ie. the proxy for English acculturation, and referred to henceforth as such).

Group	Low		Medium		High	
	N	%	N	%	N	%
English-born LGBTQ+	-	-	8	20.00%	32	80.00%
English-born non-LGBTQ+	1	2.50%	5	12.50%	34	85.00%
English-born overall	1	1.25%	13	16.25%	66	82.50%
Polish-born LGBTQ+	5	13.89%	8	22.22%	23	63.89%
Polish-born non-LGBTQ+	4	10.26%	21	53.85%	14	35.90%
Polish-born overall	9	12.00%	29	38.67%	37	49.33%
Total	10	6.45%	42	27.10%	103	66.45%

Table 4.14*. Distribution of participants by English Lifestyle (acculturation), according to participant group*

We predictably see a skew towards higher English acculturation bands across all participant groups. Overall, 66.45% participants have *High* acculturation, and this is the most popular band across all participant groups except for Polish-born non-LGBTQ+, the majority of whom (53.85%) have *Medium* acculturation. Aligning with the median results discussed prior, English-born participants' acculturation is more heavily skewed towards *High* than Polish-born participants', with 82.50% English born participants in this band, compared to 49.33% Polish-born participants. Only 1 English-born participant (non-LGBTQ+) has *Low* English acculturation, whereas 9 Polish-born participants (12%) are in this band.

LGBTQ+ status seems to interact with acculturation level across both birth countries, as a lower percentage of English-born LGBTQ+ participants have *High* English acculturation than their non-LGBTQ+ counterparts (80% compared to 85%), and, conversely, a higher percentage of Polish-born LGBTQ+ participants have *High* English acculturation compared to their non-LGBTQ+ counterparts (63.89% compared to 35.9%).

Figure 4.12 shows the percentage distribution of AJT Responses by English Lifestyle (as a measure of acculturation).

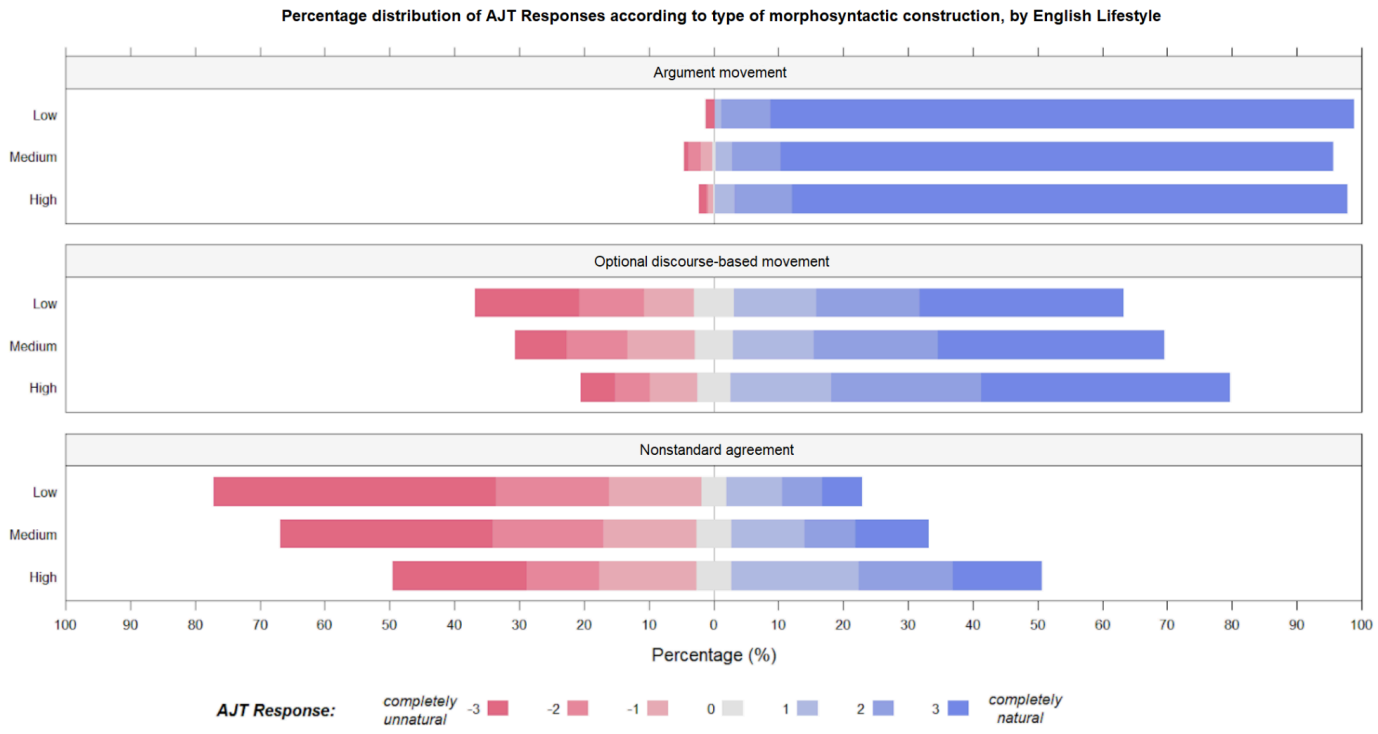


Figure 4.12*. Distribution (%) of AJT Responses according to type of morphosyntactic construction, by English Lifestyle (acculturation) band*

Participants' English acculturation positively impacts AJT acceptance of optional discourse-based movement and nonstandard past-tense BE agreement - the stacked bars for the higher acculturation levels are consistently further right than those for the lower ones. It also seems to be the case that nonstandard past-tense BE agreement undergoes a stronger cumulative increase in acceptability with each increment of acculturation level than optional discourse-based movement does. Optional discourse-based movement has 60% positive and 33.64% negative responses among *Low* acculturated participants, compared to 76.86% positive and 17.76% negative responses among *High* acculturated participants. Following a similar pattern, nonstandard past-tense BE agreement has 20.87% positive and 75.22% negative responses among *Low* acculturated participants and 47.78% positive and 46.72% negative responses among *High* acculturated participants. However, inferential statistical methods are required to assess whether the effects found here remain significant when random individual and test item effects are negated.

4.5.1.2 Polish Acculturation

As with English acculturation, Polish acculturation will first be examined in terms of median averages, incorporating the overall acculturation score, along with each of the three sub-categories - *Attitudes & Identity*, *Lifestyle*, and *Language use*. Because these median scores are derived from 7-point Likert data, the maximum possible median is 7. This data is displayed in Table 4.15 and, in addition to the results across all Polish-born participants, median values are shown by LGBTQ+ status.

Group	Polish Attitudes & Identity	Polish Lifestyle	Polish Language Use	Overall Polish Acculturation
Polish-born LGBTQ+	3	4	2	3
Polish-born non-LGBTQ+	4	5	3	4
Total across Polish-born participants	4	4.5	3	4

Table 4.15*. Median values for overall Polish acculturation and its 3 sub-categories, by LGBTQ+ status*

The median for overall Polish acculturation score across all Polish-born participants is 4. Polish-born LGBTQ+ participants have a lower median Polish acculturation score than their non-LGBTQ+ counterparts (3 compared to 4). Breaking Polish acculturation down into the three sub-categories, Polish *Language Use* has the lowest total median across all Polish-born participants (3). Unsurprisingly, Polish-born participants are comparatively less likely to need to use the Polish language while living and working in England, even if they have a strong alignment with Polish culture. Polish *Attitudes & Identity* has a higher median (4), and Polish *Lifestyle* has the highest (4.5). These medians are lower than their English acculturation equivalents among Polish-born participants, apart from English and Polish *Attitudes & Identity* which both have a median of 4. LGBTQ+ status affects the medians for all three Polish acculturation sub-categories: Polish *Attitudes & Identity* is lower for LGBTQ+ participants than for non-LGBTQ+ (3 compared to 4); Polish *Lifestyle* is also lower for LGBTQ+ participants than non-LGBTQ+ (4 compared to 5), as is Polish *Language Use* (2 compared to 3).

Table 4.16 shows participants' percentage distribution across the three bands of Polish *Lifestyle* (a proxy for Polish acculturation, and will be referred to henceforth as such),

Group	Low		Medium		High	
	N	%	N	%	N	%
Polish-born LGBTQ+	20	55.56%	14	38.89%	2	5.56%
Polish-born non-LGBTQ+	13	33.33%	19	48.72%	7	17.95%
Total	33	44.00%	33	44.00%	9	12.00%

Table 4.16*. Distribution of Polish-born participants by Polish Lifestyle (acculturation), according to LGBTQ+ status*

Across the three bands of Polish *Lifestyle*, the opposite pattern is present to that of English acculturation, in that participants skew away from higher Polish acculturation. Participants are more evenly distributed across Polish acculturation bands compared to the spread of participants across English acculturation bands (as shown in Table 4.14). Table 4.16 shows that 12% Polish-born participants have *High* Polish acculturation, compared to 44% in both the *Medium* and *Low* bands.

Aligning with the median results discussed prior, LGBTQ+ status seems to interact with Polish acculturation level. A lower percentage of Polish-born LGBTQ+ participants have *High* Polish acculturation than their non-LGBTQ+ counterparts (5.56% compared to 17.95%) and a far higher percentage of Polish-born LGBTQ+ participants have *Low* Polish acculturation than their non-LGBTQ+ counterparts (55.56% compared to 33.33%).

Figure 4.13 shows the percentage distribution of AJT responses by Polish Lifestyle (as a measure of acculturation).

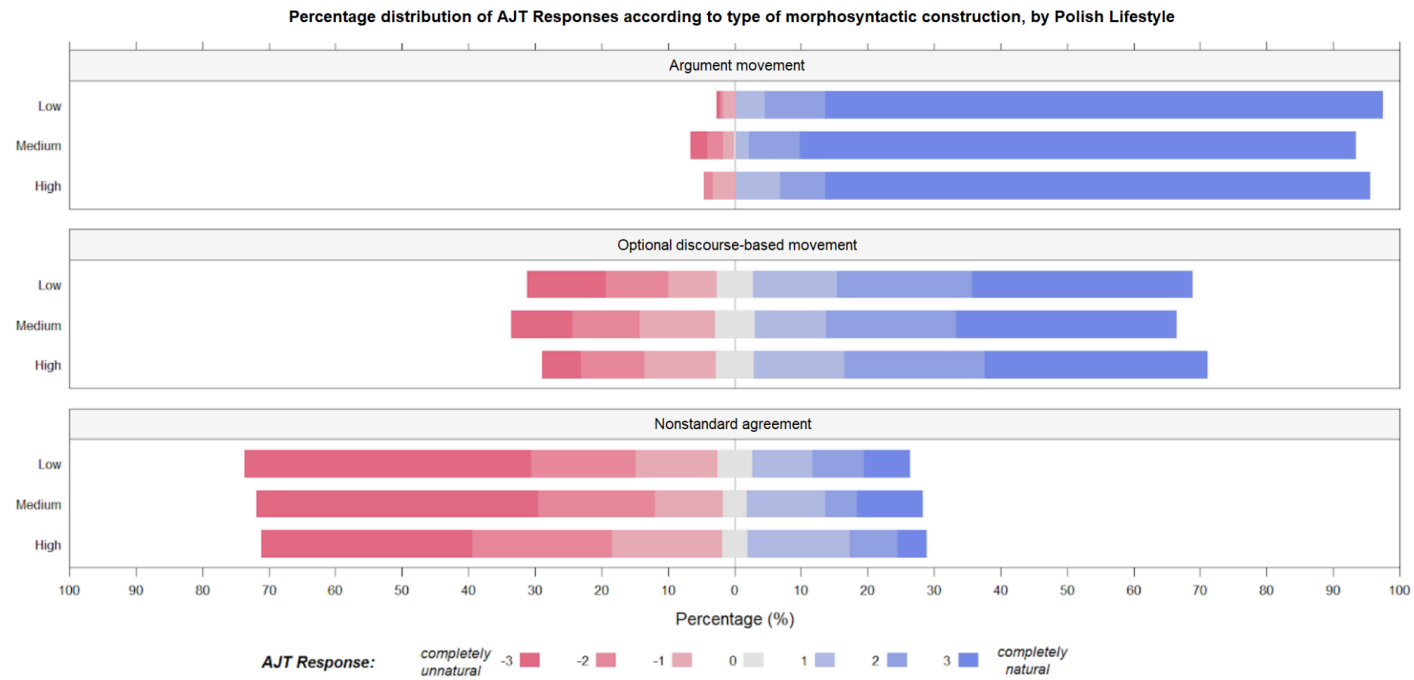


Figure 4.13. *Distribution (%) of AJT Responses according to type of morphosyntactic construction, by Polish Lifestyle (acculturation) band*

Participants' Polish acculturation level does not clearly impact AJT acceptance for any of the three types of morphosyntactic construction. Interestingly, *High* Polish-acculturated participants have higher acceptance than *Medium* or *Low* Polish-acculturated participants for both optional discourse-based movement and nonstandard past-tense BE agreement (ie. the stacked bars for the higher acculturation levels are further right). They also have the lowest proportion of minimal AJT Responses (of -3) for nonstandard past-tense BE agreement. This effect, however, appears minimal, and also the *High* band has a lower number of participants than the other two. During Likelihood Ratio testing (see Section 3.9.2.3), it was determined that Polish Lifestyle does not have a significant main effect on AJT Response ($p = .865$), nor does it significantly interact with AJT Condition ($p = .863$). Due to this, Polish acculturation is not included in further statistical modelling.

4.5.2 LGBTQ+ Community Involvement & LGBTQ+ Identity Openness

The final element of this study involved a survey which measured LGBTQ+ Community Involvement (see Section 3.5.3 for details about the indicators used to calculate this measure). This was completed by all LGBTQ+ participants. LGBTQ+ Community Involvement scores were calculated by averaging scores across all the responses from the respective survey. To provide an overall snapshot of LGBTQ+ Community Involvement results, this section will first explore median

scores, comparing by country of birth. This will be followed by an assessment of the percentage distributions of participants across the three bands of LGBTQ+ Community Involvement - *Low*, *Medium*, and *High*. Finally, the percentage distribution of AJT Responses across participants' LGBTQ+ Community Involvement bands will be explored. To explore the interplay between LGBTQ+ status and acculturation, this section will conclude with a discussion contextualising LGBTQ+ participants' results. This will take into account additional information collected about participants' levels of openness about their LGBTQ+ identities around different types of people, as well as data collected specifically from Polish-born LGBTQ+ participants' about their LGBTQ+ identity and use of the English versus the Polish language. The main motivations driving Polish-born participants' migration to England will also be discussed here.

4.5.2.1 LGBTQ+ Community Involvement

The LGBTQ+ Community Involvement measure has been derived from a set of 7 questions related to LGBTQ+ community life and activities. Table 4.17 shows the median LGBTQ+ Community Involvement score across all LGBTQ+ participants, as well as across each birth country. Because this measure is derived using a 7-point Likert-type method, the maximum possible median is 7. The median LGBTQ+ Community Involvement score across all LGBTQ+ participants is high, at 6 out of a maximum of 7.

`Group	LGBTQ+ Community Involvement
English-born LGBTQ+	5.5
Polish-born LGBTQ+	6
Total across LGBTQ+ participants	6

***Table 4.17**. Median values for LGBTQ+ Community Involvement, by birth country*

By country of birth, English-born participants have a slightly lower median LGBTQ+ Community Involvement than Polish-born (5.5 compared to 6).

From Table 4.18 we can assess participants' percentage distribution across the three bands of LGBTQ+ Community Involvement.

Group	Low		Medium		High	
	N	%	N	%	N	%
English-born LGBTQ+	8	20.00%	16	40.00%	16	40.00%
Polish-born LGBTQ+	7	19.44%	12	33.33%	17	47.22%
Total	15	19.74%	28	36.84%	33	43.42%

***Table 4.18**. Distribution of LGBTQ+ participants by LGBTQ+ Community Involvement band, according to birth country*

Aligning with the median values, participants are skewed towards higher LGBTQ+ Community Involvement, with 43.42% having *High* involvement, 36.84% having *Medium* involvement, and a minority - 19.74% with *Low* involvement. The *Low* involvement band is evenly split by birth country, but a higher percentage of Polish-born participants have *High* involvement (47.22%) compared to the percentage of English-born participants in this band (40%).

Figure 4.14 shows the percentage distribution of AJT Responses by LGBTQ+ Community Involvement.

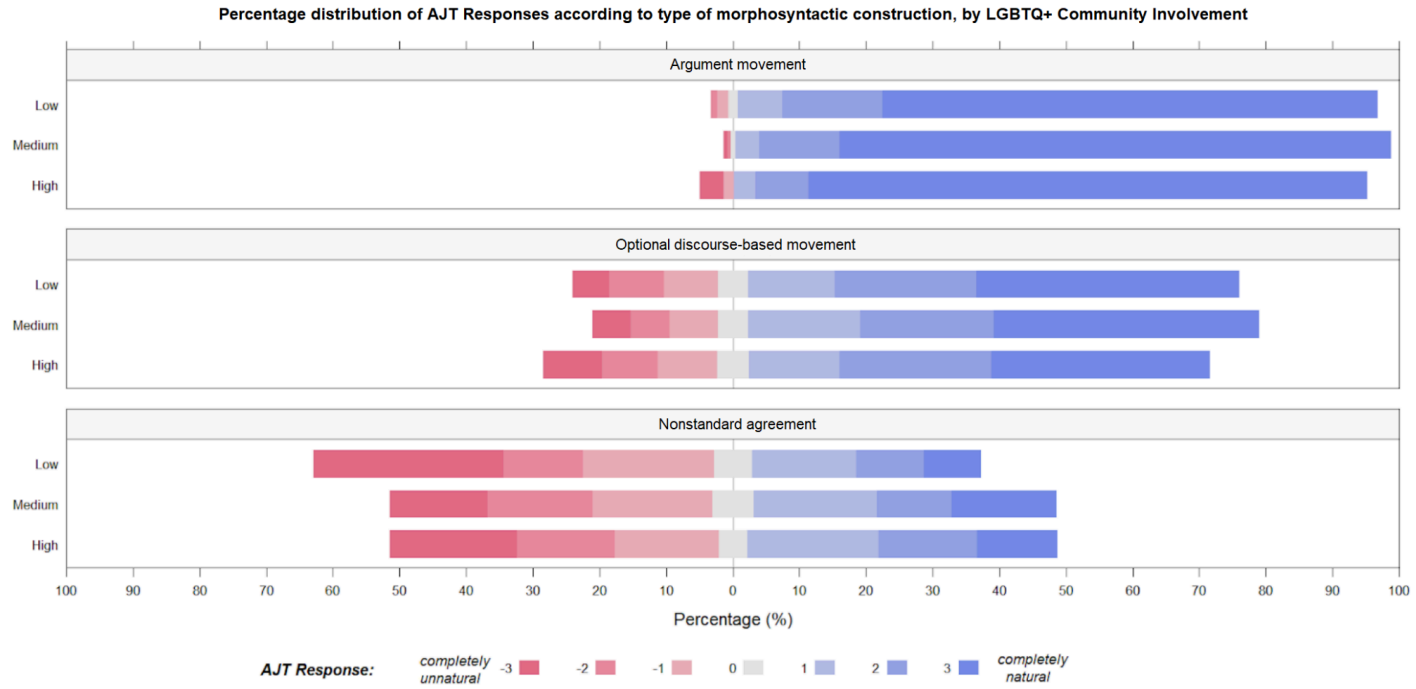


Figure 4.14. *Distribution (%) of AJT Responses according to type of morphosyntactic construction, by LGBTQ+ Community Involvement band*

From these results, it is possible that LGBTQ+ Community Involvement level is impacting AJT acceptance of optional discourse-based movement and nonstandard past-tense BE agreement, albeit minimally. Participants in the *High* band have a slightly lower acceptance of optional discourse-based movement than those in the *Medium* or *Low* bands (ie. the stacked bar is further left). This effect does not appear consistent, however, with participants in the *Low* band having lower acceptance than those in the *Medium* band. The acceptability of nonstandard past-tense BE agreement appears consistent across the *Medium* and *High* LGBTQ+ Community Involvement bands (ie. the stacked bars are horizontally parallel), with both bands providing around 45% positive and 50% negative responses to nonstandard past-tense BE agreement. *Low* involvement participants, however, are comparatively more negative in their evaluations to nonstandard past-tense BE agreement, with 34.2% positive and 60% negative evaluations. Because the *Low* band contains a smaller proportion of participants compared with the other two bands, this effect warrants testing through further statistical modelling to mitigate the random effect of individual variation. Further analysis of this is especially warranted given that, as discussed in Section 3.9.2.3, Likelihood Ratio testing did indeed reveal a significant interaction effect on AJT Response between LGBTQ+ Community Involvement and AJT Condition ($p = <.001$).

4.5.2.2 Contextualising LGBTQ+ motivations

As seen in Sections 4.1.2 and 4.4.2.1, the descriptive findings presented in this chapter suggest an interaction effect between LGBTQ+ status and birth country. Namely, LGBTQ+ status influences AJT acceptance of nonstandard past-tense BE agreement (but not optional discourse-based movement) conditions across Polish-born (but not English-born) participants; Polish-born LGBTQ+ participants are overall more positive and less negative towards these conditions compared to Polish-born non-LGBTQ+ participants.

LGBTQ+ status seems to interact with both English and Polish acculturation level, with English acculturation patterning differently with LGBTQ+ status across both birth countries. A lower percentage of English-born LGBTQ+ participants have *High* English acculturation than their non-LGBTQ+ counterparts, and English-born LGBTQ+ participants, on average, also have lower English *Attitudes & Identity* scores than English-born non-LGBTQ+ participants. Among Polish-born participants, on the other hand, the opposite relationship seems to be the case, as a higher percentage of Polish-born LGBTQ+ participants have *High* English acculturation compared to their non-LGBTQ+ counterparts. There is also a disparity in LGBTQ+ Community Involvement scores by birth country, where a higher percentage of Polish-born participants have *High* involvement compared to the percentage of English-born LGBTQ+ participants in this band. LGBTQ+ status also affects Polish acculturation, as Polish-born LGBTQ+ participants, on average have lower Polish acculturation scores than their non-LGBTQ+ counterparts, as well as lower scores for each of the three Polish acculturation sub-categories. A lower percentage of Polish-born LGBTQ+ participants have *High* Polish acculturation than their non-LGBTQ+ counterparts, and a far higher percentage of Polish-born LGBTQ+ participants have *Low* Polish acculturation than their non-LGBTQ+ counterparts.

To further explore the interplay between LGBTQ+ status and acculturation, this section will conclude with a discussion providing potential context to these results for LGBTQ+ participants. This section will take into account additional information collected about participants' levels of openness about their LGBTQ+ identities around different types of people, as well as data collected specifically from Polish-born LGBTQ+ participants' about their LGBTQ+ identity with regards to their use of the English versus the Polish language. The main motivations driving Polish-born participants' migration to England will also be discussed here, as well as their feelings towards the key sociocultural factors of religion and Brexit.

4.5.2.3 Motivations for migration

Table 4.19 shows participants' key motivations for their migration to England.

Motivation for migration	Polish-born LGBTQ+		Polish-born non-LGBTQ+		All Polish-born	
	N	%	N	%	N	%
Work/employment	20	55.56%	19	48.72%	39	52.00%
Study/education	15	41.67%	16	41.03%	31	41.33%
Cultural differences between Poland and England	16	44.44%	3	7.69%	19	25.33%
Family/friends/partner living in England	6	16.67%	6	15.38%	12	16.00%
Moved with family as a child	4	11.11%	3	7.69%	7	9.33%
Other	4	11.11%	3	7.69%	7	9.33%

Table 4.19*. Distribution of Polish-born participants' overall motivations for moving to England, by LGBTQ+ status*

At least one response was required, but participants could select up to 3 responses if applicable. Of the 75 Polish-born participants, 49 only indicated one motivation, 12 selected 2 different motivations, and 14 listed 3 motivations. This information is first discussed with regards to overall motivations for migration across all Polish-born participants, and is then discussed with regards to LGBTQ+ status.

By far the most common factor motivating participants' move to England is *Work/employment*, which was important to 52% - ie. 39 of the 75 - Polish-born participants, followed by *Study/education*, important to 41.3% (31/75) participants. Cultural differences between Poland and England motivated 25.3% (19/75) Polish-born participants. Other, less often cited factors are to join family or friends in England, motivating 16% (12/75) participants and having the decision made during childhood by family, affecting 9.3% (7/75) participants. Finally, a handful of participants (9.3%, or 7/75) also provided an *Other* motivation beside the provided selection, which largely came down to an interest in English language and/or culture, and desire to travel beyond Poland.

Comparing motivations according to LGBTQ+ status, almost all of the motivations listed in Table 4.19 are distributed roughly evenly among LGBTQ+ versus non-LGBTQ+ Poles. Noticeably, however, the distributions for *Cultural differences between Poland and England* differ drastically, with LGBTQ+ participants making up 16 of the 19 Polish-born participants who indicated this motivation as a factor for migration. In other words, of the 36 Polish-born LGBTQ+ participants, 16 (ie. 44.44%) considered cultural aspects important enough to, at least in part, motivate their decision to move country, whereas only 3 of the 39 (7.69%) non-LGBTQ+ participants felt this way.

4.5.2.4 Key sociocultural factors

I will now briefly explore two key sociocultural factors which may help contextualise the lived experiences of the Polish-born and, specifically, Polish-born LGBTQ+ participants - religious affiliation and Brexit approval (see Section 1.3 for a discussion of the relevance of these). Both measures are derived from 7-point Likert scales, therefore, have a maximum median value of 7. Firstly, the degree of religious affiliation amongst participants in general is very low, with a median response of 1 (equivalent to *Not at all religious*), and with 60% of all participants falling within this band. Polish-born participants are slightly less non-religious than English-born, with 56% responding with *1*, compared to 63.75% of

English-born participants. Comparing Polish-born participants by LGBTQ+ status, non-LGBTQ+ participants have by far the *lowest* non-religiousness, with only 43.59% in band *1*. Polish-born LGBTQ+ participants are the *most* non-religious of the four groups (including English-born LGBTQ+ participants), with 69.44% responding *1* and with no participants giving a response higher than *4* (equivalent to *Neutral/no opinion*).

As for Brexit approval, participants are generally very disapproving, with a median response of 1 and with 70.97% falling in band *1*. Interestingly, English participants are more strongly *entirely opposed* to Brexit (ie. gave a rating of *1*) than Polish-born, with 73.75% and 68% in this band, respectively. English-born LGBTQ+ participants are more strongly *entirely opposed* to Brexit than their non-LGBTQ+ counterparts, with 80% responding with *1*, compared to 67.50%. Polish-born non-LGBTQ+ participants are the least *entirely opposed*, with only 58.97% in band *1*. A far higher percentage of Polish-born LGBTQ+ participants are *entirely opposed* to Brexit compared to their non-LGBTQ+ counterparts (77.78% compared to 58.97%).

4.5.2.5 LGBTQ+ Identity Openness

As well as collecting information about LGBTQ+ participants' LGBTQ+ Community Involvement, their 'Openness' was also calculated. This was based on participants' reported likelihood that they would be open about their LGBTQ+ identity with English versus non-English strangers and friends, and was collected in order to assess whether the perceived 'welcomeness' of LGBTQ+ identities is perceived as different in English culture (whether better or worse) compared to other cultures. In addition, Polish-born participants were also asked the same questions, specifically in relation to Polish strangers and friends, in order to determine whether the same is true for Polish culture in the eyes of the Polish-born participants. The median average responses are displayed in Table 4.20, by country of birth and across all LGBTQ+ participants. The columns denote responses related to English people, Non-English people, and, for Polish-born participants only, Polish people. Each of these three categories is subdivided into columns differentiating friends from strangers, and, for each, a total value is also given, which is the average across each of these two subcategories. As these data originate from 7-point Likert scale responses, the maximum possible median is 7.

Group	English Friends	English Strangers	Total Openness English	Non-English Friends	Non-English Strangers	Total Openness Non-English	Polish Friends	Polish Strangers	Total Openness Polish
English-born LGBTQ+	7	5	5	6	3	5	-	-	-
Polish-born LGBTQ+	7	5	6	6	4	5	5	2	3
Total across LGBTQ+ participants	7	5	6	6	3	5	5	2	3

Table 4.20*. Median openness about LGBTQ+ identity with English people, non-English people, and Polish people, sub-divided by respondents' country of birth*

Comparing the total openness scores across all LGBTQ+ participants (ie. the bottom row of Table 4.20), on average, participants feel more open to disclosing their LGBTQ+ identities with English people, with a total median openness of 6, compared to non-English people, who score 5. LGBTQ+ Polish-born participants feel even less open with other Poles, with a median total openness score of 3. This is lower than Polish-born respondents' openness towards non-English people (5) and English people (6). On an individual level, only 8 Polish-born LGBTQ+ participants had an equal

openness rating for English and Polish people, and the remaining 28 rated their openness higher with English than with Polish people.

Predictably, median LGBTQ+ identity openness ratings are consistently higher for friends than strangers of the equivalent type; 7 compared to 5 for English people; 6 compared to 3 for non-English people, and 5 compared to 2 for Polish people (the latter data informed by Polish-born participants only). Polish-born participants' ratings towards fellow Polish friends versus strangers diverge more from each other than their ratings across the other two categories, with a difference of 3, compared to 2 between non-English friends versus strangers, and between English friends versus strangers. Polish-born participants are more likely to be open about their LGBTQ+ identity with English strangers (5) than Polish strangers (2). Similarly, Polish-born participants are more likely to be open with English friends (7) than Polish friends (5), though the margin of difference is smaller. Polish-born participants have a slightly higher total median openness towards English people (6) than English-born participants do (5). Breaking this down according to non-English friends versus strangers, Polish-born participants are slightly more likely to be open with non-English strangers (4) than English-born participants are (3).

4.5.2.6 LGBTQ+ Interaction & Language

Finally, Polish-born LGBTQ+ participants also provided data regarding their perceived likelihood of using the English versus the Polish language when interacting with other LGBTQ+ people. This was measured using two survey items, both with equivalent 7-point Likert-type response scales, ranging from 'entirely unlikely' to 'entirely likely'. Participants had the option to leave these questions unanswered if, for instance, they do not interact with other LGBTQ+ people at all, however, all Polish-born LGBTQ+ participants answered both questions.

On average, Polish-born LGBTQ+ participants' likelihood of using the English language to interact with other LGBTQ+ people is heavily skewed towards the upper end of the scale (towards *entirely likely*), with a median score of 7, and with 28 of the 36 (77.78%) Polish-born LGBTQ+ participants giving the maximum rating for this measure. The responses regarding reported likelihood of using the Polish language with other LGBTQ+ people, on the other hand, are much more evenly distributed, with a median response of 5. This measure does still exhibit a slight positive skew, with 7 being the mode response, though only by 27.78% (10 of 36) participants.

Additionally, comparing between the responses to these two questions per participant, 65.8% (25 of 36) Polish-born LGBTQ+ participants indicate a *higher* likelihood of using English than the comparative likelihood of using Polish when interacting with other LGBTQ+ people, while 31.6% (12 of 36) have the same likelihood rating for both languages (all ratings either 6 or 7). Only one person feels relatively *more* likely to use Polish than English with other LGBTQ+ people.

It seems the Polish-born LGBTQ+ sample is made up of people with a variety of degrees of involvement with other LGBTQ+ Poles. This could be a contributing factor to the comparatively lower 'Openness' ratings with other Poles in the prior section. Perhaps participants know comparatively fewer other LGBTQ+ Poles because they live in England and, as shown in Section 4.4.1, Polish acculturation measures are, on average, lower than that of English acculturation in these participants. This may mean that many of these participants do not have strong ties to Polish LGBTQ+ networks as well. However, this is speculative, and it is also entirely possible that participants do know other LGBTQ+ Poles but use English rather than Polish to communicate with them.

4.6 Overview of descriptive results

This chapter has outlined the descriptive results of this study. Firstly, the patterning of the dependent variable (AJT Response) was explored according to each of the AJT Conditions (ie. the morphosyntactic features of interest in this study), including discussion of AJT Response results through the lens of the three types of morphosyntactic construction investigated in this thesis - ie. argument movement; optional discourse-based movement; and nonstandard agreement - as well as a breakdown across the four participant groups (English-born LGBTQ+; English-born non-LGBTQ+; Polish-born LGBTQ+; Polish-born non-LGBTQ+). Following this, the effects of each factor of interest on the AJT Response were investigated, using stacked bar charts to compare the percentage distributions of AJT Responses across the categories of each variable, and by the type of morphosyntactic construction. Results have largely been summarised in terms of the overall percentages of positive versus negative responses (ie. responses between -3 and -1 versus those between 1 and 3), highlighting any categories with noticeably higher or lower proportions of responses at the extremities of the AJT scale. As part of the descriptive analysis, the percentage distributions of participants across categories of each factor were also evaluated to provide additional context about the participant sample and to assess the suitability of these distributions prior to statistical modelling. The remainder of this chapter is a review of the findings, which are also summarised in Table 4.21 at the end of this chapter.

4.6.1 Morphosyntactic conditions

Descriptive analysis of the effect of AJT Condition on results across all participants was presented first. Predictably, the positive control argument movement construction (raising-to-subject with the verb *seem*) is almost unanimously accepted by all participants. The optional discourse-based movement construction of left dislocation is the non-control condition with the most positive range of AJT Responses, followed by right dislocation, and then topicalisation. This is followed by the nonstandard agreement constructions of nonstandard *weren't*, nonstandard *were*, and nonstandard *was*, respectively. These findings reveal a clear pattern in relative acceptability by type of morphosyntactic construction, with optional discourse-based movement more accepted than nonstandard agreement.

4.6.2 Key sample characteristics

Country of birth affects AJT Response, with Polish-born participants, on average, tending towards more negative evaluations. This is also the case when comparing across the three types of morphosyntactic construction, each of which is relatively less accepted by Polish-born than English-born participants. The key finding here is that the relative disparity in acceptance between participants from the two birth countries is higher for nonstandard past-tense BE agreement constructions than for optional discourse-based movement constructions, which, in turn, is higher than for the positive control argument movement construction (raising-to-subject with *seem*) .

LGBTQ+ status does not seem to affect acceptability of nonstandard past-tense BE agreement constructions or optional discourse-based movement constructions in participants from either birth country. However, there is an interaction between being Polish-born and LGBTQ+ on the acceptance of nonstandard past-tense BE agreement constructions: Polish-born LGBTQ+ participants are overall more positive and less negative towards these compared to Polish-born non-LGBTQ+ participants, and a similar difference in pattern according to LGBTQ+ status is **not** found in English-born participants.

4.6.3 Macro-social factors

The macro-social (ie. demographic) factors of interest explored here were age, gender, region, and the socioeconomic factors of socioeconomic status and education status. Age has an effect on acceptance of nonstandard past-tense BE agreement, with older participants generally less accepting than younger ones. Participants whose gender is beyond the binary of male and female are comparatively more accepting of nonstandard past-tense BE agreement compared to male and female respondents. Socioeconomic status does not clearly affect participants' acceptance across any of the three types of morphosyntactic constructions. Participants' involvement in higher education does not impact acceptance of optional discourse-based movement, but seems to have an effect on acceptance of nonstandard past-tense BE agreement, with higher-educated participants more accepting than those who are not higher-educated. As for region, participants in the South have the lowest acceptance of all regions for both optional discourse-based movement and nonstandard past-tense BE agreement. The Midlands is the most positive region towards nonstandard past-tense BE agreement. By country of birth, this pattern is consistent among Polish-born participants, who, in the Midlands, also evaluate nonstandard past-tense BE agreement more positively than in other regions. Among Polish-born participants, the North is the region least accepting of

nonstandard past-tense BE agreement. Assessing participants' intra-national mobility, there is a clear divide between Polish-born and English-born participants, with the majority of Polish-born participants *not* having lived outside of their current region of residence, while the majority of English-born participants *have*.

4.6.4 L2 factors

The second language (L2) factors explored are Polish-born age of arrival in England and their English language proficiency. Age of arrival to England does not seem to have a clear effect on acceptance, however, descriptive results do suggest that participants who migrated during childhood and adolescence (between 7-18) have higher acceptance across all three types of morphosyntactic constructions than those who migrated later. English Language Exposure (in years) positively correlates with acceptance of optional discourse-based movement, but no clear effect can be seen on acceptance of nonstandard past-tense BE agreement. The formal measure of English proficiency, however, shows a clear positive correlation between higher English language proficiency and higher acceptance of all three types of morphosyntactic constructions, with a stronger relative positive effect on acceptance of nonstandard past-tense BE agreement than on optional discourse-based movement, and a stronger positive effect on acceptance of optional discourse-based movement than on the argument movement construction (raising-to-subject with *seem*).

4.6.5 Meso-social factors

The meso-social factors (ie. those related to community embeddedness) explored here have been participants' levels of English and Polish acculturation, and LGBTQ+ participants' levels of LGBTQ+ Community Involvement. English acculturation level positively correlates with acceptance of optional discourse-based movement and nonstandard past-tense BE agreement. Polish acculturation level, on the other hand, does not seem to impact acceptance for any of the three types of morphosyntactic construction in a consistent way. The main finding for LGBTQ+ Community Involvement level is that *Low* involvement participants seem to be comparatively more negative in their evaluations to nonstandard past-tense BE agreement than *Medium* or *High* involvement participants.

Finally, this chapter has also explored Polish-born LGBTQ+ participants' acculturation results. A lower percentage of English-born LGBTQ+ participants have *High* English acculturation than their non-LGBTQ+ counterparts, and English-born LGBTQ+ participants, on average, also have lower English *Attitudes & Identity* scores than English-born non-LGBTQ+ participants. Among Polish-born participants, on the other hand, the opposite relationship seems to be the case, as a higher percentage of Polish-born LGBTQ+ participants have *High* English acculturation compared to their non-LGBTQ+ counterparts. There is also a disparity in LGBTQ+ Community Involvement scores by birth country, where a higher percentage of Polish-born participants have *High* involvement compared to the percentage of English-born LGBTQ+ participants in this band. Polish-born LGBTQ+ participants, on average have lower Polish acculturation scores than their non-LGBTQ+ counterparts, as well as lower scores for each of the three Polish acculturation sub-categories. A lower percentage of Polish-born LGBTQ+ participants have *High* Polish acculturation than their non-LGBTQ+ counterparts, and a far higher percentage of Polish-born LGBTQ+ participants have *Low* Polish acculturation than their non-LGBTQ+ counterparts.

To contextualise the results found for Polish-born LGBTQ+ participants, I have explored participants' levels of openness about their LGBTQ+ identities around different types of people, finding that Polish-born participants feel more likely to be open about their LGBTQ+ identity with English people than with other Poles, and more unlikely to be open with Polish strangers than friends compared to the relative difference in likelihood between English strangers and friends. Further socio-cultural contextual findings reveal that most Polish-born participants who migrated due to cultural differences between Poland and England were LGBTQ+. Furthermore, Polish-born LGBTQ+ participants were the most strongly non-religious of the four groups, and were far more strongly opposed to Brexit than their non-LGBTQ+ counterparts.

4.6.6 Summary

Overall, Polish-born participants are less accepting of all three types of morphosyntactic construction than English-born participants (and this effect is stronger for nonstandard past tense BE than for optional discourse-based movement). Also, birth country seems to have an interaction effect with LGBTQ+ status, with Polish-born LGBTQ+ participants more accepting of nonstandard past-tense BE agreement than Polish-born non-LGBTQ+. These findings confirm that the key sample characteristics employed within this study do in fact, condition the acceptance of morphosyntactic variation. L2 factors seem to also play a part in Polish-born participants' acceptance, with, for example, migrants' English proficiency seeming to have a positive effect on acceptance, this effect being stronger for nonstandard past-tense BE agreement constructions than for optional discourse-based movement constructions.

As might be expected, acceptance of nonstandard past-tense BE agreement is more likely to pattern with macro-social factors - eg. age, education (higher-educated status), region of residence, gender - than optional discourse-based movement is. On the other hand, an interesting pattern found regarding the acceptance of optional-discourse-based movement is that it seems to be more likely to be affected by factors relating to the degree of exposure to community-related linguistic norms, such as participants' level of LGBTQ+ Community Involvement, or Polish-born participants' degree of exposure to the English language. Additionally, according to the descriptive results, English acculturation (through the proxy measure of English Lifestyle) also seems to have a positive effect on acceptance across all morphosyntactic constructions.

To reiterate the findings from this chapter in a simplified overview, the key factors explored in this chapter, and their effects on acceptance across the three types of morphosyntactic constructions are summarised in Table 4.21. Given these findings, I now move on to analysing these effects further, using statistical modelling.

Type of factor	*Factor*	Summary of descriptive results
Key sample characteristics	Birth Country	Polish-born less accepting of all three types of morphosyntactic construction than English-born (and disparity wider for nonstandard past-tense BE agreement than for optional discourse-based movement)
	LGBTQ+ Status	Polish-born LGBTQ+ more accepting of nonstandard past-tense BE agreement than Polish-born non-LGBTQ+
Macro-social (demographic)	Age	Younger age bands more accepting of nonstandard past-tense BE agreement than older bands
	Gender	*Not binary* more accepting of nonstandard past-tense BE agreement than male or female
	Region of Residence	Southerners least accepting of optional discourse-based movement and nonstandard past-tense BE agreement. Midlanders most accepting of nonstandard past-tense BE agreement (also the case among only Polish-born). Polish-born Northerners least accepting of nonstandard past-tense BE agreement.
	Socioeconomic Status	No clear effect
	Higher-educated Status	Higher-educated more accepting of nonstandard past-tense BE agreement than non higher-educated
L2 (second language)	Age of Arrival to England	Those who arrived aged 7-18 **potentially** more accepting of all morphosyntactic constructions than later arrivers
	English Language Exposure (years)	Those with longer exposure more accepting of optional discourse-based movement
	English Proficiency (CEFR)	Higher proficiency English speakers are more accepting of all morphosyntactic constructions
Meso-social (community embeddedness)	English Lifestyle (Acculturation)	Those with higher English acculturation are more accepting of all morphosyntactic constructions
	Polish Lifestyle (Acculturation)	No clear effect
	LGBTQ+ Community Involvement	Those with *Low* involvement less accepting of optional discourse-based movement than *Medium* and *High*

Table 4.21*. Summary of descriptive results concerning the effects on acceptance of the three types of morphosyntactic construction*

5. Inferential Results

5.1 Introduction

The following chapter presents an exploration of the statistical results from four proportional odds (ordinal logistic regression) models. I take a nested approach in order to segment the sample according to the key characteristics under investigation (birth country and LGBTQ+ status):

- **Model 1** (Section 5.2) incorporates all participants (both English-born and Polish-born, and both LGBTQ+ and non-LGBTQ+)
- **Model 2a** (Section 5.3) incorporates only Polish-born participants (both LGBTQ+ and non-LGBTQ+)
- **Model 2b** (Section 5.4) incorporates only Polish-born participants who completed the Versant English Speaking test of linguistic proficiency
- **Model 3** (Section 5.5) incorporates only LGBTQ+ participants (both English-born and Polish-born)

See Section 3.9.1 for an explanation of the process of model design and setup, and the decisions that have been made.

I will discuss each of the four models in turn. For each model (with the exception of Model 2b, which only concerns the CEFR English proficiency predictor), I will first discuss the proportional odds ratio results for the focal predictor of AJT Condition - ie. the morphosyntactic features of interest within the present study. These results show how the different features pattern in terms of their odds of being rated higher on the AJT Response scale. The argument movement construction (raising-to-subject with the verb *seem*), which functions as the positive control condition in the AJT, is used by the models as the reference category for the AJT Condition predictor. Therefore, this AJT Condition itself does not appear in the results tables and, instead, results for the other AJT Conditions will be discussed **in reference to it**. For Models 1, 2a, and 3, these results are also visualised using probability curves that demonstrate, for each of the significant AJT Conditions, how probable the selection of each level of AJT Response is (see Section 3.9.3 for justification of this method of visualisation). Probability is measured along the y-axes of these plots, and the model coefficient estimates are plotted along the x-axes. AJT Conditions that pattern similarly in terms of how participants use the response scale appear closer together on the x-axes of these plots, allowing for easy visual comparison.

Then, for each model, other main effects on the dependent variable of AJT Response (if applicable) are discussed, as well as any interaction effects found between significant model predictors and the focal predictor of AJT Condition. These discussions are structured similarly to the descriptive results (Chapter 4), in that, for each model, I consider factors sequentially according to the five broad types: (ii) the key sample characteristics (ie. birth country in Models 1 and 3, and LGBTQ+ status in Models 1 and 2a); (ii) the macro-social (ie. demographic) factors of interest (where significant), namely age, gender, region, and the socioeconomic factors of socioeconomic status and education status; (iii) the L2 (second language) factors influencing Polish-born participants, namely their age of arrival to England and their English language proficiency (Models 2a and 2b only), and; (iv) the meso-social factors considered in this study (ie. those related to community embeddedness) - English lifestyle (ie. acculturation) levels in Models 2a and 3, as well as levels of LGBTQ+ Community Involvement in Model 3.

I then summarise the findings for the four statistical models, comparing equivalent results between models. This is in order to explore how different factors influence participants' acceptability judgements, and enable comparison of the differences in effects found across Polish-born participants and LGBTQ+ participants with those found across all participants. I will first discuss overarching patterns in the acceptability of the morphosyntactic variants under investigation according to the key sample characteristics of birth country and LGBTQ+ status. In Section 5.6.2, I compare in more detail how different variants within the focal variable of AJT Condition pattern, especially with regard to the type of morphosyntactic construction: (i) argument movement (raising-to-subject with the verb *seem*); (ii) optional discourse-based movement (left dislocation, right dislocation, topicalisation) and (iii) nonstandard agreement (past-tense BE: nonstandard *was*, nonstandard *were*, and nonstandard *weren't*). I also discuss patterns in acceptability with regards to each of the factor groups: key sample characteristics (Section 5.6.1), macro-social factors (Section 5.6.3), L2 factors (Section 5.6.4), and meso-social factors (Section 5.6.5). For ease of reference, patterns of acceptability (by variant function) according to each of these types of factors are summarised in tables in each respective section. An overall summary and discussion is then conducted, synthesising results from across the four models. For ease of reference, the effects of each factor on the patterns on acceptability of each morphosyntactic variant are also reiterated at the end of this chapter in table format (Table 5.12).

5.1.1 Interpreting inferential results

The values under discussion in the model outputs are proportional odds ratios. Odds describe the ratio between the probability of a particular outcome and the probability of another outcome. Odds *ratios* (henceforth, ORs), in the context of the results which will be discussed here, are ratios comparing the *odds* of getting a higher outcome (AJT Response) for a given category[15] of a predictor variable, relative to that predictor's assigned reference category[16]. To exemplify this on 'Topicalisation' (one of the categories of the AJT Condition predictor), the OR for Topicalisation tells us the comparative odds of getting a higher AJT Response for topicalised AJT items than for AJT items containing Raising-to-Subject control condition - the assigned reference category of this predictor. This measure, therefore, can be used to show if a particular outcome is more or less likely than another. The ORs of each non-reference category of a predictor can then be compared alongside each other. For example, the difference in odds between getting a higher AJT Response for **Topicalisation** compared to Raising-to-Subject can be compared alongside the difference in odds between getting a higher AJT Response for **Left dislocation** compared to Raising-to-Subject. Here we are essentially comparing how much higher or lower the respective ORs are between these categories of this predictor (by comparing both against their shared reference category), which allows us to compare the relative effects of these two categories on the response.

Throughout the discussion in the following chapter, the OR values from the model outputs will be interpreted and discussed as representing **the odds of getting a higher AJT Response** with the inclusion of the given category of the predictor in question - eg. the odds of getting a higher AJT Response with Topicalisation (compared to with Raising-to-Subject AJT items). Because of the proportional odds assumption, the OR values are 'averaged' across all levels of the response scale (discussed in Section 3.9.2.3). This means that the relative increase in odds that is caused by the inclusion of the predictor in question applies **equally** across the AJT Response scale. For example, the odds of Topicalisation

[15] 'Category' and 'level' can be used interchangeably to refer to the 'options' of a categorical predictor - ie. 'topicalisation', 'left dislocation', etc. are categories (or levels) of the AJT Condition predictor, and Raising-to-Subject is the reference category (or level) for this predictor.
[16] These reference categories were manually assigned, as discussed in Section 3.9.2.3.

resulting in a higher AJT Response has the same effect between -2 and -1 on the AJT Response scale as it does between 1 and 2 on the scale. OR values for a given category (eg. Topicalisation) of a predictor (eg. AJT Condition) tell us whether the presence of this category increases or decreases the odds of getting a higher AJT Response, and by how much, and this effect is the same regardless of where you 'start' on the scale. Hence, when we talk about ORs, we can simply say these are '**the odds of getting a higher AJT Response**', or, to put this even more straightforwardly, '**the odds of acceptance**'.

The odds scale centres on the number 1 - an OR of 1 indicates that the odds of getting a higher AJT Response are *not* influenced by the given category of the predictor variable (ie. the given category patterns similarly to its reference category). An OR lower than 1 indicates that the odds of getting a higher AJT Response are *lower* for the given category of the predictor variable compared to its reference category; For example, taking the category of Topicalisation from the focal predictor of AJT Condition, an OR of 0.9 for Topicalisation indicates that the odds of getting a higher AJT Response are 0.9 times (ie. 10% lower) when Topicalisation is present in the AJT item, compared to when the reference category of Raising-to-Subject is present. Finally, an odds ratio greater than 1 indicates that the odds of getting a higher AJT Response are greater for the given category of the predictor variable compared to its reference category; For example, an OR of 1.09 for Topicalisation would indicate that the odds of getting a higher AJT Response are 9% higher when Topicalisation is present in the AJT item, compared to when the reference category of Raising-to-Subject is present. Similarly, an OR of 3.0 would indicate 3 times (300%) higher odds of getting a higher AJT Response with Topicalisation than with the reference category. ORs can also be interpreted in the opposite polarity, ie. in terms of the odds of a response being *lower*, however, this has been avoided for ease of interpretation, so as to avoid as many double negatives as possible. Interaction effects (which, in these models, are all with the focal predictor of AJT Condition), are technically *ratios* of two odds ratios, meaning they require a slightly different interpretation. Let's say, for instance, two different categories of the focal predictor of AJT Condition (eg. Topicalisation and Left Dislocation) each have an interaction effect with the predictor of birth country: If, for 'Polish-born', both Topicalisation and Dislocation have ORs < 1, this means that there is a negative effect on AJT acceptance that is stronger among Polish-born participants compared to English-born participants (the reference category for birth country). Furthermore, if the OR for the Topicalisation interaction is lower than that for the Left Dislocation interaction, then this indicates that Topicalisation has even *lower* odds than Left Dislocation of getting higher AJT Responses among Polish-born participants than English-born ones. As with all regression models, the odds ratio values are only valid if all other variables in the model are held constant.

The tables presented throughout this chapter list truncated model outputs containing only the model terms shown to have significant effects on AJT Response, along with the coefficient estimates (in terms of ordered log-transformed odds, or 'logits') as well as proportional odds ratio values, to two decimal places (exponentiated equivalents of the original log odds estimates), and other elements of the model output (standard errors and z values of the coefficient estimates). P values are given to three decimal places, and these have been Benjamini-Hochberg (BH) adjusted to control for False Discovery Rate (FDR). This is a suitable method due to the large range of predictors being tested simultaneously (see Section 3.9.3 for full justification of this). Model terms followed by '.L' indicate a linear effect. Occasionally, Confidence Intervals (CIs) for the proportional odds ratio values are discussed as part of the explanation of results (though are not reported in the tables presented here - see Appendix 9 for these). These are estimates of the degree of precision of the calculations, and are helpful in determining how confident we can be that the effects found in the recruited participant sample are found throughout the

wider populations represented by these participant groups. A larger margin between the 2.5% and 97.5% CI indicates lower precision of the OR calculation, whereas a small CI range indicates higher precision. Because these results are rather complex to parse, key takeaways will be highlighted throughout the following four sections.

5.2 Model 1 (*all participants*)

To recap from Section 5.1.1, higher ORs indicate that the given category of the predictor is associated with better odds of receiving a higher AJT Response, or, to put this even more straightforwardly, higher ORs indicate better odds of acceptance. Of the AJT Conditions, all six have significant main (ie. direct) effects on AJT Response compared to the reference category (the Raising-to-Subject condition) in Model 1, which includes all participants. Table 5.1 shows the model output for these.

Model term	Model output			p value *(BH adjusted)*	Proportional Odds Ratios (ORs)
	Estimate *(ordered log odds/logits)*	Std. Error	z value		
AJTConditionNS-was	-4.57251	0.46815	-9.76719	0.000	0.01033
AJTConditionNS-were	-4.51699	0.45529	-9.92108	0.000	0.01092
AJTConditionNS-weren't	-3.77340	0.45267	-8.33589	0.000	0.02297
AJTConditionTopicalisation	-3.48267	0.45585	-7.63995	0.000	0.03073
AJTConditionLD	-2.88477	0.46849	-6.15763	0.000	0.05587
AJTConditionRD	-2.35683	0.47068	-5.00730	0.000	0.09472

Table 5.1. *Odds ratio results for the AJT Conditions significant across all participants (Model 1)*

All 6 linguistic conditions are associated with lower odds of getting a higher AJT Response compared to the argument movement positive control condition of raising-to-subject with *seem*. The odds ratios for the non-control conditions range from around 0.01 to around 0.09, meaning that the positive control has almost 100% higher odds of being rated higher. These odds are very low because of the sheer ceiling effect in the positive control condition. Comparing how *much* lower the odds are across each of the AJT Conditions, nonstandard *was* and nonstandard *were* have the lowest odds of receiving a higher AJT Response compared to raising-to-subject (both ORs: 0.01, both with $p = 0.000$). Comparatively, nonstandard *weren't* has slightly better odds, with 99.98% lower odds than raising-to-subject (OR: 0.02; $p = 0.000$). The odds of topicalisation, left dislocation, and right dislocation receiving a higher AJT Response are comparatively better, still (ORs: 0.03, 0.05, and 0.09, respectively, all with $p = 0.000$). Overall, comparing across the six non-reference AJT conditions, these results confirm that the three optional discourse-based movement conditions are more accepted than the three nonstandard past-tense BE agreement ones.

These results are visualised using probability curves in Figure 5.1.

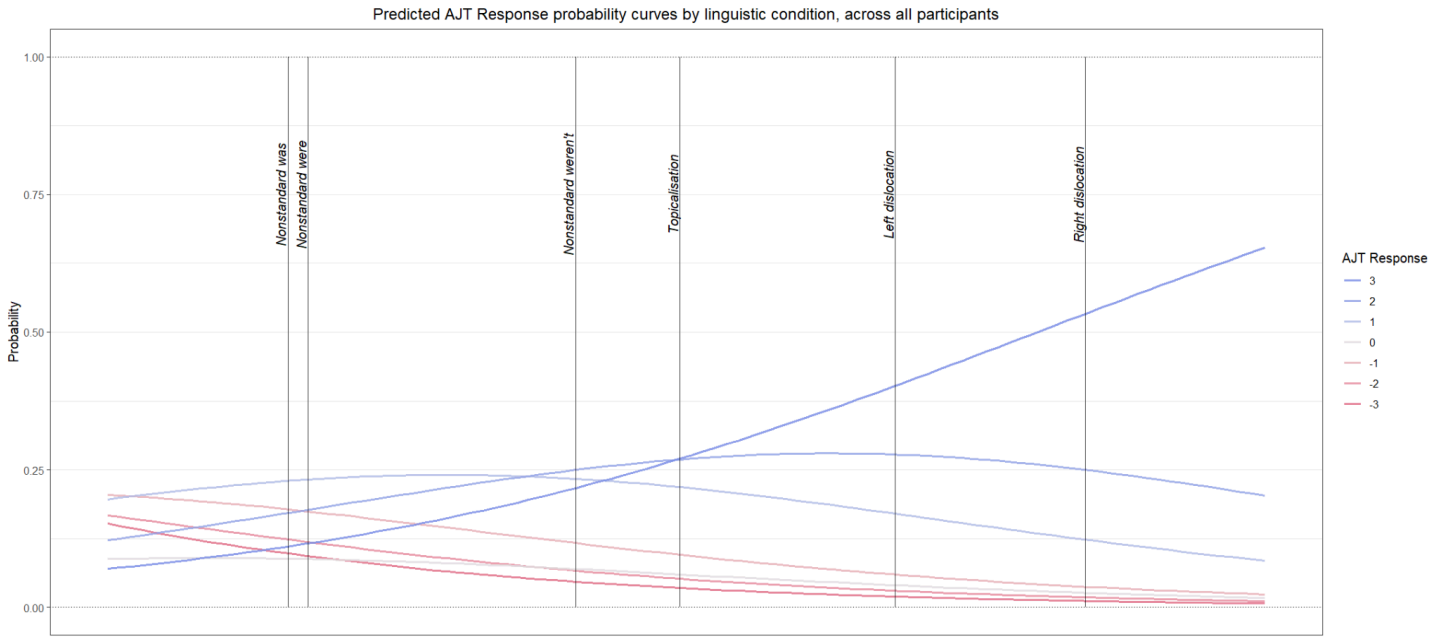


Figure 5.1*. Probability curve of predicted AJT Response by AJT Condition, across all participants*

Turning to nonstandard agreement first, nonstandard *was* and *were* are rated very similarly when factoring in all participants (ie. they are spaced very close together horizontally). Additionally, they show the least clear pattern as to which polarity of the response scale participants tend towards - across both conditions, there is much overlap between positive and negative levels, and the spread of probabilities across the levels of the response scale is very clustered together. For both, an AJT Response of *1* is most probable, with a probability of around 0.25, and responses at the extremities of the scale are least probable, all under 0.125 (though a response of *3* is more probable and a response of *-3* is less probable for nonstandard *were* than for nonstandard *was*). Nonstandard *weren't* diverges from these, with *2* being the most probable response (at a probability of around 0.25) and with a similar probability of a response of *1* or *3*. Participants clearly diverge in their use of the positive versus negative ends of the scale for nonstandard *weren't*, with the three negative responses visibly less probable in Figure 5.1 (all below 0.125) and with *-3* being the least probable.

All three optional discourse-based movement conditions have far higher probabilities of obtaining positive responses (ie. *1* or higher), and lower probabilities of obtaining negative responses (ie. *-1* or lower) compared to nonstandard past-tense BE agreement. For both left dislocation and right dislocation, responses of *3* are by far the most probable increment. Right dislocation is relatively more polarised towards the maximal increment of the scale, having a probability exceeding 0.5 of being rated *3*, compared to around 0.375 for left dislocation, and a lower probability of being rated *2* or *1* compared to left dislocation. This confirms the finding in Table 5.1 that right dislocation is the most accepted of the linguistic conditions. Maximal responses for topicalisation are less probable than for the other two optional discourse-based movement AJT Conditions, equal to those for *2* (both at around 0.25), and a response of *1* is more probable for topicalisation than for the other two AJT Conditions. Negative responses (especially response of *-1*) are also comparatively more probable for topicalisation than for the other two AJT Conditions, though, generally, negative responses are still very improbable across these AJT Conditions.

Overall, Figure 5.1 further illustrates the findings from Table 5.1, that optional discourse-based movement is more accepted than nonstandard past-tense BE agreement.

I now turn to the remaining predictors which have significant effects on participants' AJT Responses (summarised in Table 5.2).

Model term	Model output			p value *(BH adjusted)*	Proportional Odds Ratios (ORs)
	Estimate *(ordered log odds/logits)*	Std. Error	z value		
AJTConditionNS-weren't:birthCountryPolish-born	-1.66808	0.22215	-7.50883	0.000	0.18861
AJTConditionNS-was:birthCountryPolish-born	-1.12257	0.22646	-4.95709	0.000	0.32544
AJTConditionRD:birthCountryPolish-born	-1.06014	0.22812	-4.64720	0.000	0.34641
AJTConditionNS-were:birthCountryPolish-born	-0.95167	0.22168	-4.29294	0.000	0.38610
AJTConditionLD:LGBTQYes	0.54120	0.24000	2.25499	0.044	1.71806
AJTConditionNS-weren't:LGBTQYes	0.74935	0.23351	3.20900	0.004	2.11562
AJTConditionNS-were:LGBTQYes	0.78491	0.23407	3.35330	0.002	2.19220
AJTConditionNS-was:LGBTQYes	0.90469	0.23904	3.78471	0.000	2.47117
AJTConditionRD:LGBTQYes	0.93187	0.24052	3.87435	0.000	2.53926
ageBand.L	0.72376	0.29437	2.45871	0.028	2.06217
AJTConditionNS-weren't:ageBand.L	-1.58539	0.25129	-6.30905	0.000	0.20487
AJTConditionNS-was:ageBand.L	-1.32888	0.25550	-5.20104	0.000	0.26477
AJTConditionNS-were:ageBand.L	-1.21244	0.25104	-4.82972	0.000	0.29747
AJTConditionTopicalisation:ageBand.L	-0.79216	0.25132	-3.15193	0.004	0.45287
AJTConditionNS-weren't:regionNorth	-1.28595	0.35245	-3.64865	0.001	0.27639
AJTConditionNS-was:regionNorth	-1.15672	0.36156	-3.19930	0.004	0.31452
AJTConditionRD:regionNorth	-1.10000	0.36816	-2.98786	0.007	0.33287
AJTConditionNS-were:regionNorth	-0.92983	0.35418	-2.62527	0.018	0.39462
AJTConditionLD:regionNorth	-0.80782	0.36558	-2.20968	0.047	0.44583
AJTConditionNS-weren't:regionSouth	-0.97262	0.32806	-2.96473	0.007	0.37809
AJTConditionRD:regionSouth	-0.78166	0.34320	-2.27754	0.044	0.45765
IMD.L	-0.52167	0.23042	-2.26403	0.044	0.59353
AJTConditionTopicalisation:IMD.L	0.44289	0.19666	2.25211	0.044	1.55721
AJTConditionNS-was:IMD.L	0.57955	0.19818	2.92437	0.008	1.78524
AJTConditionNS-were:IMD.L	0.59301	0.19431	3.05181	0.005	1.80942

Table 5.2. *Significant terms for all participants (Model 1)*

The key sample characteristics of birth country and LGBTQ+ status are considered first. Birth country alone does not have a significant main (ie. direct) effect on AJT Response but has significant interaction effects with 4 of the 6 non-reference AJT Conditions (all except for topicalisation and left dislocation). Among Polish-born participants, the odds of

giving a higher AJT Response are lower compared to those of English-born participants for these four non-reference category AJT Conditions (all four with p = 0.000) - essentially, Polish-born participants are less accepting of these AJT Conditions than English-born participants. Nonstandard *were* has around 61% lower odds (OR: 0.39) of getting a higher AJT Response among Polish-born participants compared to English-born ones, while right dislocation has 65% lower odds (OR: 0.35), and nonstandard *was* has 67% lower odds (OR: 0.33). Nonstandard *weren't* undergoes the largest negative effect with birth country, with 81% lower odds of getting a higher AJT Response amongst Polish-born compared to English-born participants (OR: 0.19). This challenges the pattern found in the descriptive results (Section 4.1.2), where comparative acceptance between Polish-born and English-born participants diverges more for nonstandard past-tense BE agreement than for optional discourse-based movement. This will be discussed in Section 5.7.1.

LGBTQ+ status alone does not have a significant main (ie. direct) effect on AJT Response, but has significant interaction effects with all non-reference AJT conditions except for topicalisation (for which the interaction with LGBTQ+ status was not deemed statistically significant). LGBTQ+ participants have around 2 times higher odds of giving a higher AJT Response to nonstandard *was*, nonstandard *were*, nonstandard *weren't* and right dislocation compared to non-LGBTQ+ participants. Acceptance of left dislocation is affected the least by LGBTQ+ status, though LGBTQ+ participants still have 72% higher odds of giving a higher AJT Responses to left dislocation compared to non-LGBTQ+ participants (OR: 1.72, p = 0.044), and odds for nonstandard *weren't* are 2.12 times higher (OR: 2.12, p = 0.004) amongst LGBTQ+ participants. Nonstandard *were* has 2.19 times higher odds of getting a higher AJT Response among LGBTQ+ participants (OR: 2.19, p = 0.002). The conditions most affected by LGBTQ+ status are nonstandard *was* and right dislocation, both of which have around 2.5 times greater odds of receiving a higher AJT Response from LGBTQ+ participants than non-LGBTQ+ participants (ORs: 2.47 and 2.54, respectively, both with p = 0.000). Overall, the key pattern here is that LGBTQ+ participants have around twice the odds of accepting AJT items across these four AJT conditions compared to non-LGBTQ+ participants.

The macro-social factors which have significant effects on participants' AJT Responses are: age band, region of residence, and socioeconomic status.

The ordinal predictor of age band has a main (ie. direct) effect on AJT Response, and it is linearly associated with approximately two times greater odds of a higher AJT Response with each increase in age band (OR: 2.06; p = 0.028). Essentially, this means that overall acceptance of AJT items increases with age. Age band also has significant *negative* linear interaction effects with all non-reference categories of the AJT Condition predictor (all except for left dislocation and right dislocation, for which the interactions with age band are not deemed statistically significant). Nonstandard *weren't* has 80% lower odds of receiving a higher AJT Response with every increase in age band (OR: 0.20; p = 0.000), while nonstandard *was* has 74% lower odds (OR: 0.26; p = 0.000), and nonstandard *were* has 70% lower odds (OR: 0.30; p = 0.000) of receiving a higher AJT Response as age band increases. Age band has the smallest effect on topicalisation, which has 55% lower odds (OR: 0.45, p = 0.004). In short, the acceptance of nonstandard past-tense BE agreement and topicalisation decrease with age.

Region of residence alone does not have a significant main effect on AJT Response, however, it does significantly interact with AJT Condition. All the ORs for the region interaction terms are <1, so are associated with lower odds of getting a higher AJT Response compared to being in the Midlands. This means that Midlanders are more accepting of all

significant AJT Conditions than Northerners or Southerners are (ranging from around 55%-70% more accepting depending on the variant). Amongst Northerners, five of the six non-reference AJT Conditions show significant effects, with the exception of topicalisation. Nonstandard *weren't* has the lowest odds of receiving a higher AJT Response in the North compared to the Midlands - around 72% lower odds (OR: 0.28, p = 0.001) and left dislocation has the least decreased odds, though still around 55% lower in the North than in the Midlands (OR: 0.45, p = 0.047). Right dislocation, nonstandard *was*, and nonstandard *were* have 60-70% lower odds of being given a higher AJT Response in the North than in the Midlands (ORs of 0.33, 0.31, and 0.39, respectively). Unsurprisingly, participants in the North have better odds of acceptance of nonstandard *were* than nonstandard *was*. Amongst Southerners, only nonstandard *weren't* and right dislocation pattern significantly. Nonstandard *weren't* has 62% lower odds of receiving a higher AJT Response in the South compared to the Midlands (OR: 0.38, p = 0.007) and right dislocation has 54% lower odds (OR: 0.46, p = 0.044), meaning nonstandard *weren't* is less accepted in the South than right dislocation is. Only nonstandard *weren't* and right dislocation are statistically significant predictors for AJT Response in *both* the North and the South, and both conditions have higher relative odds of acceptance in the South than in the North.

Finally, the ordinal socioeconomic status predictor (IMD) has a main (ie. direct) effect on AJT Response. It is linearly associated with 41% lower odds of getting a higher AJT Response- with every increase in socioeconomic status (OR: 0.59, p = 0.044). Essentially, this means that overall acceptance of test items in the AJT decreases with higher socioeconomic status. Socioeconomic status also has a linear interaction effect with the AJT Conditions nonstandard *was*, nonstandard *were* and topicalisation. Higher socioeconomic status increases the odds of topicalisation getting a higher AJT Response by 56% (OR: 1.56, p = 0.044); increases the odds of nonstandard *was* getting a higher AJT Response by 79% (OR: 1.79, p = 0.008); and increases the odds of nonstandard *were* getting a higher AJT Response by 81% (OR: 1.81, p = 0.005). Overall, the pattern here is that acceptance of these three linguistic conditions increases with higher socioeconomic status.

5.3 Model 2a (*Polish-born participants only*)

Narrowing the participant sample down to only Polish-born participants, Table 5.3 shows that four of the six non-reference categories of AJT Condition (all except left dislocation and topicalisation) have significant main effects on AJT Response compared to the reference category (ie. Raising-to-Subject).

Model term	Model output			p value *(BH adjusted)*	Proportional Odds Ratios (ORs)
	Estimate *(ordered log odds/logits)*	Std. Error	z value		
AJTConditionNS-was	-5.23975	0.78650	-6.66215	0.000	0.00530
AJTConditionNS-weren't	-4.50961	0.76815	-5.87071	0.000	0.01100
AJTConditionNS-were	-3.74237	0.76687	-4.88003	0.000	0.02370
AJTConditionRD	-2.16853	0.79946	-2.71250	0.023	0.11435

Table 5.3*. Odds ratio results for the four AJT Conditions significant across Polish-born participants (Model 2a)*

Again, all non-reference AJT Conditions have lower odds of getting a higher AJT Response than raising-to-subject, as expected. Comparing how *much* lower the odds are across each condition, nonstandard *was* has the lowest odds of receiving a higher AJT Response compared to raising-to-subject - around 100% lower; OR: 0.00, p = 0.000). Comparatively, nonstandard *weren't* and nonstandard *were* have slightly better odds - around 99% and 98% lower than raising-to-subject, respectively; OR: 0.01 and 0.02, respectively, both with p = 0.000). Right dislocation has the highest comparative odds of receiving a higher AJT Response (OR: 0.11, p = 0.023). Overall, these results echo those found across all participants (in Model 1), in that the three nonstandard past-tense BE agreement conditions are less accepted by Polish-born participants compared to the optional discourse-based movement condition of right dislocation.

These results are visualised using probability curves in Figure 5.2.

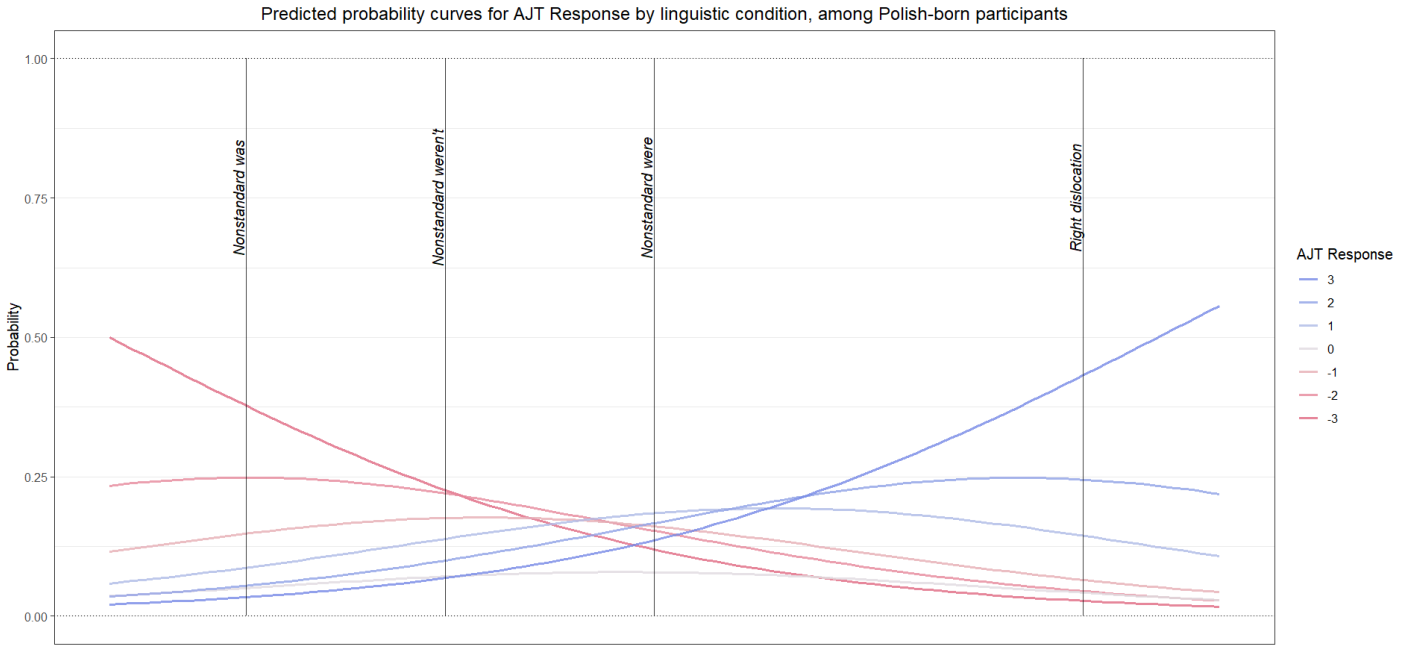


***Figure 5.2**. Probability curve of predicted AJT Response by AJT Condition, across Polish-born participants*

Turning to the nonstandard past-tense BE agreement conditions first, these are each rated differently on the AJT Response scale by Polish-born participants (ie. they are horizontally positioned fairly far apart). Nonstandard *was* has the widest spread of probabilities across the levels of the response scale, whereas nonstandard *weren't* has a smaller spread, and nonstandard *were* has very little spread. For both nonstandard *was* and nonstandard *weren't*, negative responses are more probable than positive responses are. Interestingly, this means that, when rating nonstandard *weren't*, Polish-born participants use the AJT Response scale in the opposite way to the pattern found across all participants (in Model 1), where positive responses are more probable than negative. *-3* is the most probable response to nonstandard *was* (with a probability of 0.375), and this response increment is much less probable for nonstandard *weren't*, for which *-3* and *-2* are equally probable (at under 0.25). *3* is the least probable response for both AJT Conditions, though it is slightly more probable for nonstandard *weren't* than for nonstandard *was*. For nonstandard *were*, there is a far less clear pattern as to which polarity of the response scale Polish-born participants tend, mirroring that found across all participants (in Model 1). The probabilities for the response levels are very clustered, all with probabilities of under 0.25, and an AJT Response of *1* being most probable (though not by much). Responses at the extremities of the scale are least probable, both under 0.125 (though a response of *3* is more probable for nonstandard *was* than a response of *-3* is).

Turning to the optional discourse-based movement AJT Conditions, of which only right dislocation proved significant in this model, this has a far higher probability of obtaining positive responses (ie. *1* or higher), and lower probabilities of obtaining negative responses (ie. *-1* or lower) compared to the nonstandard past-tense BE agreement AJT Conditions. The probability distribution for right dislocation mirrors that found across all participants (in Figure 5.1, from Model 1). Right dislocation is highly polarised towards the maximal increment of the scale, having a probability of around 0.44 of it being rated *3*.

Overall, Figure 5.2 further illustrates the findings from Table 5.3, that the optional discourse-based movement AJT Condition of right dislocation is more accepted by Polish-born participants than the three nonstandard past-tense BE agreement conditions are.

Next, I turn to other predictors which significantly affect Polish-born participants' AJT Responses (summarised in Table 5.4).

Model term	Model output			p value *(BH adjusted)*	Proportional Odds Ratios (ORs)
	Estimate *(ordered log odds/logits)*	Std. Error	z value		
AJTConditionLD:LGBTQYes	1.14633	0.37778	3.03438	0.010	3.14661
AJTConditionRD:LGBTQYes	1.45693	0.37610	3.87379	0.001	4.29275
AJTConditionNS-were:LGBTQYes	1.79096	0.37131	4.82329	0.000	5.99518
AJTConditionNS-was:LGBTQYes	1.85238	0.37801	4.90039	0.000	6.37497
AJTConditionNS-weren't:LGBTQYes	1.90398	0.37186	5.12016	0.000	6.71256
ageBand.L	-1.45966	0.54700	-2.66847	0.025	0.23232
AJTConditionTopicalisation:ageBand.L	1.46075	0.47992	3.04371	0.010	4.30917
AJTConditionRD:ageBand.L	1.70460	0.49159	3.46750	0.003	5.49916
AJTConditionLD:ageBand.L	1.70773	0.49296	3.46425	0.003	5.51643
AJTConditionNS-were:regionNorth	-2.13839	0.66043	-3.23786	0.006	0.11784
AJTConditionNS-weren't:regionNorth	-2.07936	0.65718	-3.16407	0.007	0.12501
AJTConditionNS-weren't:regionSouth	-1.51262	0.62222	-2.43102	0.047	0.22033
higherEdYes	1.50528	0.44683	3.36883	0.004	4.50543
AJTConditionTopicalisation:higherEdYes	-2.30915	0.38711	-5.96507	0.000	0.09935
AJTConditionLD:higherEdYes	-2.07696	0.40389	-5.14240	0.000	0.12531
AJTConditionRD:higherEdYes	-2.01399	0.39548	-5.09253	0.000	0.13346
AJTConditionNS-were:higherEdYes	-1.97822	0.38710	-5.11034	0.000	0.13832
AJTConditionNS-weren't:higherEdYes	-1.46444	0.39047	-3.75051	0.001	0.23121
AJTConditionNS-was:higherEdYes	-1.14212	0.40065	-2.85066	0.017	0.31914
AJTConditionRD:lifestyleEnglish.L	1.18605	0.38642	3.06933	0.009	3.27412
AJTConditionNS-was:lifestyleEnglish.L	1.26664	0.38757	3.26814	0.005	3.54890

Table 5.4*. Significant model terms for Polish-born participants only (Model 2a)*

The key sample characteristic of LGBTQ+ status does not have a significant main (ie. direct) effect on AJT Response amongst Polish-born participants, but has significant interaction effects with 5 of the 6 non-reference AJT Conditions (all except for topicalisation). Right dislocation has around 4.3 times higher odds of receiving a higher AJT Response among LGBTQ+ Polish-born compared to non-LGBTQ+ Polish-born participants (OR: 4.30, $p = 0.001$). Nonstandard *was*, nonstandard *were*, and nonstandard *weren't* have 6 or more times higher odds of receiving a higher AJT Response by LGBTQ+ Poles (ORs: 6.37, 6.00, and 6.71, respectively, all with $p = 0.000$). Left dislocation has the smallest

increase in odds of receiving a higher AJT Response among LGBTQ+ compared to non-LGBTQ+ Poles, though this is still over a three times increase (OR: 3.15, p = 0.010). The overall pattern here is that among Polish-born participants, being LGBTQ+ is associated with a large increase in acceptance across all significant AJT Conditions, compared to not being LGBTQ+.

Among Polish-born participants, the macro-social factors which have significant effects are: age band, region, and higher-educated status.

The ordinal predictor age band seems to have a main (ie. direct) effect on AJT Response, and it is linearly associated with 77% lower odds of a higher AJT Response with every increase in age band (OR: 0.23, p = 0.025). Essentially, this means that, among Poles, overall acceptance of AJT items decreases with age. Turning to the interaction between age band and AJT Condition, there is a significant *positive* linear association between increase in age band and odds of getting a higher AJT Responses for the left dislocation, right dislocation, and topicalisation AJT Conditions. Right dislocation and left dislocation have around 5.5 times greater odds of receiving a higher AJT Response with increase in age band (ORs: 5.50 and 5.51, respectively, both with p = 0.003). Topicalisation has slightly less raised (though still over four times greater) odds of receiving a higher AJT Response with increase in age band (OR: 4.31, p = 0.010). The overall pattern here is that, amongst Poles, the acceptance of the three optional discourse-based movement conditions increases with age.

Region does not have a significant main (ie. direct) effect on AJT Response, however, there is an interaction effect with AJT Condition. All the ORs for the region interaction terms are <1, so are associated with lower odds of getting a higher AJT Response compared to being in the Midlands, meaning Polish-born Midlanders are comparatively more accepting of all significant AJT Conditions than Northerners or Southerners are (ranging from around 80%-90% more accepting depending on the variant). This disparity in acceptance between the Midlands and the other two regions is larger among Polish-born participants than was found across all participants (in Model 1), where Midlanders are only around 55%-70% more accepting). Only nonstandard *were* and nonstandard *weren't* are significant among Polish-born participants in the North and have similar acceptance levels, both conditions having around 87-88% lower odds of receiving a higher AJT Response in the North compared to the Midlands (ORs: 0.12 and 0.13, p = 0.006 and 0.007, respectively). Nonstandard *weren't* has around 78% lower odds of receiving a higher AJT Response in the South than in the Midlands (ORs: 0.22, p = 0.047). Comparing this condition between the South and the North, being a Southern resident has a relatively smaller negative effect on acceptance of nonstandard *weren't* (compared to being in the Midlands) than being a Northern resident does - ie. Polish-born Southerners are relatively more accepting of nonstandard *weren't* than Polish-born Northerners are.

The binary predictor of higher-educated status has a main (ie. direct) effect on AJT Response, with participants who have undergone higher education having around 4.5 times higher odds of giving a higher AJT Response compared to those who have not (OR: 4.51, p = 0.004). Essentially, this means that overall acceptance of test items in the AJT is higher amongst participants who are higher-educated than those who are not. Turning to the interaction between higher education and AJT Condition, across all non-reference category AJT Conditions, there are significant *negative* associations between higher education attendance and odds of receiving a higher AJT Response. Topicalisation has the largest decrease in odds of receiving a higher AJT Response among higher educated participants compared to those who have not undergone higher education - around 90% lower; OR: 0.10, p = 0.000). Left dislocation, right dislocation, and nonstandard *were* have around 86-88% lower odds of receiving a higher AJT Response among higher educated participants compared to those who have not

undergone higher education (ORs: 0.12, 0.13 and 0.14, respectively, all with p = 0.000). Nonstandard *weren't* has around 77% lower odds (OR: 0.23, p = 0.001). Nonstandard *was* has the lowest decrease in odds of receiving a higher AJT Response by higher educated compared to non-higher educated participants - around 68% lower (OR: 0.32, p = 0.017). The overall pattern here is that higher-educated Poles are less accepting of nonstandard *were* and the three optional discourse-based movement conditions compared to non higher-educated Poles.

In terms of L2 factors, the ordinal predictor measuring the length of English language exposure - a proxy measure for English proficiency - does not have a significant main (ie. direct) effect on AJT Response, nor does it have a significant linear interaction with the focal predictor of AJT Condition[17]. Therefore, I will instead test the effect of the formal measure of English proficiency (ie. CEFR classification) in Model 2b.

As for meso-social factors, the ordinal predictor of English Lifestyle (a proxy for English acculturation) does not have a significant linear main (ie. direct) effect on AJT Response amongst Polish-born participants. However, there are significant positive linear associations between increasing English Lifestyle and greater odds of getting a higher AJT Response for two of the AJT Conditions - right dislocation and nonstandard *was*. Right dislocation has over 3 times greater odds of receiving a higher AJT Response with increasing English Lifestyle (OR: 3.27, p = 0.009), and nonstandard *was* has even greater odds (OR: 3.55, p = 0.005). Overall, the pattern here is that, amongst Poles, acceptance of right dislocation and nonstandard *was* increases with English Lifestyle (ie. acculturation).

[17] The model does suggest significant negative quadratic associations (ie. deceleration effects) between increasing length of English language exposure and lower odds of AJT acceptability for nonstandard *weren't* and topicalisation (ORs: 0.25 and 0.23, p = 0.043 and 0.023, respectively). Length of English language exposure also appears to have a significant negative cubic association with nonstandard *was*, ie. a curve with a changing rate of deceleration as length of English language exposure increases (OR: 0.30, p = 0.023). These can be seen in the Model 1 output in Appendix 9.1. However, due to the practicalities of interpretability, these are not engaged with here, in favour of the formal English proficiency measure.

5.4 Model 2b (*Polish-born participants with CEFR only*)

Next, I turn to Model 2b, which incorporates only Polish-born participants who took the Versant English Speaking Test, in order to specifically interrogate the effects of the more formal measure of English linguistic proficiency (as opposed to the proxy measure of length of English language exposure used in Model 2a). Versant test scores have been translated into their equivalent CEFR ratings, as this scale is more universally interpretable (see Section 3.8.1.2 for discussion of this). Although all other predictors were included in order to ensure that factors incorporated in the model were held constant, the English proficiency (CEFR) predictor is the only one from the Model 2b output which will be focused on here.

Model term	Model output			p value *(BH adjusted)*	Proportional Odds Ratios (ORs)
	Estimate *(ordered log odds/logits)*	Std. Error	z value		
CEFR.L	2.98407	0.99176	3.00887	0.012	19.76817
AJTConditionTopicalisation:CEFR.L	-4.27074	0.87451	-4.88358	0.000	0.01397
AJTConditionNS-was:CEFR.L	-3.12818	0.91600	-3.41505	0.005	0.04380
AJTConditionNS-weren't:CEFR.L	-2.81314	0.86983	-3.23415	0.007	0.06002
AJTConditionRD:CEFR.L	-2.80904	0.87939	-3.19430	0.007	0.06026
AJTConditionNS-were:CEFR.L	-2.26475	0.86707	-2.61196	0.034	0.10386

Table 5.5. *Significant model terms for Polish-born participants with CEFR rating (Model 2b)*

Table 5.5 shows that the ordinal predictor of English proficiency (CEFR) has a main (ie. direct) effect on AJT Response, and the model output indicates a positive linear association (the OR is >1). However, the relationship is unrealistically strong, with almost 20 times higher odds of a higher AJT Response with every increase in CEFR band (OR: 19.77, p = 0.012). Also, the upper CI for the OR calculation is enormous (2.83 to 138.09 at the 97.5% CI, to 2dp). This could be due to the relatively lower sample size for this predictor and the fact it has been collapsed into only two bands, therefore, I will focus on the interaction between CEFR and AJT Condition, which can be interpreted more reliably. There are significant negative linear associations between increase in CEFR band and odds of higher AJT Responses for all AJT Conditions except left dislocation. Nonstandard *were* has around 90% lower odds of receiving a higher AJT Response with increase in CEFR band (OR: 0.10, p = 0.034). Nonstandard *weren't* and right dislocation have around 94% lower odds of receiving a higher AJT Response with increase in CEFR band (ORs: 0.06 and 0.06, respectively, both with p = 0.007). Nonstandard *was* has around 96% lower odds of receiving a higher AJT Response with increase in CEFR band (OR: 0.04), p = 0.005). Topicalisation undergoes the strongest negative effect, with 99% lower odds of receiving a higher AJT Response with increase in CEFR band (OR: 0.01, p = 0.000). Overall, the pattern here is that higher English proficiency is associated with lower acceptance of these five significant AJT Conditions.

5.5 Model 3 (*LGBTQ+ participants only*)

Narrowing the participant sample down to only LGBTQ+ participants, Table 5.6 shows that five of the six non-reference AJT Conditions (all except right dislocation) have significant main (ie. direct) effects on AJT Response compared to the reference category (the argument movement condition of raising-to-subject with *seem*).

Model term	Model output			p value *(BH adjusted)*	Proportional Odds Ratios (ORs)
	Estimate *(ordered log odds/logits)*	Std. Error	z value		
AJTConditionNS-was	-4.28381	0.76987	-5.56436	0.000	0.01379
AJTConditionNS-were	-3.17781	0.75431	-4.21289	0.000	0.04168
AJTConditionTopicalisation	-2.80737	0.76322	-3.67832	0.001	0.06036
AJTConditionNS-weren't	-2.78297	0.75757	-3.67353	0.001	0.06186
AJTConditionLD	-1.93168	0.78684	-2.45499	0.042	0.14491

Table 5.6*. Odds ratio results for the AJT Conditions significant across LGBTQ+ participants (Model 3)*

Again, all non-reference AJT Conditions have lower odds of getting higher AJT Responses than the argument movement condition of raising-to-subject, as expected. However, it is possible to compare how *much* lower the odds are for getting higher AJT Responses across each condition. Nonstandard *was* has the lowest odds of receiving a higher AJT Response compared to Raising-to-Subject - around 99% lower; OR: 0.01, p = 0.000) and nonstandard *were* has slightly higher odds - around 96% lower (OR: 0.04, p = 0.000). Nonstandard *weren't* and topicalisation have around 94% lower odds of getting a higher AJT Response than raising-to-subject (both ORs: 0.06, both with p = 0.001). Finally, the left dislocation AJT Condition has the best comparative odds of receiving a higher AJT Response (at around 85% lower than Raising-to-Subject; OR: 0.15, p = 0.042). LGBTQ+ participants are the only sample for which there is an (albeit marginal) overlap in the relative odds of acceptance between the optional discourse-based movement and nonstandard agreement AJT Conditions - topicalisation and nonstandard *weren't* have very similar OR results, with topicalisation (an optional discourse-based movement AJT Condition) having slightly lower acceptance than the nonstandard past-tense BE agreement condition of nonstandard *weren't*.

These results are visualised using probability curves in Figure 5.3.

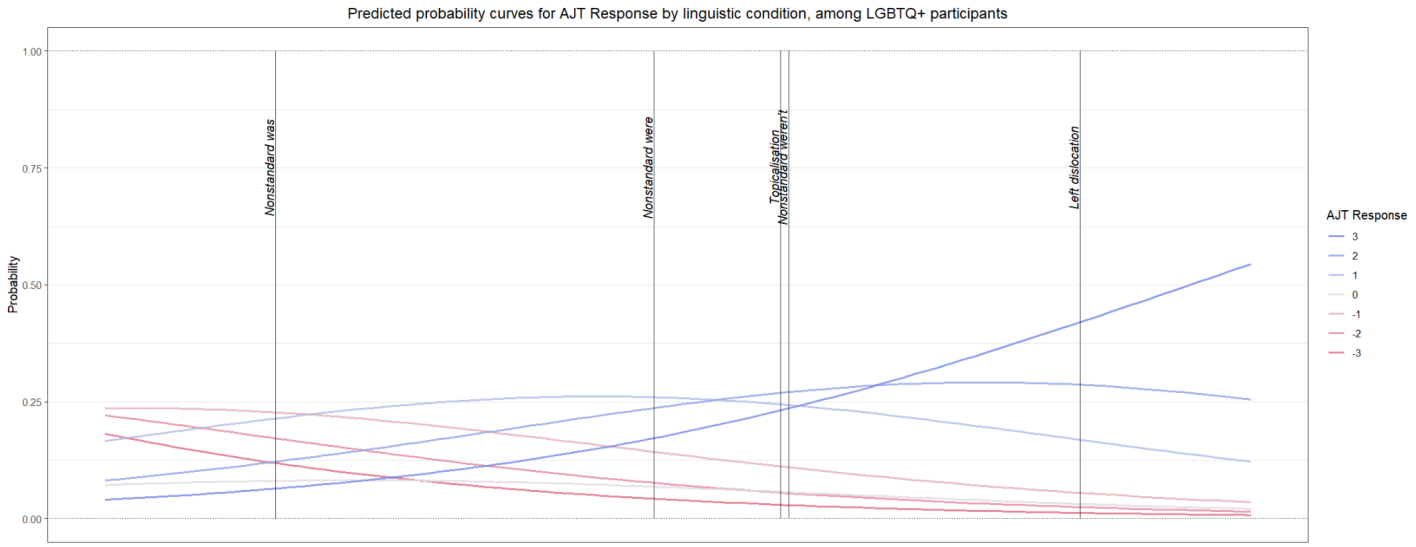


Figure 5.3*. Probability curve of predicted AJT Response by AJT Condition, across LGBTQ+ participants*

Turning to the nonstandard agreement AJT Conditions first, these are each rated differently on the AJT Response scale by LGBTQ+ participants (ie. they are horizontally positioned fairly far apart). Nonstandard *was* has the narrowest spread of probabilities across the levels of the response scale, and shows the least clear pattern as to which polarity of the response scale participants tend towards, with much overlap between positive and negative levels. It is the only AJT Condition for which *3* is the least probable response, and the only AJT Condition for which the most probable response is negative (*-1*). It is the AJT Condition with the highest probability of a response of *-3* (at 0.125). It is also the AJT Condition for which the distribution of probabilities for the levels of the response scale pattern strikingly differently among LGBTQ+ participants to that across all participants (in Model 1); In Model 1, *-3* is the least probable response for nonstandard *was* and the most probable is positive (*1*). LGBTQ+ participants diverge in their use of the positive versus negative ends of the scale for nonstandard *were* and nonstandard *weren't*, with the three negative responses visibly less probable in Figure 5.3 (all below 0.125) and with *-3* being the least probable. Nonstandard *weren't* has a higher probability of being rated *3* than nonstandard *were* does (around 0.25 compared to around 0.125), and a lower probability of receiving any of the three negative responses compared to nonstandard *were*. The most probable response for nonstandard *weren't* is *2*, while, for nonstandard *were*, it is *1*.

In terms of the optional discourse-based movement AJT Conditions, only topicalisation and left dislocation proved significant in this model. Left dislocation is highly polarised towards the maximal increment of the scale, having the highest probability of all AJT Conditions of being rated *3* (at around 0.44), and the lowest probabilities of receiving any of the three negative responses (all under 0.062). Responses of *3* are much less probable for topicalisation than for left dislocation, at around 0.25, and *2* is the most probable response for topicalisation, at just over 0.25. Additionally, topicalisation has slightly higher probabilities of receiving negative responses than left dislocation.

Overall, Figure 5.3 further illustrates the findings from Table 5.6 that, although in general, LGBTQ+ participants are more accepting of the two optional discourse-based movement conditions than the three nonstandard past-tense BE agreement ones, there is a marginal overlap, with topicalisation slightly less accepted than nonstandard *weren't*.

I now turn to other predictors which significantly affect LGBTQ+ participants' AJT Responses (summarised in Table 5.7).

Model term	Model output			p value *(BH adjusted)*	Proportional Odds Ratios (ORs)
	Estimate *(ordered log odds/logits)*	Std. Error	z value		
AJTConditionNS-weren't:birthCountryPolish-born	-1.17969	0.32020	-3.68425	0.001	0.30737
ageBand.L	1.26020	0.49046	2.56944	0.035	3.52614
AJTConditionNS-weren't:ageBand.L	-2.26088	0.43236	-5.22919	0.000	0.10426
AJTConditionNS-was:ageBand.L	-1.94492	0.43467	-4.47450	0.000	0.14300
AJTConditionNS-were:ageBand.L	-1.80878	0.43427	-4.16506	0.000	0.16385
AJTConditionTopicalisation:ageBand.L	-1.67494	0.42822	-3.91144	0.001	0.18732
genderNotBinary	1.64839	0.61744	2.66972	0.027	5.19861
AJTConditionTopicalisation:genderNotBinary	-1.73635	0.54251	-3.20060	0.007	0.17616
AJTConditionRD:genderNotBinary	-1.32279	0.54997	-2.40522	0.043	0.26639
AJTConditionNS-weren't:regionNorth	-2.19487	0.56289	-3.89928	0.001	0.11137
AJTConditionNS-were:regionNorth	-2.03410	0.56024	-3.63077	0.002	0.13080
AJTConditionNS-was:regionNorth	-1.73040	0.57006	-3.03547	0.011	0.17721
AJTConditionTopicalisation:regionNorth	-1.71596	0.57049	-3.00787	0.012	0.17979
AJTConditionRD:regionNorth	-1.58053	0.59048	-2.67668	0.027	0.20587
AJTConditionNS-weren't:regionSouth	-1.17505	0.48513	-2.42215	0.043	0.30880
AJTConditionNS-were:regionSouth	-1.14579	0.48073	-2.38343	0.043	0.31797
IMD.L	-0.74556	0.21531	-3.46274	0.003	0.47447
AJTConditionRD:higherEdYes	-1.11635	0.43835	-2.54668	0.035	0.32747
AJTConditionLD:higherEdYes	-1.02676	0.43268	-2.37304	0.043	0.35817
AJTConditionRD:lifestyleEnglish.L	2.12848	0.47864	4.44694	0.000	8.40205
AJTConditionLD:LGBTQCommunityInvolvement.L	-0.91460	0.30568	-2.99208	0.012	0.40068
AJTConditionTopicalisation:LGBTQCommunityInvolvement.L	-0.88639	0.29761	-2.97838	0.012	0.41214
AJTConditionRD:LGBTQCommunityInvolvement.L	-0.72909	0.30500	-2.39046	0.043	0.48235

Table 5.7. *Significant model terms for LGBTQ+ participants only (Model 3)*

The key sample characteristic of birth country does not have a significant main (ie. direct) effect on AJT Response among LGBTQ+ participants, but has a significant interaction with the AJT Condition of nonstandard *weren't*; Polish-born participants have 69% lower odds of giving a higher AJT Response for nonstandard *weren't* compared to English-born participants (OR: 0.31, $p = 0.001$). Essentially, Polish-born participants are less accepting of nonstandard *weren't* compared to English-born participants.

Among LGBTQ+ participants, the macro-social factors which have significant effects are: age band, gender, region, socioeconomic status, and higher-educated status.

The ordinal predictor of age band has a main (ie. direct) effect on AJT Response, with a positive linear association of 3.5 times higher odds of giving a higher AJT Response with every increase in age band (OR: 3.53, p = 0.035). Essentially, this means that, among LGBTQ+ participants, overall acceptance of AJT items increases with age. Turning to the interaction between age band and AJT Condition, there is a significant *negative* linear association between increase in age band and odds of getting a higher AJT Response for nonstandard *was*, nonstandard *were*, nonstandard *weren't*, and topicalisation. Topicalisation has around 81% lower odds of receiving a higher AJT Response with increase in age band (OR: 0.19, p = 0.001). Nonstandard *were*, nonstandard *was*, and nonstandard *weren't* have even lower odds of receiving a higher AJT Response with increase in age band - 84%, 86%, and 90% lower, respectively (ORs: 0.16, 0.14, and 0.10, respectively, all with p = 0.000). The overall pattern here is that, amongst LGBTQ+ participants, the acceptance of the three nonstandard past-tense BE agreement conditions and topicalisation decreases with age.

Gender, specifically, being outside the gender binary, has a main (ie. direct) effect on AJT Response amongst LGBTQ+ participants. Female gender is not found to have significantly different effects to male but participants who are not binary, on the other hand, have over 5 times higher odds of giving a higher AJT Response compared to male participants (OR: 5.20, p = 0.027). Right dislocation has around 73% lower odds of receiving a higher AJT Response among participants outside the binary compared to men (OR: 0.27, p = 0.043). Topicalisation has even lower odds (82% lower) of getting a higher AJT Response from participants outside the gender binary compared to men (OR: 0.18, p = 0.007). The overall pattern here is that, amongst LGBTQ+ participants, participants outside the gender binary are less accepting than men of the discourse-pragmatic conditions of right dislocation and topicalisation.

Region does not have a significant main (ie. direct) effect on AJT Response, however, it does interact with AJT Condition. All the ORs for the region interaction terms are <1, so are associated with lower odds of getting a higher AJT Response compared to being in the Midlands, meaning LGBTQ+ Midlanders are comparatively more accepting of all significant AJT Conditions than Northerners or Southerners are (ranging from around 70%-90% more accepting depending on the variant). Amongst LGBTQ+ Northerners, five of the six non-reference AJT Conditions show significant effects, with the exception of left dislocation. Of these, nonstandard *weren't* has the lowest odds of receiving a higher AJT Response by LGBTQ+ participants in the North compared to the Midlands - around 89% lower (OR: 0.11, p = 0.001). Nonstandard *were* has slightly better odds, at 87% lower (OR: 0.13, p = 0.002), and topicalisation and nonstandard *was* have even better odds - around 82% lower than the Midlands (both ORs: 0.18, p = 0.012 and 0.11, respectively). Right dislocation comparatively has the best odds of getting a higher AJT Response in the North compared to the Midlands, at 79% lower (OR: 0.21, p = 0.027). Amongst Southerners, only the nonstandard past-tense BE agreement AJT Conditions of nonstandard *were* and nonstandard *weren't* pattern significantly. Both have around 68-69% lower odds of getting a higher AJT Response in the South than the Midlands (ORs: 0.32 and 0.31, respectively, both with p = 0.043), with nonstandard *were* having marginally better odds than nonstandard *weren't*. Comparing these conditions between the South and the North, being a Southern resident results in a relatively smaller negative effect on acceptance of nonstandard *were* and nonstandard *weren't* (compared to being in the Midlands) than being a Northern resident does - ie. LGBTQ+ Southerners are relatively more accepting of nonstandard *were* and nonstandard *weren't* than LGBTQ+ Northerners are.

The ordinal socioeconomic status predictor (IMD) has a main (ie. direct) effect on AJT Response. It is linearly associated with lower odds of getting a higher AJT Responses - 53% lower odds with every increase in socioeconomic status (OR: 0.47, p = 0.003). Essentially, this means that LGBTQ+ participants of higher socioeconomic status are less accepting of AJT items, overall. The stepwise elimination process (using *buildmer*) did not deem the interaction between socioeconomic status and AJT Condition to be significant among LGBTQ+ participants, so it was not included in the model.

The binary predictor of higher-educated status does not have a significant main (ie. direct) effect on AJT Response among LGBTQ+ participants, however, there is an interaction between higher-educated status and AJT Condition. Left dislocation has around 64% lower odds of receiving a higher AJT Response among higher-educated participants compared to those who have not undergone higher education, and right dislocation has around 67% lower odds (ORs: 0.36 and 0.33, p = 0.043 and 0.035, respectively). This means that acceptance of right dislocation is more strongly affected by higher education than acceptance of left dislocation. The overall pattern here is that higher-educated LGBTQ+ participants are less accepting of left dislocation and right dislocation compared to non higher-educated LGBTQ+ participants.

In terms of meso-social factors, the ordinal predictor of English Lifestyle (a proxy for English acculturation) does not have a significant linear main (ie. direct) effect on AJT Response amongst LGBTQ+ participants. However, there is a significant positive linear association between increasing English Lifestyle and greater odds of getting a higher AJT Response for right dislocation. Right dislocation has over 8 times higher odds of getting a higher AJT Response with every increase in English Lifestyle (OR: 8.40, p = 0.000). This seems unrealistically strong, and with a raised 97.5% Confidence Interval of 21.47, it is not possible to be confident about the magnitude of this association. Nonetheless, the model suggests that an effect is present for this AJT Condition. Overall, the pattern here is that, amongst LGBTQ+ participants, acceptance of right dislocation increases with English Lifestyle (ie. acculturation).

The ordinal predictor of LGBTQ+ Community Involvement does not have a significant linear main (ie. direct) effect on AJT Response. However, there are significant negative linear associations between increasing LGBTQ+ Community Involvement and lower odds of getting a higher AJT Response for the optional discourse-based movement AJT Conditions of left dislocation, right dislocation, and topicalisation. Topicalisation and left dislocation have 59-60% lower odds of receiving a higher AJT Response with increasing LGBTQ+ Community Involvement (ORs: 0.41 and 0.40, respectively, and both with p = 0.012), while right dislocation has a lower decrease in odds - 52% lower (OR: 0.48, p = 0.043). Overall, the pattern here is that, amongst LGBTQ+ participants, acceptance of the three optional discourse-based movement AJT Conditions decreases with increasing LGBTQ+ Community Involvement.

5.6 Discussion of inferential results

I will now summarise and discuss the findings for the four statistical models, comparing equivalent results between models. This is in order to explore how different factors influence participants' acceptability judgements, and compare the differences in effects found across Polish-born participants and LGBTQ+ participants with those found across all participants. I will first discuss overarching patterns in the acceptability of the morphosyntactic features under investigation according to the key sample characteristics of birth country and LGBTQ+ status. I then compare in more detail how different features pattern across the nested models, with regard to the type of morphosyntactic construction - ie. argument movement; optional discourse-based movement; nonstandard agreement. Following this, I discuss patterns in acceptability with regards to each of the factor groups in turn - macro-social factors, L2 factors, and meso-social factors. For ease of reference, patterns of acceptability (by variant function) according to each of these types of factors are summarised in tables in each respective section. An overall conclusion is then given, synthesising results from across the four models. The effects of each factor on the patterns of acceptability for each morphosyntactic variant are also reiterated in Table form (Table 5.12).

5.6.1 Key Sample Characteristics

5.6.1.1 Birth country

Across all participants (Model 1), the key sample characteristics of birth country and LGBTQ+ status influence how participants rate specific variants. English-born participants are significantly more accepting than Polish-born participants of all significant variants - nonstandard *weren't*, nonstandard *was,* nonstandard *were*, and right dislocation. The disparity between the two birth countries is particularly noticeable in the treatment of nonstandard *weren't*: Polish-born and English-born participants have even more disparity in acceptance compared to the other significant variants. Across LGBTQ+ participants (Model 3), nonstandard *weren't* is the only variant significantly influenced by birth country: Polish-born LGBTQ+ participants are significantly less accepting compared to English-born LGBTQ+ participants.

5.6.1.2 LGBTQ+ status

Across all participants (Model 1), LGBTQ+ participants are around 2 times more accepting of all significant variants - nonstandard *was,* nonstandard *were*, nonstandard *weren't*, left dislocation, and right dislocation - compared to non-LGBTQ+ participants. Acceptance of left dislocation is the least improved by positive LGBTQ+ status. Acceptance of nonstandard *weren't* and nonstandard *were* are slightly more improved. Nonstandard *was* acceptance is the next most improved , and, finally, acceptance of right dislocation is most improved by being LGBTQ+. Among Polish-born participants (Model 2a), the same variants are significant, and LGBTQ+ status has an even stronger positive effect on acceptance of these variants among Polish-born participants than that found across all participants in Model 1. The acceptance of left dislocation is comparatively improved the *least* by Polish-born participants being LGBTQ+, followed by right dislocation. The acceptability of nonstandard *were* is improved more by being LGBTQ+ than the two dislocation conditions, but not as much as nonstandard *was* is. Finally, acceptance of nonstandard *weren't* is most positively affected by Poles being LGBTQ+.

Predictor	All participants (Model 1)	Polish-born (Models 2a & 2b)	LGBTQ+ (Model 3)
Birth country	Decrease in acceptance for Polish-born compared to English-born across nonstandard past-tense BE agreement and right dislocation	N/A	Decrease in acceptance of nonstandard *weren't* for Polish-born LGBTQ+ compared to English-born LGBTQ+
LGBTQ+ status	Increase in acceptance from those with LGBTQ+ status across nonstandard past-tense BE agreement and right dislocation	Even bigger increase in acceptance across nonstandard past-tense BE agreement and right dislocation for those with Polish-born LGBTQ+ status, compared to those without LGBTQ+ status	N/A

Table 5.8. *Summary of inferential results across key sample characteristics*

5.6.2 AJT Condition

To summarise the inferential statistics results, I first discuss how acceptability judgements pattern according to the focal variable of AJT Condition (ie. the different morphosyntactic variants under consideration). Across all participants (Model 1), all categories of AJT Condition are statistically significant, meaning that ratings for all of the morphosyntactic variants under investigation pattern significantly differently from each other (ie. they are indeed distinctly different variants). Additionally, across all participants (Model 1), where all six non-control conditions are significant, optional discourse-based movement conditions (right dislocation, left dislocation, and topicalisation) all have higher probabilities of obtaining positive ratings and lower probabilities of obtaining negative ratings compared to nonstandard agreement (nonstandard *was*, nonstandard *were*, and nonstandard *weren't*), confirming that optional discourse-based movement is, overall, considered comparatively more acceptable than nonstandard past-tense BE agreement. Across Polish-born participants (Model 2a), the nonstandard past-tense BE agreement conditions all have lower acceptability than the optional discourse-based movement condition of right dislocation, suggesting that Polish-born participants are sensitive to the patterns of acceptability based on type of morphosyntactic construction that was found across all participants (Model 1). With the exception of topicalisation being less acceptable than nonstandard *weren't* (which will be discussed in Section 5.5.4.2), nonstandard past-tense BE agreement also has lower acceptability than optional discourse-based movement across LGBTQ+ participants (Model 3). I now discuss the remaining results for the patterning of the AJT Condition predictor across the nested models, exploring each of the three variant functions sequentially.

5.6.2.1 Argument movement

The argument movement condition (raising-to-subject with the verb *seem*) functions as a positive control condition and also the reference category for the focal variable of AJT Condition (ie. it was used as the reference point by which acceptability of the other variables was measured), hence the findings for this variant are relatively brief: In all models, raising-to-subject has an almost 100% greater chance of acceptance compared to any of the other variants. As expected, this type of construction is considered practically universally acceptable across all segments of the participant sample.

5.6.2.2 Optional discourse-based movement

Across all participants (Model 1), right dislocation is most accepted among all the six significant non-control conditions, and, of all conditions, is comparatively most polarised towards maximal acceptability ratings. As for the other two optional discourse-based movement conditions, the most probable rating for left dislocation is the maximal response (3), though by a smaller margin compared to right dislocation. Topicalisation has the lowest rating consensus towards acceptance of the three optional discourse-based movement conditions, with both the lowest probability of a maximal rating, and slightly higher probabilities of negative ratings compared to the other optional discourse-based movement conditions.

Across Polish-born participants (Model 2a), the only significant optional discourse-based movement condition, right dislocation, is most accepted among all the non-control conditions. It has higher probabilities of obtaining positive responses, and lower probabilities of obtaining negative responses compared to the three nonstandard past-tense BE agreement conditions, and is most polarised towards the maximal acceptance response. Polish-born participants' acceptance of right dislocation mirrors the pattern across all participants (Model 1).

Across LGBTQ+ participants (Model 3), of the significant optional discourse-based movement conditions, left dislocation is polarised towards the maximal increment of the scale, having the highest probability of a maximal response and the lowest probabilities of negative responses compared across all significant conditions (ie. it is most accepted)[18]. Acceptability ratings of left dislocation pattern identically among LGBTQ+ participants compared to across all participants (in Model 1). Though its most probable response is positive, maximal responses are far less probable for topicalisation than for left dislocation, and topicalisation has slightly higher probabilities of receiving negative responses compared to left dislocation. These patterns for topicalisation are similar to those found across the whole participant sample (Model 1), with the exception that, among LGBTQ+ participants, topicalisation has a lower probability of maximal acceptance compared to other, lower, but still positive, responses. This difference causes topicalisation to be slightly less accepted than the most accepted nonstandard past-tense BE agreement condition, nonstandard *weren't*, meaning that the LGBTQ+ sample is the only one for which there is an overlap between acceptability of optional discourse-based movement versus nonstandard past-tense BE agreement (though, generally, other nonstandard past-tense BE agreement conditions are still less acceptable than optional discourse-based movement ones).

5.6.2.3 Nonstandard agreement

Across all participants (Model 1), the nonstandard past-tense BE agreement conditions nonstandard *was* and *were* are rated very similarly to each other. Both show little consensus among participants regarding acceptability, and maximal and minimal acceptability judgements are comparatively improbable. Nonstandard *weren't* patterns differently, with higher probabilities of positive ratings and lower probabilities of negative ratings.

Across Polish-born participants (Model 2a), for both nonstandard *was* and nonstandard *weren't*, there is a direct linear relationship amongst rating probabilities, with minimal acceptance being the most probable response and maximal acceptance being the least probable rating for both conditions. Polish-born participants have a higher rating consensus towards (negative) ratings of nonstandard *was* than they do when rating nonstandard *weren't*. Additionally, Polish-born

[18] Right dislocation was not statistically significant in Model 3, hence was not the most acceptable variant

participants' ratings of nonstandard *weren't* oppose those found across the broader participant sample, where ratings trend towards acceptance. Similarly, among Poles, nonstandard *was* is the least accepted variant, while, across the broader participant sample (in Model 1) there is little rating consensus, but positive ratings are generally more probable than negative ones. For nonstandard *were*, there is little rating consensus among Polish-born participants' acceptability ratings, mirroring the pattern found across all participants (in Model 1), though to an even greater degree. Although the most probable rating is positive, responses at the extremities of the scale are least probable.

Across LGBTQ+ participants (Model 3), participants have the lowest rating consensus towards nonstandard *was*, which is also the least accepted of all variants; the maximal response is least probable, and the most probable response is negative. LGBTQ+ participants pattern differently for nonstandard *was* compared to all participants in Model 1 (where, although there is similarly little rating consensus, the minimal response is least probable, and the most probable response is positive). Mirroring the results across all participants (in Model 1), LGBTQ+ participants' acceptance is more probable than lack of acceptance for nonstandard *were* and nonstandard *weren't*, with minimal responses being least probable. As in Model 1, nonstandard *weren't* has a higher probability of maximal ratings than nonstandard *were* and a lower probability of receiving a negative response than nonstandard *were*. Acceptability ratings of nonstandard *weren't* effectively pattern identically among LGBTQ+ participants compared to across all participants (in Model 1), while, for nonstandard *were*, all negative responses are less probable among LGBTQ+ participants than they are across all participants, meaning LGBTQ+ participants are more accepting of this variant.

5.6.3 Macro-social factors

Across all participants (Model 1), the macro-social factors of age band, region of residence, and socioeconomic status influence acceptability ratings. Across Polish-born participants (Model 2a), the macro-social factors of age band, region of residence, and higher-educated status influence acceptability ratings. Across LGBTQ+ participants (Model 3), the macro-social factors of age band, gender, region of residence, socioeconomic status, and higher-educated status influence LGBTQ+ participants' acceptability ratings. The results for each of the macro-social factors will be discussed in turn, and the overall results across all of these are summarised in Table 5.9.

Predictor	All participants (Model 1)	Polish-born (Model 2a)	LGBTQ+ (Model 3)
Age	Decreased acceptance of nonstandard past-tense BE agreement and topicalisation as participants age	Increased acceptance of optional discourse-based movement as Polish-born participants age	Decreased acceptance of nonstandard past-tense BE agreement, and right dislocation and topicalisation as LGBTQ+ participants age (even more so than across all participants)
Gender	No effect	No effect	Participants whose gender is beyond the binary have decreased acceptance of right dislocation & topicalisation compared to male participants
Region of Residence	North- & South-based participants show decreased acceptance compared to Midlands-based. Southerners are more accepting than Northerners of nonstandard *weren't* and right dislocation	North-based & South-based Polish participants show decreased acceptance compared to Midlands-based. Southerners more accepting than Northerners of nonstandard *weren't*	North- & South-based LGBTQ+ participants show decreased acceptance compared to Midlands-based. Southerners more accepting than Northerners of nonstandard *weren't* and nonstandard *were*
Socioeconomic Status	Increases acceptance of nonstandard *was*, nonstandard *were*, and topicalisation by those of higher socioeconomic status	No effect	No effect
Higher-educated Status	No effect	Decreased acceptance of all variants by Polish participants with higher-educated status	Decreased acceptance of left dislocation and right dislocation by LGBTQ+ participants with higher-educated status (even more so than across Polish-born participants)

Table 5.9*. Summary of inferential results across macro-social factors*

5.6.3.1 Age

Across all participants (Model 1), participants have better odds of providing higher ratings with increasing age band (ie. it has a direct main effect on AJT Response), however, age also influences how participants rate specific variants; as participant age band increases, acceptance of nonstandard *weren't*, nonstandard *was*, nonstandard *were*, and topicalisation actually *lowers*, with topicalisation being affected most strongly by this.

Across Polish-born speakers (Model 2a), as in Model 1, age band has a direct main effect on AJT Response. Amongst Polish-born participants, overall, older participants have *worse* odds of giving a higher AJT Response than younger ones, unlike the positive linear association found across all participants (in Model 1). However, age also influences how Polish-born participants rate specific variants, and this actually patterns the opposite way to the main effect, with acceptance of left dislocation, right dislocation, and topicalisation linearly *increasing* with age. Topicalisation is less strongly impacted by this than the two dislocation variants are.

Much like was found across all participants (in Model 1) and across Polish-born participants (in Model 2a), across LGBTQ+ participants (in Model 3), age band has a direct main effect on AJT Response - the odds of giving a higher AJT Response increase with age. The polarity of this association is in line with the positive association (linear increase with age)

found across all participants (in Model 1), unlike the negative association (linear decrease with age) found across Polish-born participants (in Model 2a). Age also influences how LGBTQ+ participants rate specific variants, and this actually patterns the opposite way to the main effect, with acceptance of nonstandard *weren't*, nonstandard *was,* nonstandard *were*, right dislocation and topicalisation linearly *decreasing* with age. This is opposite to the pattern found among Polish-born participants (in Model 2a) where increasing age is associated with higher acceptance of AJT Conditions. Among LGBTQ+ participants, the relative ordering of acceptability of these variants is identical to that found across all participants (in Model 1). However, the effects of age on LGBTQ+ participants are stronger than the same effects on all participants (with odds of acceptance around 82-90% lower with increasing age across the test conditions, compared to the whole participant sample, for which the odds are around 55-80% lower). As in Model 1, acceptance of nonstandard *weren't* is most negatively impacted as age increases, followed by nonstandard *was*, and nonstandard *were* is the least negatively impacted of the nonstandard past-tense BE agreement conditions. Finally, topicalisation is the variant least negatively impacted by increasing age.

5.6.3.2 Gender

Across LGBTQ+ participants (Model 3), we find an effect which is not found in other models - that of gender. LGBTQ+ women do not pattern significantly differently fromLGBTQ+ men in their AJT Responses, but participants with genders outside the binary of male and female have over 5 times higher odds of giving higher AJT Response compared to male LGBTQ+ participants (ie. gender has a main effect on AJT Response). Gender also influences how LGBTQ+ participants rate specific variants, and this seems to pattern in the opposite way to the main effect, with participants outside the gender binary being *less* accepting of right dislocation and topicalisation compared to LGBTQ+ men. Topicalisation is more impacted by this effect than left dislocation is.

5.6.3.3 Region of residence

In all models, Midlanders are more accepting than both Northerners or Southerners of all statistically significant variants. Across all participants (Model 1), nonstandard *weren't* is the variant with the lowest acceptance in the North compared to the Midlands, while acceptance of left dislocation is least negatively affected by residing in the North. Unsurprisingly, participants in the North are more accepting of nonstandard *were* than they are of nonstandard *was*. Nonstandard *weren't* is less accepted in the South than right dislocation is. Only nonstandard *weren't* and right dislocation are statistically significant predictors for AJT Response in *both* the North and the South, with both conditions having higher relative odds of acceptance in the South than in the North.

Across Polish-born speakers (Model 2a), the disparity in acceptance between the Midlands and the other two regions is larger than was found across all participants (in Model 1). Only the nonstandard past-tense BE agreement conditions of nonstandard *were* and nonstandard *weren't* pattern significantly among Poles in the North, and both have very similar odds of acceptance. Only nonstandard *weren't* is statistically significant in *both* the North and the South, and Southerners are relatively more accepting of nonstandard *weren't* than Northerners are, paralleling results from Model 1.

Across LGBTQ+ participants (Model 3), Northerners', acceptance of the significant optional discourse-based movement conditions is higher than of the nonstandard agreement ones (though topicalisation and nonstandard *was* have the same acceptability). Nonstandard *weren't* is the variant with the lowest acceptance in the North compared to the Midlands (as

also found in Model 1). Nonstandard *were* is marginally more accepted than nonstandard *weren't*, and nonstandard *was* is more accepted than nonstandard *were*. This is unlike the pattern found across all participants (in Model 1), where Northerners are more accepting of nonstandard *were* than they are of nonstandard *was*. Acceptance of topicalisation is lower than that of right dislocation, which is the most accepted variant by LGBTQ+ Northerners. Amongst LGBTQ+ Southerners, only the nonstandard past-tense BE agreement AJT Conditions of nonstandard *were* and nonstandard *weren't* pattern significantly. As in the North, nonstandard *were* is marginally more accepted than nonstandard *weren't*. Comparing between the North and the South, being a Southern resident results in a relatively smaller negative effect on acceptance of both nonstandard *were* and nonstandard *weren't* (compared to being in the Midlands) than being a Northern resident does. This suggests that LGBTQ+ Southerners are relatively more accepting of nonstandard *were* and nonstandard *weren't* than LGBTQ+ Northerners are, which matches the pattern found across all participants in Model 1.

5.6.3.4 Socioeconomic status

Generally, participants with higher socioeconomic status have better odds of providing higher ratings (ie. it has a direct main effect on AJT Response), however, socioeconomic status also influences how participants rate specific variants; nonstandard *was*, nonstandard *were*, and topicalisation are more accepted with increasing socioeconomic status. Nonstandard *was* and nonstandard *were* are affected more strongly by this than topicalisation, with acceptance of nonstandard *were* most strongly increased by higher socioeconomic status. The effect of socioeconomic status is found across LGBTQ+ participants too, but the odds of giving a higher AJT Response decrease with increasing socioeconomic status does not interact with AJT Condition.

5.6.3.5 Higher-educated status

Higher-educated status is only significant across Polish-born speakers (Model 2a), and LGBTQ+ participants (Model 3). Higher-educated Polish-born participants have better odds of giving a higher AJT Response than their non higher-educated counterparts (ie. higher-educated status has a direct main effect on AJT Response). However, higher-educated status also influences acceptability of specific variants, and this actually patterns the opposite way to the main effect, with higher-educated Polish-born participants being *less* accepting than non higher-educated Poles of all six non-control conditions. Topicalisation is most strongly impacted by this, and left dislocation, right dislocation, and nonstandard *were* are slightly less negatively impacted, with nonstandard *weren't* slightly less still. This means that nonstandard *weren't* is considered more acceptable by higher-educated participants than nonstandard *were* is. Among Poles, higher-educated status has the smallest negative effect on nonstandard *was*.

Across LGBTQ+ participants (Model 3), higher-educated status only significantly impacts left dislocation and right dislocation. Higher-educated participants are less accepting of these variants, and acceptance of right dislocation is more strongly hindered by higher education than acceptance of left dislocation. This is opposite to the pattern among Polish-born participants (in Model 2a), where acceptance of left dislocation is slightly more strongly hindered by higher education than right dislocation is. The effects of higher-educated status on LGBTQ+ participants are stronger than the same effects on Polish-born participants (with odds of acceptance around 87% lower with higher-educated status across the two test conditions significant here, compared to the equivalent results across the Polish-born participant sample, for which the odds

are around 64-67% lower). Additionally, the divergence in acceptability between left and right dislocation is smaller among LGBTQ+ participants than Polish-born participants.

5.6.4 L2 factors

The L2 factor of length of English exposure was not found to have any clear effects on acceptability judgements. However, Polish-born participants with higher English Proficiency (CEFR) were found more likely to give higher AJT responses than their lower proficiency counterparts. However, because of the very high CI for this model term, this result is unreliable. When acceptance of specific variants was examined, Polish-born participants with higher proficiency were actually found to be *less* accepting than lower proficiency participants of five of the six non-control conditions. This most strongly affects topicalisation (ie. this variant has the worst odds of acceptance as English proficiency increases). This is very closely followed by nonstandard *was*, nonstandard *weren't*, and right dislocation. Acceptance of nonstandard *were* is least negatively impacted by increasing English proficiency. Notably, these findings for English Proficiency contradict descriptive results (Section 4.4.2.3), which found a positive correlation between higher English language proficiency and higher acceptance of all three types of morphosyntactic construction.

Predictor	Polish-born (Models 2a & 2b)
English Language Exposure (years)	No clear effect
English Proficiency (CEFR)	Decreased acceptance as English proficiency increases, affecting topicalisation most and nonstandard *were* least

Table 5.10*. Summary of inferential results across L2 factors*

5.6.5 Meso-social factors

Across Polish-born participants (Model 2a) and also across LGBTQ+ participants (Model 3), higher English lifestyle (acculturation) increases participants' acceptance of right dislocation and nonstandard *was*. Among Polish-born participants, this effect is stronger for nonstandard *was* than for right dislocation. Interestingly, the odds of acceptance of right dislocation are relatively *more* increased by higher English lifestyle across LGBTQ+ participants (in Model 3) than they are across Polish-born participants (in Model 2a).

Descriptive results (Figure 4.14) suggested that those with *Low* LGBTQ+ Community Involvement are comparatively less accepting in their evaluations of nonstandard past-tense BE agreement, and suggest that those with *High* involvement are less accepting of optional discourse-based movement. Inferential results, however, do not find an effect for nonstandard past-tense BE agreement, but confirm the effect on optional discourse-based movement: Across LGBTQ+ participants (Model 3), higher LGBTQ+ Community Involvement decreases acceptance of optional discourse-based movement, with left dislocation affected most by this, followed by topicalisation, and right dislocation least negatively impacted by this effect.

Predictor	All participants (Model 1)	Polish-born (Model 2a)	LGBTQ+ (Model 3)
English Lifestyle (Acculturation)	No effect	Increased acceptance of right dislocation & nonstandard *was* from Polish-born participants with high English acculturation	N/A
LGBTQ+ Community Involvement	N/A	N/A	Decreased acceptance of optional discourse-based movement with higher LGBTQ+ Community Involvement

Table 5.11*. Summary of inferential results across meso-social factors*

5.7 Overview of inferential results

This section briefly reiterates the key findings from this chapter with regards to the research questions of this study. Addressing the first research question, regarding the overall pattern of acceptability judgements across the types of morphosyntactic constructions considered in this study, the inferential results confirm that all of the morphosyntactic variants under investigation pattern significantly differently from each other (ie. they are indeed distinctly different variants). The results confirm that the argument movement construction investigated in this study (raising-to-subject with *seem*) is considered universally acceptable across all segments of the participant sample. Across all participants, optional discourse-based movement is considered comparatively more acceptable than nonstandard past-tense BE agreement. The descriptive findings (from Chapter 4) that English-born participants are significantly more accepting than Polish-born participants was also confirmed by the statistical modelling. Another interesting observation is that, when Polish-born and English-born acceptability judgements are compared, nonstandard *weren't* diverges the most in terms of comparative acceptance between the two birth countries, while nonstandard *was*, right dislocation, and nonstandard *were* diverge comparatively less (with nonstandard *were* diverging the least). This challenges the pattern found in the descriptive results (Section 4.1.2), where comparative acceptance of nonstandard past-tense BE agreement diverges more than optional discourse-based movement conditions between Polish-born and English-born participants.

When only Polish-born participants are isolated, the nonstandard past-tense BE agreement conditions have lower acceptability than the statistically significant optional discourse-based movement condition of right dislocation, suggesting that Polish-born participants are also sensitive to these patterns of acceptability based on variant function. Among LGBTQ+ participants, nonstandard past-tense BE agreement conditions also generally have lower acceptability than the optional discourse-based movement ones, though with less divergence in acceptability ratings. The inferential results have uncovered some additional interesting patterns with regards to the relative acceptance between the individual test conditions: Across all participants, right dislocation is most accepted, and nonstandard *was* and *were* are rated very similarly to each other, whereas nonstandard *weren't* diverges, with far higher acceptability. While LGBTQ+ participants' ratings of nonstandard *weren't* mirror those of all participants, Polish-born participants' ratings oppose those found across the broader participant sample, and trend towards lack of acceptance. Similarly, when isolating only Polish-born participants, nonstandard *was* is the least accepted significant variant, while, across the broader participant sample, it trends towards being accepted. For

nonstandard *were*, there is little rating consensus among Polish-born participants' acceptability ratings, mirroring the pattern found across all participants, though to an even greater degree.

Addressing the second research question regarding the effect of L2 factors on Poles' acceptability ratings, inferential findings suggest that Polish-born participants with higher English language proficiency are less accepting towards all significant test conditions (both optional discourse-based movement constructions and nonstandard past-tense BE agreement constructions), meaning that English language proficiency has an effect on acceptance level.

Addressing the third research question regarding the effects of macro-social factors, participants' acceptance of significant variants generally decreases with age across all participants (with the biggest effects on acceptance of nonstandard past-tense BE agreement constructions). However, acceptance of optional discourse-based movement *increases* with age among Polish-born participants. LGBTQ+ participants whose gender is not binary are less accepting of significant optional discourse-based movement conditions. Higher-educated Polish-born participants have lower acceptance of all variants than non higher-educated ones. Region is also significant: Midlanders are more accepting than both Northerners or Southerners across all divisions of the participant sample. Southerners have higher relative acceptance than Northerners, especially for nonstandard *weren't*, which, across all participants, and also among the Polish-born as well as LGBTQ+ samples, is more accepted by Southerners than their Northern counterparts. Across all participants, acceptance of significant variants increases with socioeconomic status. Finally, I address the fourth and fifth research questions regarding the effects of meso-social factors (those related to community embeddedness) on acceptance. Among Polish-born participants, acceptance of right dislocation and nonstandard *was* increases with English acculturation level. Among LGBTQ+ participants, acceptance of optional discourse-based movement constructions decreases with LGBTQ+ Community Involvement level.

For ease of reference, Table 5.12 provides a summary of the key patterns on the acceptance of the morphosyntactic variants according to each of the factors covered in this chapter. In the next chapter, the implications of these findings are discussed.

Type of factor	*Predictor*	**Optional discourse-based movement**			**Nonstandard agreement**		
		Right dislocation	Left dislocation	Topicalisation	Nonstandard *was*	Nonstandard *were*	Nonstandard *weren't*
Key sample characteristic	Birth country *(Polish-born vs English-born)*	Decreases acceptance among all participants	-	-	Decreases acceptance among all participants	Decreases acceptance among all participants	Decreases acceptance among all participants & LGBTQ+
	LGBTQ+ status *(yes vs no)*	Increases acceptance among all participants & Polish-born	Increases acceptance among all participants & Polish-born	-	Increases acceptance among all participants & Polish-born	Increases acceptance among all participants & Polish-born	Increases acceptance among all participants & Polish-born
	Age	Decreases acceptance among LGBTQ+; Increases acceptance among Polish-born	Increases acceptance among Polish-born	Decreases acceptance among all participants & LGBTQ+; Increases acceptance among Polish-born	Decreases acceptance among all participants & LGBTQ+	Decreases acceptance among all participants & LGBTQ+	Decreases acceptance among all participants & LGBTQ+
	Gender *(Not binary vs Male)*	Decreases acceptance among LGBTQ+	-	Decreases acceptance among LGBTQ+	-	-	-
Macro-social (demographic)	Region of Residence *(North vs Midlands)*	Decreases acceptance among LGBTQ+ participants	Decreases acceptance among all participants	Decreases acceptance among LGBTQ+ participants	Decreases acceptance among all participants, & LGBTQ+	Decreases acceptance among all participants, & Polish-born, & LGBTQ+	Decreases acceptance among all participants, Polish-born, & LGBTQ+
	Region of residence *(South vs Midlands)*	Decreases acceptance among all participants	-	-	-	Decreases acceptance among LGBTQ+	Decreases acceptance among all participants, Polish-born, & LGBTQ+
	Socioeconomic Status	-	-	Increases acceptance among all participants	Increases acceptance among all participants	Increases acceptance among all participants	-
	Higher-educated Status *(yes vs no)*	Decreases acceptance among Polish-born + LGBTQ+	Decreases acceptance among Polish-born & LGBTQ+	Decreases acceptance among Polish-born	Decreases acceptance among Polish-born	Decreases acceptance among Polish-born	Decreases acceptance among Polish-born

L2 (second language)	English Language Exposure (years)	-	-	-	-	-	-
	English Proficiency (CEFR)	Decreases acceptance	-	Decreases acceptance	Decreases acceptance	Decreases acceptance	Decreases acceptance
Meso-social (community embeddedness)	English Lifestyle (Acculturation)	Increases acceptance among Polish-born	-	-	Increases acceptance among Polish-born	-	-
	LGBTQ+ Community Involvement	Decreases acceptance	Decreases acceptance	Decreases acceptance	-	-	-

Table 5.12. *Summary of the effects of all factors on acceptance of optional discourse-based movement and nonstandard agreement*

6. Discussion

6.1 Introduction

In Chapter 4, I presented and discussed the descriptive results of this study, assessing how acceptability ratings pattern across the morphosyntactic features of interest. These were considered according to the types of morphosyntactic construction: (i) argument movement (raising-to-subject with *seem*); (ii) optional discourse-based movement (left dislocation, right dislocation, topicalisation); nonstandard agreement (past-tense BE: nonstandard *was*, nonstandard *were*, and nonstandard *weren't*). The effects of individual factors on participants' judgements of these features were also explored, and split into five broad types: (i) the key sample characteristics of birth country and LGBTQ+ status; (ii) the macro-social (demographic) factors of interest, namely age, gender, region, and the socioeconomic factors of socioeconomic status and education status; (iii) the L2 (second language) factors influencing Polish-born participants, namely their age of arrival to England and their English language proficiency, and; (iv) the meso-social factors considered in this study (ie. those related to community embeddedness) - participants' levels of English and Polish acculturation, as well as their levels of LGBTQ+ Community Involvement.

In Chapter 5, the statistical findings were explored via four proportional odds (ordinal logistic regression) models, taking a nested approach in order to segment the sample according to the key characteristics under investigation (birth country and LGBTQ+ status):

- **Model 1** (Section 5.2) incorporates all participants (both English-born and Polish-born, and both LGBTQ+ and non-LGBTQ+)
- **Model 2a** (Section 5.3) incorporates only Polish-born participants (both LGBTQ+ and non-LGBTQ+)
- **Model 2b** (Section 5.4) incorporates only Polish-born participants who completed the Versant English Speaking test of linguistic proficiency
- **Model 3** (Section 5.5) incorporates only LGBTQ+ participants (both English-born and Polish-born)

The inferential statistical analysis informed our understanding of the patterns of acceptability across the individual morphosyntactic features, and the types of morphosyntactic construction considered in this thesis. Influential factors were considered sequentially, according to the same five broad types as for the descriptive results (Chapter 4), In doing so, it was possible to compare results across different segments of the participant sample and explore how different factors influence participants' acceptability judgements. More specifically, it made it possible to compare the differences in effects found across Polish-born participants and LGBTQ+ participants with those found across all participants.

In this chapter, I discuss the findings from the descriptive and inferential analysis. The first point of discussion will be the patterns of acceptability across the three types of morphosyntactic construction. I will then discuss how non-L1 Polish-born migrants integrate into these patterns of variation, in the English cultural context, and how their patterns of acceptance of the different types of morphosyntactic construction might relate to these participants' lived experiences and motivations. As part of this, I will explore Polish-born migrants' potential acquisition of indexical associations, as well as meso-level factors which might be inhibiting their acquisition of nonstandard morphosyntactic variation. Next, I consider the

meso-social factors of cultural embeddedness (acculturation), and how this interacts with Polish migrants' acquisition of sociolinguistic variation. I then explore patterns of acceptance of morphosyntactic variation, as well as divergence from sociolinguistic norms, with a focus on LGBTQ+ membership and embeddedness in the LGBTQ+ community. Finally, I explore how existing at the confluence of the two communities under investigation in the present study - ie. being a Polish-born LGBTQ+ migrant - impacts morphosyntactic variation.

6.2 Acceptance by type of morphosyntactic construction

To address the first research question proposed in the present study, the results in Chapters 5 confirm that the rating patterns across the morphosyntactic features are different to a statistically significant degree - ie. they are indeed distinctly different types of morphosyntactic constructions. Furthermore, Chapter 4 suggested a clear pattern in the relative acceptability of the three types of morphosyntactic construction proposed in this thesis - from most to least acceptable: argument movement; optional discourse-based movement; nonstandard agreement. This pattern is also backed up by the inferential results, which indicate that the argument movement construction (raising-to-subject with the verb *seem*) is considered universally acceptable across all segments of the participant sample and, generally speaking, optional discourse-based movement constructions are considered comparatively more acceptable than nonstandard past-tense BE agreement ones. Furthermore, the results indicate that it is not only L1 speakers of British English that are sensitive to the patterns of acceptability based on type of morphosyntactic construction: across Polish-born participants, nonstandard past-tense BE agreement constructions have lower acceptability than the statistically significant optional discourse-based movement construction of right dislocation (Table 5.3). Nonetheless, country of birth plays a role in *how* accepting participants are, with English-born participants more accepting of all statistically significant morphosyntactic variants than Polish-born participants. Additionally, the descriptive results (eg. Figure 4.2) also suggest that the relative disparity in acceptance between participants from the two birth countries is wider (ie. the difference in acceptability is higher) for nonstandard agreement than for optional discourse-based movement, which, in turn, is higher than for argument movement. To understand these patterns better, each type of morphosyntactic construction will be discussed in turn.

6.2.1 Argument movement

The argument movement construction (raising-to-subject with the verb *seem*) is widely accepted among all participants. The verb *seem* does not select for a subject and must therefore undergo movement in order to fulfil requirements such as case assignment (explained in more detail in Section 2.6.1). This underlying morphosyntactic process is not something that is consciously known by speakers of British-English, yet it is a fundamental part of the linguistic system for a British-English speaker, and acquired across all varieties. Hence, this construction was used as the positive control condition in this study, and results confirm that this construction is also widely accepted among Polish-born participants. The frequency of raising structures in English is cross-linguistically unusual (Givón, 2001), and the patterning of verbs like *seem* is not explicitly taught to acquirers of English. It is well documented that such raising constructions are comparatively difficult for adult L2 learners of English to acquire (Callies, 2008: 201). Callies found Polish-born L2 English speakers struggled with target-like use of raising constructions which has motivated this construction's use as a proxy for

benchmarking advanced learners' acquisition of English in the present study. Therefore, Polish-born participants' unanimous acceptance of this construction confirms that these participants are advanced speakers of English.

6.2.2 Optional discourse-based movement

Inferential results confirm that, across all participants, optional discourse-based movement is less accepted than the positive control argument movement construction, but more accepted than nonstandard agreement (Table 5.1). The relative acceptability of optional discourse-based movement is not surprising given that these types of variants exist at the intersection of grammatical and social function (Moore, 2020). Certain optional discourse-based movement constructions *can* be subject to existing socioindexical relations within the community, such as right dislocation with pronoun tags, and its association with Northern English social identities (eg. Moore & Snell, 2011). Overall, however, the discourse-based movement constructions investigated in this study are relatively widespread across British-English varieties and the communities among which these varieties are spoken. For instance, any speaker who interacts with British-English speakers will almost certainly at some point encounter left dislocated constructions, regardless of the variety of English being spoken. It is probable that this likelihood of encountering these variants across varieties of British-English bolsters their rates of acceptance, especially as most optional discourse-based movement constructions investigated here are widespread in the standard variety. This means their acceptance is less hindered by prescriptive judgments. The usage of such variants with high discourse-pragmatic utility hinges on interpersonal motivations in communication. For instance, variation within the left periphery (such as left dislocation) is associated with providing illocutionary force - ie. the speaker's intention (Cornips & Corrigan, 2005: 22). Other discourse management functions include emphasis, clarification, or focus (Snell, 2018: 10). These are functions of use to all speakers when interacting with others, so it is not surprising that the relative acceptance of these discourse-based movement constructions (which facilitate these discourse purposes) is relatively high.

As already mentioned, Polish-born participants are less accepting of the non-control variants than English-born participants. The fact that acceptance of optional discourse-based movement constructions differs by birth country is not surprising; Because optional discourse-based movement constructions function to facilitate the interactional flow or assert interspeaker positionality, it follows that the use of these features is predicated on an existing familiarity with interactional norms and behaviours of the cultural context in which they are used (in this case, British English culture). Because the usage of these features hinges on interpersonal motivations in communication, to acquire and understand the patterning of these features requires cultural engagement with the British English speech community. Descriptive results (Table 4.14) show that, although the Polish-born participant sample does generally trend towards *High* English Lifestyle (acculturation), it is far less skewed towards *High* compared to the benchmark group of English-born participants. This difference in cultural embeddedness may explain the relative disparity in acceptance of optional discourse-based movement by birth country. Nonetheless, since the Polish-born participants live and work in England, they interact with British English speakers, so will have these optional discourse-based movement constructions in their input, at least to some extent, unlike the comparatively more restricted nonstandard past-tense BE agreement. Because they are more likely to encounter these features than nonstandard past-tense BE agreement, and the optional discourse-based movement constructions are generally less prescriptively enforced, it is not surprising that Poles have a higher acceptance of optional discourse-based movement compared to nonstandard past-tense BE agreement.

6.2.3 Nonstandard agreement

In Chapter 2, I proposed that the acceptance of nonstandard past-tense BE agreement is likely to be the most restricted of the three types of morphosyntactic construction investigated in this thesis. This is confirmed by the relative low acceptability, across all participants, of nonstandard past-tense BE agreement in both descriptive (Figure 4.1) and inferential (Table 5.1) results. In addition to performing referential and discourse-pragmatic functions, variation in past-tense BE agreement is also socially meaningful within specific communities, and these nonstandard variants are indexical of particular social identities associated with these communities. This means that, in order to acquire these features, speakers must have cultural engagement with the broad British English speech community (as for the optional discourse-based movement constructions), *and* some extent of embeddedness within, or at least awareness of, the British English speech communities and associated identities for which the variants in question are socially meaningful. Variants which are associated with specific (non-hegemonic) communities and identities often violate prescriptive norms and can be met with pejorative judgments. This explains why, even across all participants (of which approximately half are L1 British English speakers), nonstandard past-tense BE agreement has low acceptability compared to the more widespread optional discourse-based movement constructions.

This pattern of more widespread variants having higher acceptability is further supported by examining how acceptability ratings pattern for nonstandard past-tense BE agreement. As discussed in Chapter 2, nonstandard *weren't* is subject to two paradigms of variation (*was/weren't*, and levelled *were* - see Section 2.6.3), and is the nonstandard past-tense BE agreement construction that is most widespread across British Englishes (eg. Schilling-Estes & Wolfram, 1994; Britain, 2002), therefore, having the highest opportunity for acquisition. Comparing across all participants, inferential results (Table 5.1) confirm that nonstandard *weren't* is the most accepted of the nonstandard agreement constructions, albeit marginally.

It is also worth noting that, generally speaking, inferential results show that Midland-based participants are consistently more accepting of all significant variants compared to both Northern- or Southern-based participants. A tentative suggestion as to why this pattern is found could perhaps be due to these more centrally-located participants being more likely to encounter a range of different varieties and linguistic features. In particular, of all significant variables, acceptance of nonstandard *weren't* in the North and the South diverges most strongly from the Midlands (Table 5.2). The nonstandard past-tense BE agreement constructions are all regionally distributed, to different extents, and participants living in the Midlands are situated at the cross-section of these patterns of variation. This supports the view that proximity (cf. Montgomery, 2012) and, therefore, exposure, has a positive effect on acceptability - ie. variants that are more widespread across British English speech communities are also more likely to be accepted. Because the use of nonstandard past-tense BE agreement constructions hinges on speakers being able to interpret these variants' indexical associations, it is not surprising that, in order to accept these variants, speakers must have access to the communities within which the variant in question is both socially meaningful and productive in everyday discourse. This latter point is particularly crucial, as this is what is most likely to condition patterns of positive acceptance of a variant. Mere awareness of a variant could include, for instance, abstract stereotypical perceptions of variants based on their associations with specific 'other' social identities, rather than being reflective of one's own identities, values, or attitudes.

6.3 Acceptability in Poles

In the previous section, I explored patterns of acceptability according to the type of morphosyntactic construction, and across each of the morphosyntactic features considered in this thesis. The following section will discuss results with regards to the second part of the first research question proposed in the present study, regarding how the non-L1 Polish-born migrants integrate into these patterns of variation in the English cultural context. To explore these results, I will discuss how these participants' patterns of acceptance of the different types of morphosyntactic constructions might relate to these participants' lived experiences and motivations. In doing so, I also draw upon the second research question, regarding the effects of these participants' English proficiency, and on the third research question, regarding the effects of certain macro-social factors on Polish-born participants' acceptance.

6.3.1 Acquiring Indexical Associations

Inferential results confirm that Polish-born migrants do align with some broad patterns found across all participants. For instance, they are sensitive to the patterns of acceptability based on type of morphosyntactic construction: among Polish-born participants, nonstandard past-tense BE agreement has lower acceptability than the optional discourse-based movement construction of right dislocation (Table 5.3). Another pattern in common is that Midland-based Poles consistently have higher acceptance of the variants under investigation than their Northern or Southern-based counterparts. This means that Polish-born migrants are also, to some extent, sensitive to the proximity (and, therefore, exposure) effects resulting from being situated at the cross-section of several distributions of nonstandard past-tense BE agreement. Interestingly, this disparity in acceptance between the Midlands and the other two regions is larger among Polish-born participants (ranging from around 80%-90% more acceptance in the Midlands, depending on the variant), compared to the equivalent pattern across all participants (where there is only around 55%-70% more acceptance in the Midlands). As argued in Section 2.3.3, at a macro-level, the notion of 'region' can be theorised not as a fixed concept, but a system arising as a result of the 'culturally constructed valances' (Ochs 1996: 417) reflective of the processes occurring within the social network structures of the people in those areas. In this view, participants in the Midlands, have more direct proximity than those in the North and the South to multiple intersecting areas where instantiated uses of the *supralocal* (Britain, 2010) variants under investigation here can occur. These instantiated uses both generate (Ochs, 1996: 417) and draw upon indexical associations between the variants and their ideological associations. Therefore, although speculative, it is worth noting that participants situated in the Midlands are more likely to have internalised these indexical associations through exposure, and therefore are more accepting of these variants. The fact that Polish-born migrants also replicate this pattern is significant because these indexical associations are below a level of conscious awareness, so this means they have, at least to some extent, internalised these features' social meanings. Conversely, Poles in the North and South are not situated in the same sort of confluence point. As descriptive analysis (Table 4.6) shows, 73.3% Polish-born participants have *not* lived outside of their current region of residence, while 71.25% English-born participants *have*, and this reduced mobility among the Polish-born sample may help explain the relatively higher disparity in acceptance between Midland-based and Northern-based or Southern-based Poles.

6.3.2 Prescriptive Ideology & Hegemonic Norms

Polish-born migrants also *diverge* from some of the patterns of acceptability found across all participants. Inferential results confirm that Polish-born participants are less accepting of all types of morphosyntactic features investigated, compared to English-born participants. Given that participants from both birth countries rate the argument movement construction similarly - ie. The Polish-born participants have acquired fundamental underlying morphosyntactic processes, confirming that they are proficient users of English - this suggests that something about the properties associated with these variants, to an extent, inhibits non-L1 participants' acceptance.

Based on the findings from this study, Polish-born migrants' acceptability ratings are influenced by their lived experiences within educational settings, experiences of formal English instruction, and the prescriptive ideologies associated with these. In this section, I mainly explore the effects of the macro-level factor of higher-educated status, and the impact of the L2 factor of English linguistic proficiency. Factors concerning Polish-born participants' education background and linguistic proficiency are associated with decreased acceptance of the morphosyntactic variants under investigation. Higher-educated status has a significant effect on Polish-born participants' acceptance across all six non-control morphosyntactic features (but does not have a significant effect across all participants), with higher-educated Poles having lower acceptance of all variants than non higher-educated Poles (Table 5.4). Most (around 80%) Polish-born participants are higher-educated (Table 4.8) and, in the descriptive findings, over 40% of Polish-born participants listed study or education as a key motivation for their migration to England (Table 4.19). Nonstandard morphosyntactic variation has been the subject of much overt proscription from "standard ideology" (Eckert, 2019: 758) in the educational setting (Moore & Spencer, 2021: 2). Only 8% of Poles have not had any formal English instruction (Table 4.10), meaning the vast majority of participants have had some level of exposure to prescriptive English grammar norms, particularly if they began learning English formally at school in Poland (of whom 85% did), before they had lived experience of the English cultural context. Therefore, not only have these migrants been exposed to general metalinguistic judgements and stigmatisation of these nonstandard variants that L1 speakers are, but have likely been overtly proscribed the "correct" forms to use during their formal English instruction. From the inferential results, it is clear that English language proficiency has an effect on acceptance ratings among Polish-born participants, but the finding is surprising in that higher proficiency Poles are *less* accepting than those with lower proficiency for all significant variants, both optional discourse-based movement and nonstandard past-tense BE agreement. This goes against descriptive results (Figure 4.11) which found a positive effect of English proficiency on acceptance. Since understanding the social mechanisms of variation is a key component of being able to use the language in real-world social contexts (Bayley & Regan, 2004: 325), this finding supports the notion that acceptance of native-like morphosyntactic variation does not hinge on attaining maximum linguistic proficiency. A potential co-linear variable resulting in the negative effect for proficiency could be education-related; participants who are more proficient may have also acquired aspects of standard language ideology, perhaps linked to the high levels of exposure to higher education among the participant sample (which over 80% of Polish-born participants have accessed - Table 4.8). Polish-born participants are also shown to reject the pattern found across all participants with regards to the spread of nonstandard past-tense BE agreement; comparing the acceptance of nonstandard past-tense BE agreement across all participants by birth country (Table 5.2), the relative acceptance of nonstandard *weren't* (of the four statistically significant variants) is most negatively impacted by Polish birth country, while, among all participants (Table 5.1), it is the most accepted. Since nonstandard *weren't* is the most

geographically widespread of the nonstandard past-tense BE agreement constructions, it seems that, across Polish-born participants, higher proximity to a variant *reduces* its chance of acceptance. Other than the factors I have already discussed in this section, it is possible that this, and the negative effect of proficiency, result from a desire to orient to hegemonic norms and overt prestige. In order to explore migrants' motivations further, I now turn to meso-social factors.

6.4 The Meso-level

Because the participant sample is diverse in its makeup (ie. the Polish-born participants also include LGBTQ+ participants), I will now explore the fourth and fifth research questions from the present study, examining the effects of the meso-social factors of cultural embeddedness (acculturation) for Polish-born migrants, and LGBTQ+ embeddedness for the LGBTQ+ participant sample. This will allow us to probe deeper into the relationship between community membership, and the internalisation of socioindexical variation.

6.4.1 Cultural Embeddedness

The meso-social factor of acculturation reveals interesting findings about the ways in which migrants interact with the different types of morphosyntactic construction. Across Polish-born participants, inferential results (Table 5.4) show that acceptance rating increases with acculturation level (ie. that *degree* of embeddedness in English cultural practices impacts acceptance of morphosyntactic variants), confirming the presence of an acculturation effect on at least two variants. This supports prior findings that migrants' attitudes towards integrating into the target culture's speech community positively affects their acquisition (eg. Jiang et al., 2009: 481; Masgoret & Gardener, 1999) and sociolinguistic competence (Schleef et al., 2011). Crucially, this finding tells us about how migrants' identities influence their internalisation of morphosyntactic variation; a migrant having a higher level of English acculturation is a consequence of an internal process, whereby they are able to situate themself within the wider English population - ie. It is a proxy measure representing abstract information about migrant identities, not simply the activities they do in their day-to-day lives. Acculturation is an ongoing process throughout a migrant's lifetime (eg. Golden, 1988), therefore, it is worth noting that Polish-born participants' acceptance of optional discourse-based movement also increases with age. As this pattern opposes that found across all participants, and across LGBTQ+ participants, where age has a negative effect on acceptability. This age effect among Poles could potentially be a side effect of the acculturation process, since acquisition of a language occurs in tandem with that of the associated culture, and establishing a deeper internalisation of a culture takes time.

The nonstandard past-tense BE agreement construction of nonstandard *was*, which is normally the least accepted variant among Poles, undergoes the strongest increase in acceptability with increasing English acculturation level, whereas the optional discourse-based movement construction of right dislocation, which is normally the most accepted, undergoes a weaker positive effect. This supports the pattern found in the descriptive results (Figure 4.12) that, among Polish migrants, nonstandard past-tense BE agreement undergoes a stronger cumulative increase in acceptability with each increment of acculturation level than optional discourse-based movement. This is not surprising given that optional discourse-based movement constructions are more widely available and acquired by a broader range of people and communities, while nonstandard past-tense BE agreement constructions are more restricted in their usage. Therefore, we would expect their

acceptability to pattern more strongly with acculturation than that of optional discourse-based movement, as the acceptance of nonstandard past-tense BE agreement hinges on a deeper level of involvement with the communities within which these nonstandard variants are socially meaningful. Given that both the formation of socioindexical associations and becoming acculturated are both predicated on processes of internalisation and embeddedness, it is not surprising that one process might bolster the other.

6.4.2 LGBTQ+ Embeddedness

Speakers' identities (and their values and attitudes) reflect in their use of language variation, therefore, it is of interest to variationists to understand how different types of communities fit into our paradigms of sociolinguistic variation. I will now explore patterns of acceptance of morphosyntactic variation according to LGBTQ+ membership and embeddedness. Although I specifically examine the LGBTQ+ community here, this is as a case study for exploring how interconnected indexical links may pattern across non-hegemonic communities based around a shared set of particular values. The investigation of this community affords us a more nuanced understanding of the social diversity across the participant sample and reflects a key element of the identities of many of the participants represented here. It also allows us to compare how membership to such a community, and the influence of its associated community norms, impacts participants' acceptance of the different morphosyntactic features under investigation in this study, with a focus on how this patterns according to the type of morphosyntactic construction (ie. argument movement; optional discourse-based movement; and nonstandard agreement). As part of this, I will also discuss results about LGBTQ+ status, considering the influences of LGBTQ+ community membership, in order to nuance what we can learn from participants' depth of embeddedness within this community.

6.4.2.1 Acceptance of nonstandard variation

In some ways, LGBTQ+ participants' acceptance behaviours pattern similarly to the whole participant sample. For instance, inferential results (Table 5.6) confirm that, among LGBTQ+ participants, nonstandard past-tense BE agreement generally has lower acceptability than optional discourse-based movement, suggesting that LGBTQ+ participants are sensitive to the patterns in acceptability according to these different types of morphosyntactic construction. Additionally, Midland-based LGBTQ+ participants consistently have higher acceptance of the variants under investigation than their Northern or Southern-based counterparts, meaning LGBTQ+ participants are sensitive to the proximity (and, therefore, exposure) effects resulting from being situated at the cross-section of several distributions of nonstandard past-tense BE agreement. LGBTQ+ participants also diverge from Poles in some ways, behaving more similarly to the whole participant sample. For instance, they are generally accepting of nonstandard *weren't* (Table 5.7), while Poles are generally not.

In fact, inferential results (Table 5.2) confirm that LGBTQ+ participants are around *twice* as accepting of **all** statistically significant variants - nonstandard *was*, nonstandard *were*, nonstandard *weren't*, left dislocation, and right dislocation - compared to their non-LGBTQ+ counterparts. Membership to the LGBTQ+ community is associated with specific practices, values, and a sense of shared solidarity - aspects which I have argued throughout this thesis serve to reinforce aspects of identity. Shared values (such as solidarity; Snell, 2018) can provide motivation to bolster the use of nonstandard variation. Communities that face marginalisation and pressure from hegemonic power structures (such as the

LGBTQ+ community) can become resistant to these, instead orienting *towards* the internal values of the community (Milroy, 1980: 60-61 discusses a similar phenomenon occurring in local communities). The LGBTQ+ community is predicated on values such as inclusivity, nuance over rigid categorisation (ie. right vs wrong), and blurring the lines of hegemonic norms. Furthermore, the community is a highly non-homogenous melting pot, meaning members will likely have experience with people of different backgrounds, potentially from different countries or regions of England. Therefore, the results found here suggest that the shared values of acceptance and open-mindedness prevalent within the LGBTQ+ community are also instilled in participants' acceptance of nonstandard variation.

As well as LGBTQ+ participants generally being more accepting of the morphosyntactic features investigated in this study, it is important to explore the *degree* to which being in the LGBTQ+ community bolsters acceptance of these features, with regards to the different types of morphosyntactic construction under consideration in this thesis. Across all participants (Table 5.2), acceptance of left dislocation is *least* boosted (of the five statistically significant variants) by being LGBTQ+ compared to *not* being LGBTQ+, while right dislocation is *most* boosted. While both left dislocation and right dislocation have discourse-pragmatic functionality, right dislocated structures have been found to be inextricably linked to the *practices* enacted by groups (Moore, 2003; 2020), and may have socioindexical links with macro-social identity factors such as gender and age (Durham, 2011). It is possible that LGBTQ+ participants are orienting to the socioindexical properties of right dislocation, especially given that test items for this condition included right dislocated pronoun tags, which have been argued to carry socioindexical links connected to shared values of solidarity (eg. Snell, 2018).

To probe deeper into this, I will now explore what more can be learned from the meso-level (ie. community-embeddedness) factor of LGBTQ+ Community Involvement.

6.4.2.2 Divergence from sociolinguistic norms

LGBTQ+ participants are the only segment of the participant sample in which there is an (albeit marginal) overlap in the relative acceptance of optional discourse-based movement and nonstandard past-tense BE agreement; LGBTQ+ participants accept topicalisation (a form of discourse-based movement) slightly *less* readily than nonstandard *weren't* (Figure 5.3). As mentioned in Section 6.4.2.1, nonstandard *weren't* effectively patterns identically among LGBTQ+ participants compared to across all participants, therefore topicalisation is the variant which patterns differently here. Before I examine what this effect might mean, it is worth noting that this finding reinforces the fact that, although the three type of morphosyntactic construction (argument movement; optional discourse-based movement; and nonstandard agreement) generally pattern in a cohesive way, the relative acceptability of morphosyntactic variants can be somewhat fluid, varying, for instance, according to the norms and practices among certain communities. Given this, I now turn to the acceptability of optional discourse-based movement among LGBTQ+ participants.

Exploring LGBTQ+ participants' patterns of acceptability at the meso-social level (ie. according to their level of community-embeddedness), it is clear that optional discourse-based movement patterns differently among LGBTQ+ participants, compared to across all participants. Inferential results (Table 5.7) confirm that participants with higher levels of LGBTQ+ Community Involvement are *less* accepting of optional discourse-based movement compared to those with lower involvement. A further interesting dimension to this pattern is that the degree of LGBTQ+ Community Involvement seems to affect more socially meaningful optional discourse-based movement *less*: The acceptance of left dislocation is most strongly

impacted by higher LGBTQ+ Community Involvement, while right dislocation is *least* negatively impacted by this effect. This parallels the pattern explored in Section 6.4.2.1 where, across all participants, the acceptance of left dislocation is *least* boosted by being LGBTQ+ compared to not being LGBTQ+, while acceptance of right dislocation is *most* boosted.

I first explore the general pattern that more highly involved LGBTQ+ community members are less accepting of the optional discourse-based movement constructions under investigation. Because the usage of constructions with such discourse-pragmatic utility hinges on participants' interpersonal motivations in communication, the acceptance of these features could be tied to individuals' cultural engagement with the British English speech community, and participation in similar community and communicative norms. A potential explanation for this pattern across LGBTQ+ members, therefore, is that the lived experiences of members of this community, especially ones more deeply embedded in the community, mean they orient counter-culturally in some way, or otherwise in opposition to hegemonic norms. Another suggestion is that this may be tied to the types of discourses and practices prevalent within the community, as certain optional discourse-based movement constructions can perform socioindexical functions due to speakers making greater use of particular pragmatic functions, by virtue of the dominant discourses within the community. This builds on previous work which distinguishes between how different types of morphosyntactic constructions accrue social meaning. For instance, Moore (forthcoming) argues that the social meanings of certain morphosyntactic constructions, such as the optional discourse-based movement construction of right dislocation, are mitigated by their discourse-pragmatic utility (in this case, providing end-focus and emphasis), which then provides the potential for social meanings to develop. If, hypothetically, members of a certain community (the LGBTQ+ community, in this case) happen to perform these discourse-pragmatic actions in a different way, or use different constructions to achieve similar aims, we might then see this reflected in these communities' perceptions of these forms.

The fact that the relative acceptability of optional discourse-based movement differs according to socioindexical properties of the variant, and that this patterns with level of embeddedness with the community suggests that members of this community do, in fact, orient to the socioindexical properties of right dislocation. The pattern found here further bolsters the suggestion raised in Section 6.4.2.1 that members of the LGBTQ+ community may be orienting towards the internal values of the community (potentially, though not necessarily, solidarity), as those who are more embedded in this community show this pattern more strongly.

This view is further supported by the fact that, in the present study, the acceptability of optional discourse-based movement among LGBTQ+ participants is found to pattern with the macro-social factors of gender and age, which have been considered as potential constraints on right dislocation (Durham, 2011). Participants whose gender is beyond the binary of male and female also undergo this effect, having lower acceptance of right dislocation & topicalisation compared to male LGBTQ+ participants (though, with right dislocation less affected by this than topicalisation). Speakers with non-normative genders have been argued to present "challenges to the linguistic status quo" (Zimman, 2016, referring to transgender people's use of pronouns), so it would follow that these speakers may desire to orient against dominant cultural norms. Although it is difficult to say anything definitively about this effect, it is interesting, but not surprising, that gender has an effect among LGBTQ+ participants, given that gender has often been implicated in analyses of discourse, especially within the LGBTQ+ community (eg. Jones, 2022). The second macro-social factor of age is also interesting, as older LGBTQ+ participants are less accepting of all statistically significant variants, and the effect of age on LGBTQ+ people is stronger

than the effect of age is across all participants. It is known, for instance, that members of transgender youth groups discursively construct their identities and stances of opposition towards normative power structures (Jones, 2022), therefore, it is not surprising that these macro-social factors can be seen to have an effect on participants in this study.

Finally, the cumulative effect found of LGBTQ+ Community Involvement on the relative acceptance of right dislocation could be similar to the effect seen for acculturation among the Polish migrants (Section 6.4.1), where nonstandard past-tense BE agreement undergoes a stronger cumulative increase in acceptability with each increment of acculturation level than optional discourse-based movement. I theorised that this could be because both the formation of socioindexical associations and becoming acculturated are both predicated on processes of internalisation and embeddedness, and LGBTQ+ Community Involvement follows a similar process. I tentatively propose that, instead of affecting nonstandard past-tense BE agreement (perhaps because these socioindexical links are less meaningful to LGBTQ+ participants, or because participants are from different birth countries, regions, and differing levels of acculturation), this pattern is perhaps due to an association between the formation of socioindexical links (with the optional discourse-based movement construction of right dislocation) and of becoming more embedded into a particular community.

As already mentioned, the participant groups discussed thus far are from different cross-sections of the communities under analysis, and thus pattern very differently. This is confirmed in the patterns of acculturation captured in the descriptive analysis: (Table 4.14) shows that a lower percentage of English-born LGBTQ+ participants have *High* English acculturation than their non-LGBTQ+ counterparts, while a **higher** percentage of Polish-born LGBTQ+ participants have *High* English acculturation compared to their non-LGBTQ+ counterparts. In order to understand how existing at the confluence of the two communities under investigation here impacts participants' morphosyntactic variation, I now focus specifically on Polish-born LGBTQ+ migrants.

6.4.2.3 At the cross-section

There is not a statistical model specific to the Polish-born LGBTQ+ sample, meaning it is not possible to directly discuss their levels of LGBTQ+ Community Involvement. However, it is possible to infer how these participants' acceptance patterns by examining the effects of LGBTQ+ status amongst Polish-born participants. From the descriptive findings (Figure 4.3), it seemed that LGBTQ+ status does not affect the acceptability of optional discourse-based movement - only nonstandard past-tense BE agreement, with Polish-born LGBTQ+ participants more accepting of these variants compared to Polish-born non-LGBTQ+ participants. However, inferential results confirm that being LGBTQ+ positively affects Polish-born participants' acceptance of all morphosyntactic variants. Given what we know about the Polish LGBTQ+ cultural context, this is not surprising. Although Polish-born participants are generally not very Polish-acculturated, LGBTQ+ Poles are even less so; A lower percentage of Polish-born LGBTQ+ participants have *High* Polish acculturation than their non-LGBTQ+ counterparts, and a far higher percentage of Polish-born LGBTQ+ participants have *Low* Polish acculturation than their non-LGBTQ+ counterparts. Descriptive patterns show that Polish-born LGBTQ+ migrants feel more likely to be open about their LGBTQ+ identity with English people than with other Poles, and more unlikely to be open with Polish strangers than friends compared to the relative difference in likelihood between English strangers and friends. Further socio-cultural contextual findings reveal that most Polish-born participants who migrated due to cultural differences between Poland and England were LGBTQ+. These factors together suggest that the Polish-born LGBTQ+ participant

group do not have strong ties or affiliation with Polish culture, and possibly feel a sense of detachment with their Polish identities. This may be the reason why Polish-born LGBTQ+ participants are orienting more towards behaviours, values and norms associated with the LGBTQ+ community which is predicated on shared values. As mentioned previously, these shared values can provide motivation to bolster the use of nonstandard variation, which would explain this increased acceptance across the morphosyntactic variants considered in this study. However, this is only speculation as to why we might be seeing the group of LGBTQ+ Poles pattern more similarly to the LGBTQ+ community, rather than the Polish-born community.

6.5 Conclusion

In this chapter, I have discussed the findings from the descriptive and inferential analysis. I first discussed the patterns of acceptability across the three types of morphosyntactic construction considered in this thesis, and outlined the key patterns found. I then discussed how non-L1 Polish-born migrants integrate into these patterns, in the English cultural context, and how their patterns of acceptance of the different morphosyntactic constructions might relate to these participants' lived experiences and motivations. As part of this, I explored Polish-born migrant's potential acquisition of indexical associations, as well as meso-level factors which might be inhibiting their acquisition of nonstandard morphosyntactic variation. Next, I considered the meso-social factors of cultural embeddedness (acculturation), and how this interacts with Polish migrants' acquisition of sociolinguistic variation. I then explored patterns of acceptance of morphosyntactic variation, as well as divergence from sociolinguistic norms, with a focus on LGBTQ+ membership and embeddedness in the LGBTQ+ community. Finally, I explored how existing at the confluence of the two communities under investigation in the present study - ie. being a Polish-born LGBTQ+ migrant - impacts morphosyntactic variation.

7. Conclusions

This thesis has considered factors on participants' perceptions of morphosyntactic variation in a non-L1 (ie. non-native language). Often, the various factors relating to identities, community ties, and other lived experiences - which have been shown in L1 speakers to influence or constrain the use of language variation - are overlooked in studies on L2 migrants' acquisition of language, or are not considered in depth, in favour of focusing on language through the lens of participants' proficiency. While these approaches have their merits, it is known that extralinguistic factors play a complex and crucial role on the patterns of variation that we find. All types of communities are sites for the production of identity (Eckert & McConnell-Ginet, 2007: 29); migrants, as members of their target communities, are no exception to the effects of this, therefore, it is important to take into consideration migrant identities and the communities within which they are embedded when studying their language variation. A key benefit to exploring migrants' patterns of acceptance (ie. perception) of morphosyntactic variation across several dimensions - ie. from the perspective of the macro-social (demographic) level; in terms of L2 constraints on acquisition; and from the perspective of the meso-social (community-embeddedness) level - is that we can cluster these factors together, comparing between them to build up a richer picture of the complex interplay of effects that are influential, for a more nuanced understanding.

This has been done by taking a 'top-down' approach to the analysis of variation, aggregating data across a participant sample, and assessing the influence of various factors that have been reported to influence and constrain participants' acquisition of socially meaningful variation. Three methodological elements have been implemented to inform the results: (i) an Acceptability Judgement Task (AJT); (ii) the Versant English Language Speaking Test; and (iii) several sociological surveys. The factors that have been explored are split into five broad types: (i) the key sample characteristics of birth country and LGBTQ+ status; (ii) the macro-social (demographic) factors of interest, namely age, gender, region, and the socioeconomic factors of socioeconomic status and education status; (iii) the L2 (second language) factors influencing Polish-born participants, namely their age of arrival to England and their English language proficiency, and; (iv) the meso-social factors considered in this study (ie. those related to community embeddedness) - participants' levels of English and Polish acculturation, as well as their levels of LGBTQ+ Community Involvement.

Statistical effects of these factors on participants' acceptability judgements have been explored via four proportional odds (ordinal logistic regression) models, taking a nested approach in order to segment the sample according to the key characteristics under investigation (birth country and LGBTQ+ status). I assessed how acceptability ratings pattern across the morphosyntactic features of interest, according to the three types of morphosyntactic construction under investigation: Argument movement (raising-to-subject with the verb *seem*); ii. Optional discourse-based movement (left dislocation, right dislocation, topicalisation), and iii. Nonstandard agreement (past-tense BE: nonstandard *was*, nonstandard *were*, and nonstandard *weren't*).

7.1 Key Findings

Here, key findings from this study are reported by answering the research questions that were proposed in Chapter 1.

1. How do participants pattern in their acceptability judgement ratings of British-English morphosyntactic features by type of morphosyntactic construction - argument movement; optional discourse-based movement; nonstandard agreement? How do Polish-born participants' acceptability judgements compare to those of English-born participants?

The findings from both descriptive and inferential analysis reveal a clear pattern in the relative acceptability of the three types of morphosyntactic construction - from most to least acceptable: argument movement; optional discourse-based movement; nonstandard agreement. It is not only L1 speakers of British-English that are sensitive to the patterns of acceptability based on type of morphosyntactic construction: Polish-born participants acceptance is also similarly stratified, though English-born participants are more accepting of all statistically significant morphosyntactic variants than Polish-born participants. The relative acceptability of variants is found to be somewhat fluid, varying, for instance, according to the norms and practices among certain communities (such as the LGBTQ+ community).

2. To what extent can factors relating to formal linguistic accounts (eg. degree of prior linguistic input; linguistic proficiency) capture the patterns of variation found in Polish-born participants' acceptability judgement ratings?

It is clear that English language proficiency has an effect on acceptance ratings among Polish-born participants, but the finding is surprising in that higher proficiency Poles are *less* accepting than those with lower proficiency for all significant morphosyntactic constructions, both optional discourse-based movement and nonstandard agreement. This could be due to prescriptive ideology impacting higher-educated migrants' judgements. The categorical measure of years of English exposure was not found to have a clear effect.

3. What more can we learn about participants' acceptance of morphosyntactic variants by considering macro-social factors (eg. age; gender; socioeconomic class)?

Macro-social factors are beneficial for broadening understanding of particular effects observed. For instance, the acceptability of optional discourse-based movement among LGBTQ+ participants was found to be influenced by the macro-social factors of gender and age, which are known to be potential constraints on the use of right dislocation. Polish-born participants pattern differently to LGBTQ+ participants in that age increases their acceptance, suggesting a potential acculturation side-effect.

4. Do participants' acceptability judgement ratings pattern differently according to participants' British or Polish acculturation levels?

Results indicate the presence of an acculturation effect on at least two of the morphosyntactic constructions, one an optional discourse-based movement construction (right dislocation) and one a nonstandard agreement construction

(nonstandard *was*). Among Polish migrants, it seems that nonstandard agreement undergoes a stronger cumulative increase in acceptability with each increment of acculturation level than optional discourse-based movement does.

5. How does participants' acceptance of morphosyntactic variation vary according to participants' membership to, and level of embeddedness within, the LGBTQ+ community, compared with the overall Polish-born and English-born populations?

LGBTQ+ participants are around twice as accepting of optional discourse-based movement and nonstandard agreement than non-LGBTQ+ participants. I suggest this is reflective of the practices, values, and a sense of shared solidarity that this community is predicated upon. Inferential results confirm that participants with higher levels of LGBTQ+ Community Involvement are less accepting of optional discourse-based movement compared to those with lower involvement, but that right dislocation is the least negatively impacted by this. I propose that, because the use of these features requires cultural engagement with the dominant speech community, the lived experiences of LGBTQ+ community members - especially ones more deeply embedded - mean they orient in opposition to hegemonic norms. Inferential results confirm that being LGBTQ+ has a positive effect on Polish-born participants' acceptability across all morphosyntactic variants. Polish acculturation was not found to have a statistically significant effect on participants' acceptability judgements.

7.2 Future Directions

There are a number of ways future work could expand on the findings seen here.

Because morphosyntactic variants can possess both discourse-pragmatic and socioindexical properties, varying according to requirements from the communities using them, this information can be used to make predictions about the future of sociolinguistic variation. For instance, we could assess whether the patterning of different optional discourse-based movement constructions is narrowing, in terms of the functions they are being used to carry out. If we see the distributions in ways they are being used becoming more restricted, this suggests these variants are shifting towards primarily socioindexical usage as they accrue more social meaning. It is not that they are losing their discourse-pragmatic functions necessarily, but could shift in emphasis as socioindexical functions take priority.

In terms of the linguistic features under investigation, it would be beneficial to test the relative effects of different *types* of the features investigated here - for instance, right dislocation with pronoun tags, versus other, less socioindexically linked forms of right dislocation, to assess whether we are indeed seeing that LGBTQ+ community members orient to the socioindexical properties of right dislocation. It would also be worthwhile to consider participants' acceptance of a wider range of morphosyntactic constructions, in particular, other argument movement constructions than the one considered in this study, and to perhaps integrate these more fully into the analysis, as opposed to functioning as a reference category for measuring the effects of the other two types of constructions.

AJT methodologies have been problematised for their predisposition towards focus on form, which is especially important when testing the effects of variants which are most commonly used in discourse. Future research could build on the results from this study, assessing whether we can verify these patterns of acceptance using elicitation methodologies (which would also mitigate the known caveats to relying on participants' self-reported data), or through working with participants more closely (ie. ethnographically).

8. References

Ackerman, L. (2019) *Ordinal Data*, *lmackerman.com*. Available at: https://lmackerman.com/notebooks/ordinal_data.html (Accessed: 17 July 2023).

Acosta, K.L. (2008) 'Lesbianas in the Borderlands: Shifting Identities and Imagined Communities', *Gender & Society*, 22(5), pp. 639–659. Available at: https://doi.org/10.1177/0891243208321169.

Adamson, H.D. and Regan, V.M. (1991) 'The Acquisition of Community Speech Norms by Asian Immigrants Learning English as a Second Language: A Preliminary Study', *Studies in Second Language Acquisition*, 13(1), pp. 1–22. Available at: https://doi.org/10.1017/S0272263100009694.

Adger, D. (2006) 'Combinatorial Variability', *Journal of Linguistics*, 42(3), pp. 503–530. Available at: http://dx.doi.org.sheffield.idm.oclc.org/10.1017/S002222670600418X.

Adger, D. and Smith, J. (2005) 'Variation and the minimalist programme', in L.M.E.A. Cornips and K.P. Corrigan (eds) *Syntax and Variation: Reconciling the Biological and the Social*. John Benjamins, pp. 149–178. Available at: http://eprints.gla.ac.uk/6174/.

Adger, D. and Trousdale, G. (2007) 'Variation in English syntax: theoretical implications', *English Language & Linguistics*, 11(2), pp. 261–278.

Ameijeiras-Alonso, J., Crujeiras, R.M. and Rodríguez-Casal, A. (2021) 'multimode: An R Package for Mode Assessment', *Journal of Statistical Software*, 97(9), pp. 1–32. Available at: https://doi.org/10.18637/jss.v097.i09.

Anderson, B. (1986) *Imagined communities: Reflections on the origin and spread of nationalism*. London: Verso.

Anderwald, L. (2001) 'Was/Were-variation in non-standard British English today', *English World-Wide*, 22(1), pp. 1–21. Available at: https://doi.org/10.1075/eww.22.1.02and.

Anwyl-Irvine, A. *et al.* (2019) 'Gorilla in our Midst: An online behavioral experiment builder', *bioRxiv*, p. 438242. Available at: https://doi.org/10.1101/438242.

Arnett, J.J. (2000) 'Emerging adulthood: A theory of development from the late teens through the twenties', *American Psychologist*, 55, pp. 469–480. Available at: https://doi.org/10.1037/0003-066X.55.5.469.

Ashman, K.A., Bird, C.M. and Zepf, S.E. (1994) 'Detecting bimodality in astronomical datasets', *The Astronomical Journal*, 108, p. 2348. Available at: https://doi.org/10.1086/117248.

Baayen, R. (2008) *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baker, P. (2003) *Polari-the lost language of gay men*. Routledge.

Barrett, R. (2018) 'Speech play, gender play, and the verbal artistry of queer argots', *Suvremena Lingvistika; Zagreb*, 44(86), pp. 215–242. Available at: http://dx.doi.org/10.22210/suvlin.2018.086.03.

Bassetti, B. and Cook, V. (2011) *Language and bilingual cognition*. New York: Psychology Press.

Bayley, R. and Regan, V. (2004) 'Introduction: The acquisition of sociolinguistic competence', *Journal of Sociolinguistics*, 8(3), pp. 323–338.

Beal, J. (2008) 'English dialects in the North of England: morphology and syntax', in *1 The British Isles*. De Gruyter Mouton, pp. 373–403.

Becker, K., Khan, S. ud D. and Zimman, L. (2022) 'Beyond binary gender: creaky voice, gender, and the variationist enterprise', *Language Variation and Change*, 34(2), pp. 215–238. Available at: https://doi.org/10.1017/S0954394522000138.

Becker, K. and Stoddard, M. (2018) 'Variation in /s/-fronting in a diverse sample of s/g/s/', in. *New Ways of Analyzing Variation 47*. Available at: https://drive.google.com/file/d/0B0PyzGlyqnkXYTE5SE9oTFBMM2otUjJWQzZMTkF3QXg0cjl3/view?usp=sharing (Accessed: 22 October 2018).

Bender, E.M. (2001) *Syntactic variation and linguistic competence: The case of AAVE copula absence*. Stanford University.

Berry, J.W. (1979) 'Research in multicultural societies: Implications of cross-cultural methods', *Journal of Cross-Cultural Psychology*, 10(4), pp. 415–434.

Berry, J.W. (1980) 'Acculturation as varieties of adaptation', *Acculturation: Theory, models and some new findings*, 9, p. 25.

Berry, J.W. *et al.* (2006) 'Immigrant Youth: Acculturation, Identity, and Adaptation', *Applied Psychology*, 55(3), pp. 303–332. Available at: https://doi.org/10.1111/j.1464-0597.2006.00256.x.

Bever, T. (1970) 'The Cognitive Basis for Linguistic Structures', in, pp. 279–352. Available at: https://doi.org/10.1093/acprof:oso/9780199677139.003.0001.

Birman, D. and Trickett, E.J. (2001) 'Cultural transitions in first-generation immigrants: Acculturation of Soviet Jewish refugee adolescents and parents', *Journal of Cross-Cultural Psychology*, 32, pp. 456–477. Available at: https://doi.org/10.1177/0022022101032004006.

Birman, D. and Tyler, F.B. (1994) 'Acculturation and alienation of Soviet Jewish refugees in the United States.', *Genetic, social, and general psychology monographs* [Preprint].

Blommaert, J. (2010) *The sociolinguistics of globalization*. Cambridge University Press.

Blommaert, J. (2014) 'Commentary: Superdiversity old and new', *Language & Communication*, 44, pp. 82–88.

Britain, D. (2002a) 'Diffusion, levelling, simplification and reallocation in past tense BE in the English Fens', *Journal of Sociolinguistics*, 6(1), pp. 16–43. Available at: https://doi.org/10.1111/1467-9481.00175.

Britain, D. (2002b) 'Surviving "Estuary English": innovation diffusion, koineisation and local dialect differentiation in the English Fenland'. Available at: http://repository.essex.ac.uk/152/1/errl_41c.pdf.

Britain, D. (2007) 'Grammatical Variation in England', in D. Britain (ed.) *Language in the British Isles*. Cambridge: Cambridge University Press, pp. 75–104.

Britain, D. (2010) 'Grammatical variation in the contemporary spoken English of England', in A. Kirkpatrick (ed.) *Handbook of World Englishes*. Routledge, pp. 37–58.

Britain, D. (2020) 'What happened to those relatives from East Anglia? A Multilocality Analysis of Dialect Levelling in the Relative Marker System', in K. Beaman et al. (eds) *Advancing Socio-grammatical Variation and Change: In Honour of Jenny Cheshire*. New York: Routledge, pp. 93–114.

Brook, M. *et al.* (2018) '"I don't come off as timid anymore": Real-time change in early adulthood against the backdrop of the community', *Journal of Sociolinguistics*, 22(4), pp. 351–374.

Bross, F. (2019) 'Acceptability Ratings in Linguistics: A Practical Guide to Grammaticality Judgments, Data Collection, and Statistical Analysis. Version 1.02'. Available at: www.fabianbross.de/acceptabilityratings.pdf (Accessed: 3 July 2022).

Bucholtz, M. (1999) '"Why be normal?": Language and identity practices in a community of nerd girls', *Language in Society*, pp. 203–223.

Bucholtz, M. and Hall, K. (2005) 'Identity and interaction: A sociocultural linguistic approach', *Discourse studies*, 7, pp. 4–5.

Calder, J. (2020) 'Language and Sexuality: Language and LGBTQ+ Communities', *The International Encyclopedia of Linguistic Anthropology*, pp. 1–7.

Callies, M. (2008) 'Easy to understand but difficult to use? Raising constructions and information packaging in the advanced learner variety', in *Linking up contrastive and learner corpus research*. Brill, pp. 199–226.

Cameron, D. and Kulick, D. (2003) *Language and Sexuality*. Cambridge: Cambridge University Press. Available at: https://doi.org/10.1017/CBO9780511791178.

Campbell-Kibler, K. *et al.* (2006) 'The elements of style', in *Poster presented at New Ways of Analyzing Variation*.

Carter, R. and McCarthy, M. (1995) 'Grammar and the Spoken Language', *Applied Linguistics*, 16(2), pp. 141–58.

Chambers, J.K. (1992) 'Dialect acquisition', *Language*, 68(4), pp. 673–705.

Chambers, J.K. (2003) 'Sociolinguistics of immigration', in D. Britain and J. Cheshire (eds) *Social Dialectology: In Honour of Peter Trudgill*. Amsterdam: John Benjamins, pp. 97–114.

Chambers, J.K. (2009) 'Cognition and the linguistic continuum from vernacular to standard', *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*, pp. 19–32.

Chen, S.-Y., Feng, Z. and Yi, X. (2017) 'A general introduction to adjustment for multiple comparisons', *Journal of Thoracic Disease*, 9(6), pp. 1725–1729. Available at: https://doi.org/10.21037/jtd.2017.05.34.

Cheshire, J. (1982) 'Linguistic Variation and Social Function', *Cambridge Studies in Linguistics London*, 37.

Cheshire, J. (1996) 'Syntactic variation and the concept of prominence', in J. Klemola, M. Kyto, and M. Rissanen (eds) *Speech past and present: essays in English dialectology in memory of Ossi Ihalainen*. Frankfurt: Peter Lang, pp. 1–17.

Cheshire, J. (2005) 'Syntactic variation and spoken language', in L. Cornips and K.P. Corrigan (eds) *Syntax and Variation: Reconciling the Biological and the Social*. John Benjamins Publishing Company (Current Issues in Linguistic Theory), pp. 81–106. Available at: https://doi.org/10.1075/cilt.265.05che.

Cheshire, J. and Edwards, V. (1998) 'Lessons from a survey of British dialect grammar', *Links & letters*, pp. 61–73.

Cheshire, J., Edwards, V. and Whittle, P. (1993) 'Non-standard English and dialect levelling', *Real English: the grammar of English dialects in the British Isles*, pp. 53–96.

Cheshire, J. and Fox, S. (2009) 'Was/Were Variation: A Perspective from London', *Language Variation and Change*, 21(1), pp. 1–38.

Cheshire, J., Kerswill, P. and Williams, A. (2005) 'Phonology, grammar, and discourse in dialect convergence', *Dialect change: Convergence and divergence in European languages*, pp. 135–167.

Choe, J. and Deen, K. (2016) 'Children's difficulty with raising: A performance account', *Language Acquisition*, 23(2), pp. 112–141. Available at: https://doi.org/10.1080/10489223.2015.1047097.

Christensen, R.H.B. (2015) 'Analysis of ordinal data with cumulative link models—estimation with the R-package ordinal', *R-package version*, 28, p. 406.

Christensen, R.H.B. (2018) 'Cumulative Link Models for Ordinal Regression with the R Package ordinal', p. 40.

Christensen, R.H.B. (2023) 'ordinal'. Available at: https://github.com/runehaubo/ordinal (Accessed: 9 August 2023).

Christensen, K.R., Kizach, J. and Nyvad, A.M. (2013) 'Escape from the Island: Grammaticality and (Reduced) Acceptability of wh-island Violations in Danish', *Journal of Psycholinguistic Research*, 42(1), pp. 51–70. Available at: https://doi.org/10.1007/s10936-012-9210-x

Cornips, L. (2006) 'Variation and Formal Theories of Syntax, Chomskyan', in *Encyclopedia of Language & Linguistics, 2nd Edition*. Elsevier BV, pp. 330–332.

Cornips, L. (2014) 'Language contact, linguistic variability and the construction of local identities', in T.A. Åfarli and B. Maehlum (eds) *The Sociolinguistics of Grammar*. John Benjamins Publishing Company, pp. 67–90.

Cornips, L. (2015) 'The No Man's Land between Syntax and Variationist Sociolinguistics: The Case of Idiolectal Variability', in A. Adli, M. García García, and G. Kaufmann (eds) *Variation in Language: System- and Usage-Based Approaches*. Germany: de Gruyter, pp. 147–172.

Cornips, L.M.E.A. and Corrigan, K.P. (eds) (2005) *Syntax and Variation: Reconciling the Biological and the Social*. Amsterdam: John Benjamins Publishing Company (Current Issues in Linguistic Theory).

Cortés, D.E., Rogler, L.H. and Malgady, R.G. (1994) 'Biculturality among puerto rican adults in the united states', *American journal of community psychology*, 22(5), pp. 707–721.

Coupland, N. (2007) *Style: Language variation and identity*. Cambridge University Press.

Cuéllar, I., Arnold, B. and González, G. (1995) 'Cognitive referents of acculturation: Assessment of cultural constructs in Mexican Americans', *Journal of Community Psychology*, 23(4), pp. 339–356. Available at: https://doi.org/10.1002/1520-6629(199510)23:4<339::AID-JCOP2290230406>3.0.CO;2-7.

Cuellar, I., Arnold, B. and Maldonado, R. (1995) 'Acculturation Rating Scale for Mexican Americans-II: A Revision of the Original ARSMA Scale', *Hispanic Journal of Behavioral Sciences*, 17(3), pp. 275–304. Available at: https://doi.org/10.1177/07399863950173001.

D'Arcy, A. (2005) *Like: Syntax and development*. Unpublished PhD dissertation. University of Toronto.

D'Arcy, A. (2007) 'Like and language ideology: Disentangling fact from fiction', *American speech*, 82(4), pp. 386–419.

Domino, G. and Acosta, A. (1987) 'The relation of acculturation and values in Mexican Americans', *Hispanic Journal of Behavioral Sciences*, 9(2), pp. 131–150.

Doran, M. (2004) '3. Negotiating Between Bourge and Racaille: Verlan as Youth Identity Practice in Suburban Paris', in *Negotiation of Identities in Multilingual Contexts*, pp. 93–124. Available at: https://doi.org/10.21832/9781853596483-006.

Dörnyei, Z. and Taguchi, T. (2010) *Questionnaires in second language research: construction, administration, and processing*. 2nd ed. New York ; London: Routledge (Second language acquisition research. Monographs on research methodology).

Drummond, R.J. (2010) *Sociolinguistic variation in a second language: The influence of local accent on the pronunciation of non-native English speakers living in Manchester*. PhD dissertation. The University of Manchester.

Durham, M. (2011) 'Right dislocation in Northern England: Frequency and use — perception meets reality', *English World-Wide*, 32(3), pp. 257–279. Available at: https://doi.org/10.1075/eww.32.3.01dur.

Eckert, P. (2008) 'Variation and the indexical field', *Journal of Sociolinguistics*, 12(4), pp. 453–476.

Eckert, P. (2011) 'Where does the social stop?', in F. Gregersen, J.K. Parrott, and P. Quist (eds) *Language Variation - European Perspectives III: Selected papers from the 5th International Conference on Language Variation in Europe (ICLaVE 5), Copenhagen, June 2009*. John Benjamins Publishing Company (Studies in Language Variation), pp. 13–30. Available at: https://doi.org/10.1075/silv.7.02eck.

Eckert, P. (2012) 'Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation', *Annual Review of Anthropology*, 41, pp. 87–100.

Eckert, P. (2019) 'The limits of meaning: Social indexicality, variation, and the cline of interiority', *Language*, 95(4), pp. 751–776. Available at: https://doi.org/10.1353/lan.2019.0072.

Eckert, P. and McConnell-Ginet, S. (1992) 'Think practically and look locally: language and gender as community-based practice', *Annual review of*, 21(1), pp. 461–88.

Eckert, P. and McConnell-Ginet, S. (2007) 'Putting communities of practice in their place', *Gender & Language*, 1(1), pp. 27–37.

Eisikovits, E. (1991) 'Variation in subject–verb agreement in Inner Sydney English', in J. Cheshire (ed.) *English around the World: Sociolinguistic Perspectives*. Cambridge: Cambridge University Press, pp. 235–255. Available at: https://doi.org/10.1017/CBO9780511611889.017.

Ellis, A.J. (1869a) *On Early English Pronunciation*. London: Trübner.

Ellis, A.J. (1869b) *On early English pronunciation. 5 vols*. London: Asher & Co. and Trübner & Co.

Erickson, F. (1973) 'What Makes School Ethnography "Ethnographic?"', *Council on Anthropology and Education Newsletter*, 4(2), pp. 10–19.

Farver, J.A.M., Narang, S.K. and Bhadha, B.R. (2002) 'East meets West: Ethnic identity, acculturation, and conflict in Asian Indian families', *Journal of Family Psychology*, 16, pp. 338–350. Available at: https://doi.org/10.1037/0893-3200.16.3.338.

Fasold, R.W. (1996) '4. The quiet demise of variable rules', in R. Singh (ed.) *Towards a Critical Sociolinguistics*. John Benjamins Publishing Company (Current Issues in Linguistic Theory), p. 79. Available at: https://doi.org/10.1075/cilt.125.07fas.

Feagin, C. (1979) *Variation and change in Alabama English: a sociolinguistic study of the White community*. Washington, D.C.: Georgetown University Press.

Fillmore, C.J. (1968) 'The Case for Case', in E. Bach and R.T. Harms (eds) *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston, pp. 1–88.

Fruehwald, J. (2017) 'Generations, lifespans, and the zeitgeist', *Language Variation and Change*, 29(1), pp. 1–27.

Gardner, R.C. and Lambert, W.E. (1959) 'Motivational variables in second-language acquisition', *Canadian Journal of Psychology/Revue canadienne de psychologie*, 13(4), p. 266.

Gardner, R.C. and Lambert, W.E. (1972) *Attitudes and motivation in second-language learning*. Massachusetts: Newbury House Publishers, Inc.

Garrett, P. (2010) *Attitudes to Language*. Cambridge: Cambridge University Press.

Giles, H. and Powesland, P.F. (1975) *Speech style and social evaluation*. Oxford, England: Academic Press (Speech style and social evaluation), pp. viii, 218.

Givon, T. (2001) *Syntax: An Introduction*. John Benjamins Publishing Company. Available at: https://benjamins.com/catalog/z.syn2 (Accessed: 4 June 2023).

Golden, J.G. (1988) *Assimilation or Biculturalism – Which Way To Adapt to American Life?* Denver, CO: ERIC, p. 12

Golebiowska, E.A. (2014) *The Many Faces of Tolerance: Attitudes toward Diversity in Poland*. New York; London: Routledge.

GOV.UK (2019) *English indices of deprivation 2019*, *GOV.UK*. Available at: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019 (Accessed: 19 July 2022).

Greenbaum, S. (1977) 'Contextual Influence on Acceptability Judgments', *Linguistics* [Preprint].

Greenberg, G.R. (1984) 'Left dislocation, topicalization, and interjections', *Natural Language and Linguistic Theory*, 2(3), pp. 283–287. Available at: https://doi.org/10.1007/BF00133789.

Guo, X. *et al.* (2009) 'Some evidence for multidimensional biculturalism: Confirmatory factor analysis and measurement invariance analysis on the Bicultural Involvement Questionnaire–Short Version.', *Psychological Assessment*, 21(1), pp. 22–31. Available at: https://doi.org/10.1037/a0014495.

Hammer, K. (2017) 'Language through a prism: Patterns of L2 internalisation and use in acculturated bilinguals', *Journal of Pragmatics*, 117, pp. 72–87. Available at: https://doi.org/10.1016/j.pragma.2017.06.006.

Hay, J. and Schreier, D. (2004) 'Reversing the trajectory of language change: Subject–verb agreement with be in New Zealand English', *Language Variation and Change*, 16(3), pp. 209–235.

Hazenberg, E.N.L. (2012) *Language and Identity Practice: A Sociolinguistic Study of Gender in Ottawa, Ontario*. PhD dissertation. Memorial University of Newfoundland.

Hazenberg, E.N.L. (2016) 'Walking the straight and narrow: linguistic choice and gendered presentation', *Gender and Language*, 10(2), pp. 270–294.

Hebdige, D. (2012) *Subculture: The meaning of style*. Routledge.

Henry, A. (1995) *Belfast English and Standard English: Dialect Variation and Parameter Setting*. Oxford University Press.

Henry, A. (2005) 'Non-standard dialects and linguistic data', *Lingua*, 115(11), pp. 1599–1617. Available at: https://doi.org/10.1016/j.lingua.2004.07.006.

Hodge, D.R. and Gillespie, D.F. (2007) 'Phrase completion scales: a better measurement approach than Likert scales?', *Journal of Social Service Research*, 33(4), pp. 1–12.

Howley, G. (2015) *The acquisition of Manchester dialect variants by adolescent Roma migrants*. PhD dissertation. University of Salford.

Huddleston, R. and Pullum, G.K. (2002) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press. Available at: https://doi.org/10.1017/9781316423530.

Hudson, R.A. (2007) 'English Dialect Syntax in Word Grammar', *English Language and Linguistics*, 11(2), pp. 383–405.

Hulstijn, J.H. (2011) 'Language Proficiency in Native and Nonnative Speakers: An Agenda for Research and Suggestions for Second-Language Assessment', *Language Assessment Quarterly*, 8(3), pp. 229–249. Available at: https://doi.org/10.1080/15434303.2011.565844.

Hymes, D. (1972) 'On Communicative Competence', in J.B. Pride and J. Holmes (eds) *Sociolinguistics. Selected Readings*. Hammondsworth: Penguin, pp. 269–293.

Jamieson, E., Smith, J., Adger, D., Heycock, C. and Thomas, G. (2023) 'When intuitions (don't) fail: Combining syntax and sociolinguistics in the analysis of Scots', *English Language and Linguistics* [Preprint].

Jespersen, O. (1928) *A Modern English Grammar on Historical Principles-Part 1*. Carl Winters Univ.

Jiang, M. *et al.* (2009) 'Acculturation in relation to the acquisition of a second language', *Journal of Multilingual and Multicultural Development*, 30(6), pp. 481–492. Available at: https://doi.org/10.1080/01434630903147898.

Johnson, J.S. and Newport, E.L. (1989) 'Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language', *Cognitive psychology*, 21(1), pp. 60–99.

Johnstone, B. (2010) 'Locating language in identity', in *Language and identities*, pp. 29–36.

Jones, L. (2012) *Dyke/Girl: Language and Identities in a Lesbian Group*. Basingstoke: Palgrave Macmillan.

Jones, L. (2022) '"I'm a boy, can't you see that?": Dialogic embodiment and the construction of agency in trans youth discourse', *Language in Society*, pp. 1–22. Available at: https://doi.org/10.1017/S0047404522000252.

Kang, S-M. (2006) 'Measurement of Acculturation, Scale Formats, and Language Competence: Their Implications for Adjustment', *Journal of Cross-Cultural Psychology*, 37(6), pp. 669–693. Available at: https://doi.org/10.1177/0022022106292077.

Kerswill, P. (1996) 'Children, adolescents, and language change', *Language Variation and Change*, 8(2), pp. 177–202. Available at: https://doi.org/10.1017/S0954394500001137.

Kerswill, P. (2006) 'Migration and language', in *Sociolinguistics/Soziolinguistik. An international handbook of the science of language and society*, pp. 1–27.

Kirkham, S. (2013) *Ethnicity, social practice and phonetic variation in a Sheffield secondary school*. PhD Thesis. University of Sheffield.

Kirkham, S. and Moore, E. (2013) 'Adolescence', in J.K. Chambers and N. Schilling-Estes (eds) *The Handbook of Language Variation and Change*. 2nd edn. Malden, MA: Wiley-Blackwell, pp. 277–296. Available at: http://goo.gl/T6lpHu.

Konnelly, L. and Cowper, E. (2020) 'Gender diversity and morphosyntax: An account of singular they', *Glossa: a journal of general linguistics*, 5(1). Available at: https://doi.org/10.5334/gjgl.1000.

Konnelly, L. and Tagliamonte, S.A. (unpublished manuscript) 'LGBTQ+ Youth within the Speech Community: Enriching Large Scale Studies of Variation and Change'.

Krashen, S. (1985) 'The Input Hypothesis. Beverly Hills', *Laredo Publishing Company. Krashen, S.(1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. Modern Language Journal*, 73(4), pp. 440–464.

Krashen, S. (1994) 'The pleasure hypothesis', *Georgetown University round table on languages and linguistics*, pp. 299–302.

Krashen, S.D. (1999) 'Seeking a Role for Grammar. A Review of Some Recent Studies', *Foreign Language Annals*, 32(2), pp. 245–254. Available at: https://doi.org/10.1111/j.1944-9720.1999.tb02395.x.

Kroch, A.S. (1989) 'Reflexes of grammar in patterns of language change', *Language Variation and Change*, 1(3), pp. 199–244. Available at: https://doi.org/10.1017/S0954394500000168.

Labov, W. (1964) 'Phonological correlates of social stratification', *American Anthropologist*, 66(6), pp. 164–176.

Labov, W. (1966) *The Social Stratification of English in New York City*. Washington: Center.

Labov, W. *et al.* (1968) 'A Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York City. Volume I: Phonological and Grammatical Analysis.'

Labov, W. (1972) *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1993) 'The unobservability of structure and its linguistic consequences', in *NEW WAYS IN ANALYZING VARIATION CONFERENCE*.

Labov, W. (2001) *Principles of Linguistic Change, Volume 2: Social factors*. Blackwell.

Lamont, M. and Molnár, V. (2002) 'The study of boundaries in the social sciences', *Annual review of sociology*, 28(1), pp. 167–195.

Lavandera, B.R. (1978) 'Where does the sociolinguistic variable stop?', *Language in Society*, 7(2), pp. 171–182. Available at: https://doi.org/10.1017/S0047404500005510.

Lee, C. (2001) *Sociocultural identity and second language learning: A study of Korean students in an American university*. The Ohio State University.

Lenneberg, E.H. (1967) 'The Biological Foundations of Language', *Hospital Practice*, 2(12), pp. 59–67. Available at: https://doi.org/10.1080/21548331.1967.11707799.

Levon, E. (2011) 'Teasing Apart to Bring Together: Gender and Sexuality in Variationist Research', *American Speech*, 86(1), pp. 69–84. Available at: https://doi.org/10.1215/00031283-1277519.

Levon, E. (2015) 'Integrating intersectionality in language, gender, and sexuality research', *Language and Linguistics Compass*, 9(7), pp. 295–308.

Levon, E. and Buchstaller, I. (2015) 'Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing', *Language Variation and Change*, 27(3), pp. 319–348. Available at: https://doi.org/10.1017/S0954394515000149.

Lewis, M.M. (1969) *Language and the Child (Exploring Education)*. Slough: National Foundation for Educational Research.

Liu, D. and Zhang, H. (2018) 'Residuals and Diagnostics for Ordinal Regression Models: A Surrogate Approach', *Journal of the American Statistical Association*, 113(522), pp. 845–854. Available at: https://doi.org/10.1080/01621459.2017.1292915.

Livia, A. and Hall, K. (1997) *Queerly phrased: Language, gender, and sexuality*. Oxford: Oxford University Press.

Lopez-Class, M., Castro, F.G. and Ramirez, A.G. (2011) 'Conceptions of acculturation: A review and statement of critical issues', *Social science & medicine*, 72(9), pp. 1555–1562.

Luna, D., Ringberg, T. and Peracchio, L.A. (2008) 'One individual, two identities: Frame switching among biculturals', *Journal of Consumer Research*, 35, pp. 279–293. Available at: https://doi.org/10.1086/586914.

Lybeck, K. (2002) 'Cultural Identification and Second Language Pronunciation of Americans in Norway', *The Modern Language Journal*, 86(2), pp. 174–191.

Macaulay, R.K.S. (1977) *Language, Social Class and Education: A Glasgow Study*. Edinburgh: Univ. Edinb. Press

Macaulay, R.K.S. (1989) 'He was some man him: Emphatic pronouns in Scottish English', in T. J. Walsh (ed) *Synchronic and Diachronic Approaches to Linguistic Variation and Change*, pp. 179–187. Washington, DC: Georgetown University Press.

Macaulay, R.K.S. (1991) *Locating Discourse in Dialect: The Language of Honest Men and Bonnie Lasses in Ayr*. New York: Oxford University Press.

Macaulay, R.K.S. (2005) *Talk that counts: age, gender, and social class differences in discourse*. Oxford: University Press. Available at: http://dx.doi.org/10.1093/acprof:oso/9780195173819.001.0001 (Accessed: 5 May 2023).

Martin, R.R. and Haroldson, S.K. (1992) 'Stuttering and Speech Naturalness', *Journal of Speech, Language, and Hearing Research*, 35(3), pp. 521–528. Available at: https://doi.org/10.1044/jshr.3503.521.

Masgoret, A. –M. and Gardner, R.C. (2003) 'Attitudes, Motivation, and Second Language Learning: A Meta–Analysis of Studies Conducted by Gardner and Associates', *Language Learning*, 53(1), pp. 123–163. Available at: https://doi.org/10.1111/1467-9922.00212.

Mateu, V. and Hyams, N. (2021) 'On children's late acquisition of raising seem and control promise: Is a unified account possible?Victoria Mateu and Nina Hyams', in A. Belletti and C. Collins (eds) *Smuggling in Syntax*. Oxford University Press, p. 0. Available at: https://doi.org/10.1093/oso/9780197509869.003.0009.

Mcdonald, J.L. (2000) 'Grammaticality judgments in a second language: Influences of age of acquisition and native language', *Applied Psycholinguistics*, 21(3), pp. 395–423. Available at: https://doi.org/10.1017/S0142716400003064.

McMillan, D.W. and Chavis, D.M. (1986) 'Sense of community: A definition and theory', *Journal of community psychology*, 14(1), pp. 6–23.

Meechan, M. and Foley, M. (1994) 'On resolving disagreement: Linguistic theory and variation – There's bridges', *Language Variation and Change*, 6(1), pp. 63–85. Available at: https://doi.org/10.1017/S0954394500001587.

Mesthrie, R. (1996) 'Putting some linguistics into applied linguistics: a sociolinguistic study of left dislocation in South African black English', *Stellenbosch papers in linguistics plus*, 1996(29), pp. 260–283.

Mesthrie, R. (2014) 'Analyzing Sociolinguistic Variation in Multilingual Contexts', in J. Holmes (ed.) *Research methods in sociolinguistics: a practical guide*. Hoboken: Wiley Blackwell, pp. 276–289.

Meyerhoff, M. and Klaere, S. (2017) *Big issues with smaller samples in language variation: A case for clustering speakers and linguistic variables*, *Language Variation - European Perspectives VI: Selected papers from the Eighth International Conference on Language Variation in Europe (ICLaVE 8), Leipzig, May 2015*. Edited by I. Buchstaller and B. Siebenhaar. John Benjamins Publishing Company (Studies in Language Variation), pp. 23–46. Available at: https://doi.org/10.1075/silv.19.02mey.

Meyerhoff, M. and Walker, J.A. (2013) 'An existential problem: The sociolinguistic monitor and variation in existential constructions on Bequia (St. Vincent and the Grenadines)', *Language in Society*, 42(4), pp. 407–428. Available at: https://doi.org/10.1017/S0047404513000456.

Miller, M.J. (2007) 'A bilinear multidimensional measurement model of Asian American acculturation and enculturation: Implications for counseling interventions.', *Journal of counseling psychology*, 54(2), p. 118.

Milroy, J. and Milroy, L. (1985) 'Linguistic change, social network and speaker innovation', *Journal of Linguistics*, 21, pp. 339–384.

Milroy, J. and Milroy, L. (1997) 'Network structure and linguistic change', in *Sociolinguistics*. Springer, pp. 199–211.

Milroy, L. (1980) *Language and Social Networks*. Oxford: Basil Blackwell.

Milroy, L. (1987) *Language and Social Networks*. 2nd edn. Malden, MA: Blackwell.

Mokrosz, E. (2022) 'The Puzzles of Topic Dislocations in Polish', *Studies in Polish Linguistics*, 17(4), pp. 145–176.

Montgomery, C. (2012) 'The effect of proximity in perceptual dialectology', *Journal of Sociolinguistics*, 16(5), pp. 638–668. Available at: https://doi.org/10.1111/josl.12003.

Moore, E. (2003) *Learning style and identity: A sociolinguistic analysis of a Bolton high school*. PhD Thesis. University of Manchester.

Moore, E. (2004) 'Sociolinguistic Style: A Multidimensional Resource for Shared Identity Creation', *The Canadian Journal of Linguistics / La revue canadienne de linguistique*, 49(3), pp. 375–396. Available at: https://doi.org/10.1353/cjl.2006.0006.

Moore, E. (2010) 'Communities of practice and peripherality', in *Language and identities*, pp. 123–33.

Moore, E. (2011) 'Interaction between social category and social practice: explaining was/were variation', *Language Variation and Change*, 22(3), pp. 347–371. Available at: https://doi.org/10.1017/S095439451000013X.

Moore, E. (2020) 'The Role of Syntax in the Study of Sociolinguistic Meaning: Evidence from an Analysis of Right Dislocation', in *Advancing Socio-grammatical Variation and Change*. Routledge, pp. 73–90.

Moore, E. (2021) 'The Social Meaning of Syntax', in L. Hall-Lew, E. Moore, and R.J. Podesva (eds) *Social Meaning and Linguistic Variation: Theorizing the Third Wave*. Cambridge: Cambridge University Press, pp. 54–79.

Moore, E. (2023) *Socio-syntax: Exploring the Social Life of Grammar*. Cambridge: Cambridge University Press. Available at: DOI: 10.1017/9781108921299.

Moore, E. (forthcoming) 'Grammatical variation and social meaning', in C. Montgomery & E. Moore (eds) *The Oxford Handbook of British Englishes*. Oxford: Oxford University Press.

Moore, E. and Podesva, R.J. (2009) 'Style, indexicality, and the social meaning of tag questions', *Language in Society*, 38(4), pp. 447–485.

Moore, E. and Snell, J. (2011) '"Oh, they're top, them": Right dislocated tags and interactional stance', in F. Gregersen, J.K. Parrott, and P. Quist (eds) *Studies in Language Variation*. Amsterdam: John Benjamins Publishing Company, pp. 97–110. Available at: https://doi.org/10.1075/silv.7.08moo.

Moore, E. and Spencer, S. (2021) '"It just sounds proper common": Exploring the social meanings expressed by nonstandard grammar', *Linguistics and Education*, 63, p. 100933. Available at: https://doi.org/10.1016/j.linged.2021.100933.

Mori, S. and Gobel, P. (2006) 'Motivation and gender in the Japanese EFL classroom', *System*, 34(2), pp. 194–210.

Motschenbacher, H. (2011) 'Taking Queer Linguistics further: sociolinguistics and critical heteronormativity research', *International Journal of the Sociology of Language*, 2011(212), pp. 149–179.

Mougeon, R., Rehner, K. and Nadasdi, T. (2004) 'The learning of spoken French variation by immersion students from Toronto, Canada', *Journal of sociolinguistics*, 8(3), pp. 408–432.

Nance, C. (2013) *Phonetic variation, sound change, and identity in Scottish Gaelic*. PhD Thesis. University of Glasgow.

Nycz, J. (2018) 'Stylistic variation among mobile speakers: Using old and new regional variables to construct complex place identity', *Language Variation and Change*, 30(2), pp. 175–202. Available at: https://doi.org/10.1017/S0954394518000108.

Ochs, E. (1991) 'Socialization through language and interaction: A theoretical introduction', *Issues in Applied Linguistics*, 2(2).

Ochs, E. (1992) 'Indexing Gender', in A. Duranti and C. Goodwin (eds) *Rethinking context: Language as an interactive phenomenon*. Cambridge: Cambridge University Press.

Ochs, E. (1993) 'Constructing social identity: A language socialization perspective', *Research on Language and Social Interaction*, 26, pp. 287–306. Available at: https://doi.org/10.1207/s15327973rlsi2603_3.

Ochs, E. (1996) *Linguistic resources for socializing humanity*. Cambridge University Press.

ONS (2011) *2011 Census: Key Statistics and Quick Statistics for Local Authorities in the United Kingdom*, *Office for National Statistics*. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/keystatisticsandquickstatisticsforlocalauthoritiesintheunitedkingdom/2013-12-04 (Accessed: 26 June 2023).

ONS (2017) *Living abroad - migration between Britain and the EU8*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/livingabroad/migrationbetweenbritainandtheeu8 (Accessed: 9 August 2023).

ONS (2021) *Ethnic group, national identity and religion - Office for National Statistics*. Available at: https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion (Accessed: 20 July 2022).

Orfitelli, R. (2012) *Argument Intervention in the Acquisition of A-movement*. UCLA. Available at: https://escholarship.org/uc/item/524202gk.

Orfitelli, R. and Grüter, T. (2013) 'Do null subjects really transfer?', in J. Cabrelli Amaro, T. Judy, and D. Pascual y Cabo (eds) *Proceedings of the 12th Generative Approaches to Second Language Acquisition Conference (GASLA 2013). 12th Generative Approaches to Second Language Acquisition Conference (GASLA 2013)*, University of Florida, pp. 145–154. Available at: https://eprints.whiterose.ac.uk/97341/1/paper2991.pdf.

Papapolydorou, M. (2010) 'Educational inequalities in Greece, Sweden and the United Kingdom: A comparative analysis of the origins', in D. Mattheou (ed.). Dordrecht, Netherlands / Heidelberg, Germany / London, UK / New York, USA: Springer Netherlands, pp. 119–134. Available at: https://doi.org/10.1007/978-90-481-8534-4_8.

Pavlenko, A. (2014) *The Bilingual Mind: And What it Tells Us about Language and Thought*. Cambridge: Cambridge University Press. Available at: https://doi.org/10.1017/CBO9781139021456.

Payne, A. (1980) 'Factors controlling the acquisition of the Philadelphia dialect by out-of-state children', *Locating language in time and space*, 1, pp. 143–78.

Pearson (2011) 'Versant English Test: test description and validation summary'. Palo Alto, CA: Pearson Knowledge Technologies. Available at: http://www.versanttest.com.

Petyt, K.M. (1985) *'Dialect' and 'Accent' in Industrial West Yorkshire*. John Benjamins Publishing.

Pfister, R. *et al.* (2013) 'Good things peak in pairs: a note on the bimodality coefficient', *Frontiers in Psychology*, 4. Available at: https://doi.org/10.3389/fpsyg.2013.00700.

Phinney, J.S. and Ong, A.D. (2007) 'Conceptualization and measurement of ethnic identity: Current status and future directions', *Journal of Counseling Psychology*, 54, pp. 271–281. Available at: https://doi.org/10.1037/0022-0167.54.3.271.

Platt, J.T., Weber, H. and Lian, H.M. (1984) *The New Englishes*. Routledge & Kegan Paul.

Podesva, R.J. (2007) 'Phonation type as a stylistic variable: The use of falsetto in constructing a persona', *Journal of Sociolinguistics*, 11(4), pp. 478–504.

Podesva, R.J. and Van Hofwegen, J. (2014) 'How Conservatism and Normative Gender Constrain Variation in Inland California: The Case of /s/', in *University of Pennsylvania Working Papers in Linguistics*. Article (2).

Podesva, R.J. and Van Hofwegen, J. (2016) '/s/exuality in small-town California: Gender normativity and the acoustic realization of /s/', in E. Levon and R.B. Mendes (eds) *Language, Sexuality, and Power: Studies in Intersectional Sociolinguistics*. Oxford: Oxford University Press, pp. 168–188.

Quirk, R. (1985) *A Comprehensive grammar of the English language*. London: Longman. Available at: http://www.gbv.de/dms/hebis-mainz/toc/061277584.pdf (Accessed: 3 June 2023).

Race Disparity Unit (2021) *Ethnicity facts and figures*, *GOV.UK*. Available at: https://www.ethnicity-facts-figures.service.gov.uk/ (Accessed: 13 July 2023).

Rainbow Europe (2021) *Rainbow Map*, *Rainbow Europe*. Available at: https://rainbow-europe.org/ (Accessed: 22 April 2021).

Regan, V. (1995) 'The acquisition of sociolinguistic native speech norms', *Second language acquisition in a study abroad context*, pp. 245–267.

Roberts, J. (2005) 'Acquisition of sociolinguistic variation', in M.J. Ball (ed.) *Clinical sociolinguistics*. John Wiley & Sons, pp. 153–164.

Robinson, P. (1995) 'Aptitude, awareness, and the fundamental similarity of implicit and explicit second language learning', *Attention and awareness in foreign language learning*, 9, pp. 303–357.

Roeper, T. and Green, L. (2007) 'Node labels and features: Stable and unstable dialects and variation in acquisition', *Linguistic Variation Yearbook*, 7(1), pp. 1–26.

Ross, C. (2012) 'Imagined communities: initiatives around LGBTQ ageing in Italy', *Modern Italy*, 17(4), pp. 449–464. Available at: https://doi.org/10.1080/13532944.2012.706997.

RStudio Team (2020) *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, PBC. Available at: http://www.rstudio.com/.

Ryan, S.D. (2021) '"I just sound Sco[ʔ]ish now": The acquisition of word-medial glottal replacement by Polish adolescents in Glasgow', *English World-Wide. A Journal of Varieties of English*, 42(2), pp. 145–174. Available at: https://doi.org/10.1075/eww.00066.dur.

Sankoff, D. (1988) 'Variable Rules', in *Sociolinguistics: An International Handbook of the Science of Language and Society*. Berlin ; New York: Walter de Gruyter, pp. 984–997.

Sankoff, D. and Laberge, S. (1978) 'The linguistic market and the statistical explanation of variability', *Linguistic variation: Models and methods*, 239, p. 250.

Sankoff, G. (2004) 'Adolescents, young adults and the critical period: Two case studies from "Seven Up"', *Sociolinguistic variation: Critical reflections*, pp. 121–39.

Schilling-Estes, N. and Wolfram, W. (1994) 'Convergent explanation and alternative regularization patterns: Were/weren't leveling in a vernacular English variety', *Language Variation and Change*, 6(3), pp. 273–302. Available at: https://doi.org/10.1017/S0954394500001691.

Schleef, E., Meyerhoff, M. and Clark, L. (2011) 'Teenagers' acquisition of variation: A comparison of locally-born and migrant teens', in *realisation of English (ing) in Edinburgh and London' English World-Wide*. (2), pp. 206–236.

Schleef, E. (2013) 'Written surveys and questionnaires in sociolinguistics', *Research methods in sociolinguistics: A practical guide*, pp. 42–57.

Schütze, C.T. and Sprouse, J. (2013) 'Judgment data', *Research methods in linguistics*, pp. 27–50.

Schwartz, S.H. *et al.* (2012) 'Refining the theory of basic individual values', *Journal of Personality and Social Psychology*, 103, pp. 663–688. Available at: https://doi.org/10.1037/a0029393.

Schwartz, S.J. *et al.* (2010) 'Rethinking the concept of acculturation: Implications for theory and research.', *American Psychologist*, 65(4), pp. 237–251. Available at: https://doi.org/10.1037/a0019330.

Schwartz, S.J. *et al.* (2014) 'Effects of Language of Assessment on the Measurement of Acculturation: Measurement Equivalence and Cultural Frame Switching', *Psychological Assessment*, 26(1), pp. 100–114. Available at: https://doi.org/10.1037/a0034717.

Shorrocks, G. (1999) *A grammar of the dialect of the Bolton area*. Lang.

Silverstein, M. (2003) 'Indexical order and the dialectics of sociolinguistic life', *Language & Communication*, 23(3–4), pp. 193–229. Available at: https://doi.org/10.1016/S0271-5309(03)00013-2.

Smith, J., Durham, M. and Fortune, L. (2007) '"Mam, ma troosers is fa'in doon!" Community, caregiver and child in the acquisition of variation in Scottish dialect', *Language Variation and Change*, 19(1), pp. 63–99. Available at: https://doi.org/10.1017/S0954394507070044.

Smith, J. and Tagliamonte, S. (1998) '"We were all thegither . . . I think we was all thegither":was regularization in Buckie English', *World Englishes*, 17(2), pp. 105–126. Available at: https://doi.org/10.1111/1467-971X.00086.

Snell, J. (2018) 'Solidarity, stance, and class identities', *Language in Society*, 47(5), pp. 665–691.

Sorace, A. (2011) 'Pinning down the concept of "interface" in bilingualism', *Linguistic Approaches to Bilingualism*, 1(1), pp. 1–33. Available at: https://doi.org/10.1075/lab.1.1.01sor.

Sprouse, J. and Almeida, D. (2012) 'Assessing the reliability of textbook data in syntax: Adger's Core Syntax', *Journal of Linguistics*, 48(3), pp. 609–652. Available at: https://doi.org/10.1017/S0022226712000011.

Sprouse, J., Schütze, C.T. and Almeida, D. (2013) 'A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001-2010', *Lingua*, 134, pp. 219–248. Available at: https://doi.org/10.1016/j.lingua.2013.07.002.

Stychin, C. (2003) *Governing sexuality: The changing politics of citizenship and law reform*. Hart Publishing.

Sullivan, G.M. and Artino, A.R. (2013) 'Analyzing and Interpreting Data From Likert-Type Scales', *Journal of Graduate Medical Education*, 5(4), pp. 541–542. Available at: https://doi.org/10.4300/JGME-5-4-18.

Swirrl IT Ltd. (2019) 'English indices of deprivation 2019: Postcode Lookup'. Available at: https://imd-by-postcode.opendatacommunities.org/imd/2019 (Accessed: 19 July 2022).

Szapocznik, J. *et al.* (1978) 'Theory and measurement of acculturation', *Interamerican journal of Psychology*, 12(2), pp. 113–130.

Szapocznik, J., Kurtines, W.M. and Fernandez, T. (1980) 'Bicultural involvement and adjustment in Hispanic-American youths', *International Journal of intercultural relations*, 4(3–4), pp. 353–365.

Szulc, L. (2019) *Queer #PolesinUK: identity, migration and social media*, p. 29. Available at: http://eprints.lse.ac.uk/id/eprint/101767 (Accessed: 13 September 2020).

Tabouret-Keller, A. and Le Page, R.B. (1985) *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge University Press.

Tagliamonte, S. (1998) 'Was/were variation across the generations: View from the city of York', *Language Variation and Change*, 10(2), pp. 153–191. Available at: https://doi.org/10.1017/S0954394500001277.

Tagliamonte, S. and Smith, J. (1999) 'Analogical Leveling in Samaná English: The Case of Was and Were', *Journal of English Linguistics*, 27(1), pp. 8–26. Available at: https://doi.org/10.1177/00754249922004390.

Tagliamonte, S.A. (2006) *Analysing Sociolinguistic Variation*. Cambridge University Press (Key Topics in Sociolinguistics). Available at: https://doi.org/10.1017/CBO9780511801624.

Tagliamonte, S.A. (2012) *Variationist sociolinguistics: Change, observation, interpretation*. John Wiley & Sons.

Tagliamonte, S.A. and Baayen, R.H. (2012) 'Models, forests, and trees of York English: Was/were variation as a case study for statistical practice', *Language Variation and Change*, 24(2), pp. 135–178. Available at: https://doi.org/10.1017/S0954394512000129.

Tagliamonte, S.A. and D'Arcy, A. (2009) 'Peaks beyond phonology: Adolescence, incrementation, and language change', *Language*, pp. 58–108.

Timmis, I. (2010) '"Tails" of Linguistic Survival', *Applied Linguistics*, 31(3), pp. 325–345. Available at: https://doi.org/10.1093/applin/amp028.

Trudgill, P. (1974) *The social differentiation of English in Norwich*. CUP archive.

Trudgill, P. (1986) *Dialects in contact*. Oxford: Blackwell.

Umaña-Taylor, A.J., Yazedjian, A. and Bámaca-Gómez, M. (2004) 'Developing the ethnic identity scale using Eriksonian and social identity perspectives', *Identity: An International Journal of Theory and Research*, 4, pp. 9–38. Available at: https://doi.org/10.1207/S1532706XID0401_2.

Van Moere, A. (2012) 'A psycholinguistic approach to oral language assessment', *Language Testing; London*, 29(3), pp. 325–344. Available at: http://dx.doi.org.sheffield.idm.oclc.org/10.1177/0265532211424478.

Vincent, B.W. (2016) *Non-Binary Gender Identity Negotiations: Interactions with Queer Communities and Medical Practice*. PhD dissertation. University of Leeds.

Voeten, C.C. (2020) 'buildmer: Stepwise elimination and term reordering for mixed-effects regression', *R package version*, 1(6).

Watt, D. (2002) '"I don't speak with a Geordie accent, I speak, like, the Northern accent": Contact-induced levelling in the Tyneside vowel system', *Journal of Sociolinguistics*, 6(1), pp. 44–63. Available at: https://doi.org/10.1111/1467-9481.00176.

Watt, D.J.L. (1998) *Variation and change in the vowel system of Tyneside English*. PhD Thesis. Newcastle University.

Weiner, E.J. and Labov, W. (1983) 'Constraints on the agentless passive', *Journal of Linguistics*, 19(1), pp. 29–58. Available at: https://doi.org/10.1017/S0022226700007441.

Weinreich, U., Labov, W. and Herzog, M. (1968) *Empirical foundations for a theory of language change*. University of Texas Press.

Wilson, J. and Henry, A. (1998) 'Parameter Setting within a Socially Realistic Linguistics', *Language in Society*, 27(1), pp. 1–21.

Winer, C. (2022) 'Solidarity, Disdain, and the Imagined Center of the Gay Imagined Community', *Sociological Inquiry*, 92(S1), pp. 710–732. Available at: https://doi.org/10.1111/soin.12403.

Winship, C. and Mare, R.D. (1984) 'Regression Models with Ordinal Variables', *American Sociological Review*, 49(4), pp. 512–525. Available at: https://doi.org/10.2307/2095465.

Wolfram, W. and Sellers, J. (1999) 'Ethnolinguistic marking of past be in Lumbee Vernacular English', *Journal of English Linguistics*, 27(2), pp. 94–114.

Wolfram, W.A. (1969) 'A Sociolinguistic Description of Detroit Negro Speech. Urban Language Series, No. 5.'

Wong, M. (2021) 'Graphical Check for Proportional Odds Assumption [R package pomcheckr version 0.1.0]'.

Wright, J. (1892) *A grammar of the dialect of Windhill, in the West Riding of Yorkshire*. English dialect society.

www.gorilla.sc (2020) *Walkthrough: Scripting in Questionnaire Builder 1*, *Gorilla Support*. Available at: https://support.gorilla.sc/support/tools/legacy-tools/questionnaire-builder-1/questionnaire-scripting#thequestionnaire (Accessed: 4 August 2023).

Young, R. and Bayley, R. (1996) 'VARBRUL analysis for second language acquisition research', in *Second language acquisition and linguistic variation*, pp. 253–306.

Yuan, B. and Dugarova, E. (2012) 'Wh-Topicalization at the Syntax-Discourse Interface in English Speakers' L2 Chinese Grammars', *Studies in Second Language Acquisition*, 34(4), pp. 533–560.

Zane, N. and Mak, W. (2003) 'Major approaches to the measurement of acculturation among ethnic minority populations: A content analysis and an alternative empirical strategy.', in *Acculturation: Advances in theory, measurement, and applied research*. Washington, DC, US: American Psychological Association, pp. 39–60. Available at: https://doi.org/10.1037/10472-005.

Zimman, L. (2016) *Pronouns have always been political | Trans Talk*, *Medium*. Available at: https://medium.com/trans-talk/pronouns-have-always-been-political-a32c753a1539 (Accessed: 18 May 2023).

Zimman, L. (2021) 'Gender diversity and the voice', in *The Routledge Handbook of Language, Gender, and Sexuality*. Routledge.

9. Appendices

CONTENTS

9.1 Acceptability Judgement Task

9.1.1 AJT test items

The following is a list of all the test items included within the AJT, divided into the three types of morphosyntactic constructions investigated in this thesis, as well as by AJT Condition. During the task, these items were all randomised. Negative control and training items were excluded from analysis. Additionally, several test items were ultimately excluded from analysis, and these are greyed out.

Morphosyntactic construction	AJT Condition	Item
Argument movement	Raising-to-subject with the verb *seem*	*" John seemed to like Mary "*
		" She seemed angry "
		" He seemed to enjoy the movie "
		" Mary seemed to be bored "
		" Mary seems to hate John "
		" John seems to be working "
		" He seems to like Mary "
		" She seems happy "
Nonstandard past-tense BE agreement	Nonstandard *was*	*" We was at the shop yesterday "*
		" We wasn't really enjoying the film "
		" There was dogs barking at us "
		" Schools was closed for a month "
		" Was we at the park last Tuesday? "
		" Was there two girls at the park? "
		" Was schools closed for long? "
		" We was like 'What are you doing?' "
	Nonstandard *were*	*" She were in the shop for ages "*
		" Everyone were busy yesterday "
		" There were a lot of food to eat "
		" School were closed "
		" Were she alright after she fell over? "
		" Were it raining all day? "
		" Were school closed for long? "
		" She were like 'I love your shoes' "
	Nonstandard *weren't*	*" She weren't at the cinema "*
		" There weren't much to do "
		" School weren't closed on Friday "
		" Weren't she ill last week? "
		" She was lying, weren't she? "
		" Weren't it raining? "

		" It was raining, weren’t it? "
		" Weren’t school closed yesterday? "
Optional discourse-based movement	Right dislocation	*" I'm so clumsy, me "*
		" You’re so nosy, you "
		" He's so clingy, him "
		" He’s waiting for us at the park, is John "
		" She’s a great dancer, Mary "
		" They’re really friendly, dogs "
		" It’s really good, that "
		" John’s working hard, he is "
	Left dislocation	*" Him, he’s awful "*
		" Coffee, I can’t live without it "
		" That new Netflix show, John said Mary told him it’s really good "
		" The man in the yellow hat, he was running away "
		" Those shoes with the red laces, I wore them yesterday "
		" This hat of yours, it’s so warm "
		" An honest politician, I’d like to meet one "
		" My dad, I love visiting him and my mum on weekends "
	Topicalisation	*" These chocolate biscuits, I could eat for hours "*
		" This red car, I'd love to buy "
		" That outfit, I’m not too sure about "
		" Those rules, I don't agree with "
		" For school, I read a lot of books "
		" On the way to school, I saw an ambulance "
		" By the front door, I left my coat "
		" Behind my house, there’s a park "
Negative control		*" I really like themself "*
		" He ranned a long way "
		" I will flew to Japan tomorrow "
		" He saw they leave "
		" You am hungry "
		" Him ate the sandwich "
		" Man the walked a dog "
		" I hated movie the "
Training		*" I don’t want no dinner "*
		" There’s snakes at the zoo "
		" I like watching TV "
		" John likes Mary "
		" I going the park "
		" I studies exam "

9.1.2 AJT test items (pilot study)

Morphosyntactic construction	AJT Condition	Item
Argument movement	Raising-to-subject with the verb *seem*	*John seems to like Mary*
		She seems to be happy
		John seems to Mary to be annoyed
		Mary seems to John to be happy
		John believed it to be raining
		Mary believed John to be studying
		Mary seems easy to talk to
		John seems hard to please
Nonstandard past-tense BE agreement	Nonstandard *was*	*We was at the shop yesterday*
		We wasn't really enjoying the film
		There was dogs barking at us
		Schools was closed for a month
		Was we at the park last Tuesday?
		Was there two girls at the park?
		Was schools closed for long?
		We was like 'What are you doing?'
	Nonstandard *were*	*She were in the shop for ages*
		Everyone were busy yesterday
		There were a lot of food to eat
		School were closed
		Were she alright after she fell over?
		Were it raining all day?
		Were school closed for long?
		She were like 'I love your shoes'
	Nonstandard *weren't*	*She weren't at the cinema*
		There weren't much to do
		School weren't closed on Friday
		Weren't she ill last week?
		She was lying, weren't she?
		Weren't it raining?
		It was raining, weren't it?
		Weren't school closed yesterday?
	Right dislocation	*I'm so clumsy, me*
		You're so nosy, you
		He's so clingy, him
		He's waiting for us at the park, is John
		She's a great dancer, Mary

Optional discourse-based movement		*They're really friendly, dogs*
		It's really good, that
		John's working hard, he is
	Left dislocation	*Him, he's awful*
		Coffee, I can't live without it
		That new Netflix show, John said Mary told him it's really good
		The man in the yellow hat, he was running away
		Those shoes with the red laces, I wore them yesterday
		This hat of yours, it's so warm
		An honest politician, I'd like to meet one
		My dad, I love visiting him and my mum on weekends
	Topicalisation	*The dog, I played with until I got tired*
		A house, I'd love to buy in the future
		That outfit, I'm not too sure about
		Him, I can't take seriously
		For school, I read a lot of books
		On the way to school, I saw an ambulance
		By the front door, I left my coat
		Behind my house, there's a park
Negative control		*I really like themself*
		He ranned a long way
		I will flew to Japan tomorrow
		He saw they leave
		You am hungry
		Him ate the sandwich
		Man the walked a dog
		I hated movie the

9.2 Versant Test

9.2.1 Versant Test - Instructions

9.2.2 Versant Test - Sample Score Report

SCORE REPORT

VERSANT™

Versant English Test

Test Identification Number: 12345678
Test Completion Date: January 1, 2019
Test Completion Time: 1:23 PM (UTC)

OVERALL SCORE
48

SKILL AREA	SCORE	20 30 40 50 60 70 80
Overall Score	48	
Sentence Mastery	48	
Vocabulary	53	
Fluency	48	
Pronunciation	45	

	DESCRIPTION
Overall	The Overall Score of the test represents the ability to understand spoken English and speak it intelligibly at a native-like conversational pace on everyday topics. Scores are based on a weighted combination of four diagnostic subscores. Scores are reported in the range from 20 to 80.
Candidate's Capabilities	Candidate can handle many utterances using a variety of words and structures, and can follow and sometimes participate in a native-paced conversation. Pronunciation is mostly intelligible; candidate can express some composite information on familiar topics to a cooperative listener.

For more information, visit us online at www.VersantTest.com

PEARSON

ORDINATE TECHNOLOGY VERSANT APPROVED LANGUAGE TESTING

SCORE REPORT

VERSANT™

Test Identification Number: 12345678

EXPLANATION OF SUBSKILL SCORES

SKILL AREA	UNDERSTANDING THE SKILLS	CURRENT CAPABILITIES
Sentence Mastery	Sentence Mastery reflects the ability to understand, recall and produce English phrases and clauses in complete sentences. Performance depends on accurate syntactic processing and appropriate usage of words, phrases and clauses in meaningful sentence structures.	Candidate can understand, recall and produce many English phrases and clauses in sentence context. Candidate produces a range of meaningful sentences.
Vocabulary	Vocabulary reflects the ability to understand common everyday words spoken in sentence context and to produce such words as needed. Performance depends on familiarity with the form and meaning of everyday words and their use in connected speech.	Candidate usually understands and can produce everyday English words when they are used in clear speech.
Fluency	Fluency reflects the rhythm, phrasing and timing evident in constructing, reading and repeating sentences.	Candidate speaks with uneven or staccato pacing, although speech may contain some smooth runs containing several words. Frequent obvious pauses result in an irregular speech rate and some disconnected phrases.
Pronunciation	Pronunciation reflects the ability to produce consonants, vowels and stress in a native-like manner in sentence context. Performance depends on knowledge of the phonological structure of everyday words.	Candidate consistently pronounces certain consonants and vowels in a non-native manner. Stress may be placed incorrectly in some words, or stress placement may be ambiguous. Speech is mostly intelligible, but many listeners will need to adjust to the accent.

SCORE REPORT

VERSANT™

Test Identification Number: 12345678

DETAILED DESCRIPTION OF LANGUAGE CAPABILITIES

This section allows users to form an idea of the spoken language tasks and interactions that average candidates scoring above a certain Versant English Test score are likely to be able to do. This information is based on the results of a study in which experienced raters listened to and rated candidates responding to the variety of prompts implemented in the Versant English Test, including responses to open questions that ask for personal opinions and views on different issues.

The score interpretations are based on large samples of speakers representing 22 languages from East Asia, the Middle East, South America, and from both Eastern and Western European countries. Male and female speakers were equally represented, and the age of candidates ranged from 12 to 65 with an average age of 25.

The section refers to the Common European Framework of Reference for Languages (Council of Europe, 2001). Descriptors from this publication were used by the raters while listening to candidates' performances and deciding on their levels of competence.

The information presented in this section is based on the average ratings of large samples of candidates. It refers to the average candidate; individual candidates may perform at a slightly higher or lower level than indicated in this guide. (See "Current Capabilities" on pages 1 and 2 for your personal results.)

SKILL AREA	A CANDIDATE WITH THIS OVERALL SCORE TYPICALLY:
Listening	• Can understand straightforward factual information about common everyday or job-related topics, identifying both general messages and specific details, provided speech is clearly articulated in a generally familiar accent. • Can generally follow the main points of extended discussion around him/her, provided speech is clearly articulated in standard dialect. • Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.
Speak Production	• Can give straightforward descriptions on a variety of familiar subjects within his/her field of interest. • Can describe dreams, hopes, and ambitions. • Can describe events, real or imagined. • Can briefly give reasons and explanations for opinions, plans, and actions.
Spoken Interaction	• Can communicate with some confidence on familiar routine and nonroutine matters related to his/her interests and professional field. • Can express thoughts on more abstract, cultural topics such as films, books, music, etc. • Can follow clearly articulated speech directed at him/her in everyday conversation, though will sometimes have to ask for repetition of particular words and phrases. • Can enter unprepared into conversations on familiar topics. • Can follow much of what is said around him/her on general topics and topics related to his/her field provided interlocutors avoid very idiomatic usage and articulate clearly. • Can follow what is said, though he/she may occasionally have to ask for repetition or clarification if the other people's talk is rapid or extended.

SCORE REPORT

VERSANT™

Test Identification Number: 12345678

SKILL AREA	A CANDIDATE WITH THIS OVERALL SCORE TYPICALLY:
Language Quality	• Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events, but lexical limitations cause repetition and even difficulty with formulation at times. • Shows good control of elementary vocabulary but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations. • Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur. • Can link a series of shorter, discrete simple elements into a connected, linear sequence of points. • Can express the main point he/she wants to make comprehensibly.
Strategies & Skills	• Can identify unfamiliar words from the context on topics related to his/her field and interests. • Can repeat back part of what someone has said to confirm mutual understanding and help keep the development of ideas on course. • Can ask someone to clarify or elaborate what he/she has just said. • Can work out how to communicate the main point(s) he/she wants to get across, exploiting any resources available and limiting the message to what he/she can recall or find the means to express. • Can convey meaning by qualifying a word meaning something similar (e.g., a truck for people = bus).

TO IMPROVE, A CANDIDATE AT THIS LEVEL SHOULD:

- Practice listening to conversations or presentations likely to be encountered in social, professional, or academic life and identifying speaker viewpoints and attitudes as well as the information content.
- Practice keeping up with language spoken at a normal speed by watching and summarizing TV news and current affairs programs, documentaries, live interviews, talk shows, plays, and films.
- Practice providing clear, detailed descriptions on a wide range of subjects related to your field of interest.
- Practice explaining a viewpoint on a topical issue giving the advantages and disadvantages of various options.
- Practice delivering announcements or talks on general topics, departing spontaneously from the prepared text as needed and following up on interesting points raised by friends or classmates.
- Practice communicating spontaneously with good grammatical control, being careful to adopt a level of formality appropriate to the circumstances.
- Actively participate in conversations to practice language use on a wide range of general, academic, vocational, or leisure topics.
- Practice conveying degrees of emotion and highlighting the personal significance of events and experiences.
- Practice identifying arguments supporting and opposing points of view while listening to an animated discussion.
- Expand your repertoire of vocabulary items and phrases to be able to avoid frequent repetition, broadening your range of stock phrases (e.g., "I mean...", "That is to say...", "Let me think about that...") to gain time and keep the turn while formulating what to say.
- Focus on developing a clear pronunciation and intonation.
- Practice using a variety of linking words efficiently to mark clearly the relationships between ideas.

SCORE REPORT

VERSANT™

Test Identification Number: 12345678

RELATIONSHIP TO OTHER SCORES AND LEVELS

Research has been conducted to explore how a Versant English Test overall score relates to other scales that measure or describe language proficiency. Note that the corresponding scores or levels provided are based on the relationships observed in our studies; the information does not guarantee a score on other tests or in other evaluations.

TEST/SCALE	SCORE/RANGE
GSE	Corresponding GSE score: **45** The Global Scale of English (GSE) is reported on a scale from 10 to 90. The GSE has been aligned to the Common European Framework of Reference for Languages and provides a common, empirically validated equivalence for Pearson tests measuring English ability.
CEFR	Corresponding level in the Common European Framework of Reference (CEFR): **B1 – Independent User** CEF-R global-level descriptor: Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise while traveling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of a personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
TOEFL Speaking	Corresponding TOEFL iBT Speaking score range: **16 - 21** The TOEFL iBT Speaking subscore is drawn from performance on the Speaking section of the TOEFL. Candidates perform six tasks where they receive written and aural input and respond in a single, long turn. TOEFL iBT Speaking responses receive a 0 - 4 rating and are then converted to a 0 - 30 scale.
TOEFL Total	Corresponding TOEFL iBT Total score range: **61 - 89** The TOEFL iBT Total score is drawn from performance on four sections (Reading, Listening, Speaking, and Writing). It is claimed that the TOEFL assesses the academic language domain drawing on authentic university materials and tasks. TOEFL iBT Total scores are reported on a 0 - 120 scale.
TOEIC	Corresponding TOEIC score range: **719 - 762** The TOEIC Listening and Reading test is a paper-and-pencil, multiple-choice assessment that elicits responses in two sections (Listening and Reading). It is claimed that the TOEIC measures the everyday English skills of people working in an international environment. TOEIC total scores are reported on a 10 - 990 scale.

NOTE: The Versant English Test and other tests/scales address different constructs of language proficiency. Therefore, predictions are approximate. More information about these concordance studies is available upon request.

9.3 Acculturation Surveys

9.3.1 English Acculturation Survey

English Culture

In the following section, you will be asked about your English language use, attitudes, and lifestyle.

Please answer the questions as if you were **under normal circumstances**, and not in quarantine/lockdown.

Language use

Living in England, how likely are you to use the **English** language in the following settings?

If you only know the English language, select 'entirely likely' (7) for all

At your workplace or place of education

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Within your household

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

With an interest group (club/hobby activity)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

With your peer group (main group of friends)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Living in England, how likely are you to use the **English** language when performing the following tasks?

If you only know the English language, select 'entirely likely' (7) for all

Calculating/counting

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Note-taking (for personal use/synthesising information/learning)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Writing out a to-do list for work, or a personal action plan

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Writing shopping lists or informal reminders

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Attitudes and Identity

How **important** to you are the following things?

Being perceived as an English/British person or native speaker of English

Not at all important	1	2	3	4	5	6	7	**Extremely important**

Interacting with British people

Not at all important	1	2	3	4	5	6	7	**Extremely important**

What is your intended length of residence in England?

I want to leave England as soon as possible	1	2	3	4	5	6	7	**I never want to leave England**

To what extent do you consider yourself English/British?

Not at all	1	2	3	4	5	6	7	**Completely**

In England, to what extent do you consider yourself a member of the local community in which you live?

Not at all	1	2	3	4	5	6	7	**Completely**

To what extent do you think your personal values align with typical British values?

(according to your personal interpretation of what 'British values' are)

Not at all	1	2	3	4	5	6	7	**Completely**

To what extent do you feel a sense of contentness and 'fitting in' in England?

Not at all	1	2	3	4	5	6	7	**Completely**

Lifestyle

Living in England, how often do you typically interact with the following types of people, either in person or remotely?

Key:
1 = Never/ Less than once a year
2 = Several times a year
3 = Several times a month
4 = Once a week
5 = Several times a week
6 = Every day
7 = Several times a day

Someone who was born in Britain

Never/ Less than once a year	1	2	3	4	5	6	7	**Several times a day**

Someone who was born in the English county in which you currently live
(county = larger geographical region, eg. South Yorkshire, Greater London, Kent, etc.)

Never/ Less than once a year	1	2	3	4	5	6	7	**Several times a day**

Living in England, how likely are you to do the following?

Note: *examples are only provided to demonstrate what is meant, so don't worry if you do not partake in all/any of the given examples*

Participate in typical British traditions/customs
(eg. Christmas day, Boxing day, Bonfire night, etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Partake in typically British cultural behaviours *(eg. queuing)* or activities that are popular in Britain
(eg. going to the local pub; following British sports or sports teams; etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Consume British meals, snacks and drinks
(eg. fish and chips; tea with milk; McVitie's biscuits; etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Consume British media
(eg. TV, films, books, magazines, radio, etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Participate regularly in a group activity with British people, either in person or digitally
(eg. study group, sports team, book club, special interest group, etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Be close personal friends with a British person

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

9.3.2 Polish Acculturation Survey

Polish Culture

In the following section, you will be asked about your Polish language use, attitudes, and lifestyle.

Please answer the questions as if you were **under normal circumstances**, and not in quarantine/lockdown.

Language use

Living in England, how likely are you to use the **Polish** language in the following settings?

At your workplace or place of education

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Within your household

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

With an interest group (club/hobby activity)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

With your peer group (main group of friends)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Living in England, how likely are you to use the **Polish** language when performing the following tasks?

Calculating/counting

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Note-taking (for personal use/synthesising information/learning)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Writing out a to-do list for work, or a personal action plan

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Writing shopping lists or informal reminders

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Attitudes and Identity

How **important** to you are the following things?

Being perceived as a Polish person or native speaker of Polish

Not at all important	1	2	3	4	5	6	7	**Extremely important**

Interacting with Polish people

Not at all important	1	2	3	4	5	6	7	**Extremely important**

When (if at all) do you intend to return to Poland?

(to live there, not just for a visit/holiday)

I never want to return to Poland	1	2	3	4	5	6	7	**I want to return to Poland as soon as possible**

To what extent do you consider yourself Polish?

Not at all	1	2	3	4	5	6	7	**Completely**

In Poland, to what extent did you consider yourself a member of the local community in which you lived?

*(If you have lived in several places in Poland, answer for the place in which you lived the **longest**)*

Not at all	1	2	3	4	5	6	7	**Completely**

To what extent do you think your personal values align with typical Polish values?

(according to your personal interpretation of what 'Polish values' are)

Not at all	1	2	3	4	5	6	7	**Completely**

When you lived in Poland, to what extent did you feel a sense of contentness and 'fitting in' there?

Not at all	1	2	3	4	5	6	7	**Completely**

Lifestyle

Living in England, how often do you typically interact with the following types of people, either in person or remotely?

Key:
1 = Never/ Less than once a year
2 = Several times a year
3 = Several times a month
4 = Once a week
5 = Several times a week
6 = Every day
7 = Several times a day

Someone who was born in Poland

Never/ Less than once a year	1	2	3	4	5	6	7	**Several times a day**

Someone who was born in the Polish county ('powiat') in which you lived
*(If you have lived in several counties, answer for the county in which you lived the **longest**)*

Never/ Less than once a year	1	2	3	4	5	6	7	**Several times a day**

Living in England, how likely are you to do the following?

Note: *examples are only provided to demonstrate what is meant, so don't worry if you do not partake in all/any of the given examples*

Participate in typical Polish traditions/customs
(eg. Wigilia, Andrzejki, celebrating 'imieniny', etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Partake in typically Polish cultural behaviours
(eg. Thanking company after a meal even if nobody at the table cooked the food)
or activities that are popular in Poland
(eg. following Polish sports teams; cooking meals for your household from scratch; etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Consume Polish meals, snacks and drinks
(eg. Pierogi; 'open' sandwiches; Princessa chocolate; etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Consume Polish media
(eg. TV, films, books, magazines, radio, etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Participate regularly in a group activity with Polish people, either in person or digitally
(eg. study group, sports team, book club, special interest group, etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Be close personal friends with a Polish person

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

9.3.3 Acculturation Survey feedback

9.3.3.1 Feedback (*High English acculturation*)

Your results

'**Acculturation**' is a term used to describe the extent to which people adapt to the culture(s) they are exposed to. Typically, it is used in reference to people being exposed to cultures different from their native culture (such as migrants).

This study explores the modes of acculturation undergone by Polish-born adults living in England.

Your survey responses about your **English language use, attitudes & identity, and lifestyle**, indicate:

high English acculturation

This suggests that you very strongly affiliate yourself with social, psychological, and cultural markers associated with English culture.

Some possible explanations for your score are listed below:

- You likely associate with most aspects of English culture
- Your English or British identity might be personally significant to you
- You may orient towards English or British social networks, or have many influences from English culture

Your results

'**Acculturation**' is a term used to describe the extent to which people adapt to the culture(s) they are exposed to. Typically, it is used in reference to people being exposed to cultures different from their native culture (such as migrants).

This study explores the modes of acculturation undergone by Polish-born adults living in England.

Your survey responses about your **English language use, attitudes & identity, and lifestyle**, indicate:

moderate English acculturation

This suggests that you somewhat, but not entirely, affiliate with social, psychological, and cultural markers associated with English culture.

Some possible explanations for your score are listed below:

- Your English or British identity might not be of utmost personal significance to you
- You may orient towards more multicultural social networks, or have some other influences from non-English cultures
- You may not associate with certain aspects of English culture (eg. due to political factors, etc.)

Your results

'**Acculturation**' is a term used to describe the extent to which people adapt to the culture(s) they are exposed to. Typically, it is used in reference to people being exposed to cultures different from their native culture (such as migrants).

This study explores the modes of acculturation undergone by Polish-born adults living in England.

Your survey responses about your **English language use, attitudes & identity, and lifestyle**, indicate:

low English acculturation

This suggests that you do not affiliate very much with social, psychological, and cultural markers associated with English culture.

Some possible explanations for your score are listed below:

- Your English or British identity might not be of great personal significance to you
- You may orient towards multicultural or Polish social networks, or have other strong influences from non-English cultures
- You may not associate with a lot of aspects of English or British culture (eg. due to political factors, etc.)

9.3.3.4 Feedback (*High Polish acculturation*)

Your results

'**Acculturation**' is a term used to describe the extent to which people adapt to the culture(s) they are exposed to. Typically, it is used in reference to people being exposed to cultures different from their native culture (such as migrants).

This study explores the modes of acculturation undergone by Polish-born adults living in England.

Your survey responses about your **Polish language use, attitudes & identity, and lifestyle**, indicate:

high Polish acculturation

This suggests that you very strongly affiliate yourself with social, psychological, and cultural markers associated with Polish culture.

Some possible explanations for your score are listed below:

- You likely associate with most aspects of Polish culture
- Your Polish identity might be very significant to you
- You may orient towards Polish social networks in England, or have many continuing influences from Polish culture

Click 'Next' to find out about what modes of acculturation may apply to you.

Your results

'**Acculturation**' is a term used to describe the extent to which people adapt to the culture(s) they are exposed to. Typically, it is used in reference to people being exposed to cultures different from their native culture (such as migrants).

This study explores the modes of acculturation undergone by Polish-born adults living in England.

Your survey responses about your **Polish language use, attitudes & identity, and lifestyle**, indicate:

moderate Polish acculturation

This suggests that you somewhat, but not entirely, affiliate with social, psychological, and cultural markers associated with Polish culture.

Some possible explanations for your score are listed below:

- Although you likely have some continuing influence from Polish culture, you may orient towards more multicultural social networks in England, or have some other influences from non-Polish cultures
- Your Polish identity might not be of utmost personal significance to you
- You may not associate with certain aspects of Polish culture (eg. due to political factors, etc.)

Click 'Next' to find out about what modes of acculturation may apply to you.

Your results

'**Acculturation**' is a term used to describe the extent to which people adapt to the culture(s) they are exposed to. Typically, it is used in reference to people being exposed to cultures different from their native culture (such as migrants).

This study explores the modes of acculturation undergone by Polish-born adults living in England.

Your survey responses about your **Polish language use, attitudes & identity, and lifestyle**, indicate:

low Polish acculturation

This suggests that you do not affiliate very much with social, psychological, and cultural markers associated with Polish culture.

Some possible explanations for your score are listed below:

- Your Polish identity might not be of great personal significance to you
- You may orient towards multicultural or English/British social networks, or have other strong influences from non-Polish cultures
- You may not associate with a lot of aspects of Polish culture (eg. due to political factors, etc.)

Click 'Next' to find out about what modes of acculturation may apply to you.

Your results

There are different **modes of acculturation** which can be categorised by comparing between your English acculturation and your Polish acculturation scores.

A few of these are described below - see which one (if any) fits you the best! (or just scroll to the bottom and click 'Next').

- **High acculturation** to both English and Polish culture:

Integration

In this mode of acculturation, individuals have a strongly developed bicultural identity, meaning they have incorporated the language, behaviours, customs, and values, of both their original culture and their new culture into their lives.

- **Low acculturation** to both English and Polish culture:

Marginalisation

In this mode of acculturation, individuals feel distanced from the language, behaviours, customs, and values of both their original culture as well as their new culture. This can be a temporary phase, especially if the individual is still adapting to life in a new culture.

- **Higher Polish acculturation than English acculturation**:

Separation

In this mode of acculturation, individuals have a strong attachment to their original culture and remain oriented towards its language, behaviours, customs, and values, preferring not to acculturate very much to the new culture.

- **Lower Polish acculturation than English acculturation**:

Assimilation

In this mode of acculturation, individuals identify more closely with their new culture than their original culture, and adapt their language, behaviours, and values, to that of the new culture.

How accurate do you consider your **English** acculturation result to be?

Completely inaccurate	1	2	3	4	5	6	7	**Completely accurate**

How accurate do you consider your **Polish** acculturation result to be?

Completely inaccurate	1	2	3	4	5	6	7	**Completely accurate**

If you have any feedback regarding your acculturation scores, please type this below:

9.4 LGBTQ+ Community Involvement Survey

9.4.1 LGBTQ+ Community Involvement Survey (English-born)

LGBTQ+ identity

Because you indicated that you identify as LGBTQ+, please answer the following questions

In general, how likely are you to do the following?

Interact with other LGBTQ+ people

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Attend national/local LGBTQ+ pride events or meetups

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Access LGBTQ+-related support
(eg. peer support, support groups, therapy)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Access LGBTQ+ digital spaces
(eg. forums, dating apps, online pages/groups)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Consume media that is aimed at an LGBTQ+ audience
(eg. music, podcasts, films, TV, books, magazines, etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Engage in LGBTQ+ activism or actively help/promote the LGBTQ+ community
(eg. belong to LGBTQ+ activist groups, attend protests, share articles/posts, sign/share petitions)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Visit LGBTQ+-associated venues
(eg. clubs, bars, cafés, etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

In general, how likely are you to be open about your LGBTQ+ identity with the following types of people?

If you do not know anyone from any of the listed groups, answer hypothetically

English/British friends

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

non-British friends

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Polish friends

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

English/British strangers or acquaintances

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

non-British strangers or acquaintances

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Polish strangers or acquaintances

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

When interacting with other LGBTQ+ people, either in person or remotely, how likely are you to use the following languages?

(If you do not interact with other LGBTQ+ people, leave this section blank)

English

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Polish

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

9.4.2 LGBTQ+ Community Involvement Survey (Polish-born)

LGBTQ+ identity

Because you indicated that you identify as LGBTQ+, please answer the following questions

In general, how likely are you to do the following?

Interact with other LGBTQ+ people

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Attend national/local LGBTQ+ pride events or meetups

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Access LGBTQ+-related support
(eg. peer support, support groups, therapy)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Access LGBTQ+ digital spaces
(eg. forums, dating apps, online pages/groups)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Consume media that is aimed at an LGBTQ+ audience
(eg. music, podcasts, films, TV, books, magazines, etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Engage in LGBTQ+ activism or actively help/promote the LGBTQ+ community
(eg. belong to LGBTQ+ activist groups, attend protests, share articles/posts, sign/share petitions)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Visit LGBTQ+-associated venues
(eg. clubs, bars, cafés, etc.)

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

In general, how likely are you to be open about your LGBTQ+ identity with the following types of people?

If you do not know anyone from any of the listed groups, answer hypothetically

English/British friends

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

non-British friends

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Polish friends

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

English/British strangers or acquaintances

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

non-British strangers or acquaintances

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Polish strangers or acquaintances

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

When interacting with other LGBTQ+ people, either in person or remotely, how likely are you to use the following languages?

(If you do not interact with other LGBTQ+ people, leave this section blank)

English

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

Polish

entirely unlikely	1	2	3	4	5	6	7	**entirely likely**

9.4.3 LGBTQ+ Community Involvement Survey feedback

9.4.4 Feedback (*High involvement)*

Your result

This research explores the modes of acculturation that LGBTQ+ people undergo, and how this interacts with their language use.

As such, an **LGBTQ+ community-involvement score** has been calculated based on your answers to the previous survey.

Your results indicate:

high LGBTQ+ community-involvement

This suggests that you strongly involve yourself with the LGBTQ+ community, and this is reflected in your behaviours, and also in your openness about your LGBTQ+ identity.

How accurate do you consider this result to be?

Completely inaccurate	1	2	3	4	5	6	7	**Completely accurate**

If you have any feedback regarding your LGBTQ+ community-involvement score, please type this below:

Your result

This research explores the modes of acculturation that LGBTQ+ people undergo, and how this interacts with their language use.

As such, an **LGBTQ+ community-involvement score** has been calculated based on your answers to the previous survey.

Your results indicate:

moderate LGBTQ+ community-involvement

This suggests that you somewhat involve yourself with the LGBTQ+ community, and this is reflected in your behaviours. You may be more open about your LGBTQ+ identity with some people than others.

How accurate do you consider this result to be?

Completely inaccurate	1	2	3	4	5	6	7	**Completely accurate**

If you have any feedback regarding your LGBTQ+ community-involvement score, please type this below:

Your result

This research explores the modes of acculturation that LGBTQ+ people undergo, and how this interacts with their language use.

As such, an **LGBTQ+ community-involvement score** has been calculated based on your answers to the previous survey.

Your results indicate:

low LGBTQ+ community-involvement

This suggests that you only slightly (or not at all) involve yourself with the LGBTQ+ community, and this is reflected in your behaviours. You might not be open about your LGBTQ+ identity with everyone in your life, or just prefer to involve yourself with other communities/groups and focus on other aspects of your life.

How accurate do you consider this result to be?

Completely inaccurate	1	2	3	4	5	6	7	**Completely accurate**

If you have any feedback regarding your LGBTQ+ community-involvement score, please type this below:

9.5 Demographic Information Form

9.5.1 Demographic Information Form (English-born)

About you

In order to further inform this research, please answer the following questions

How long (to the nearest year) have you lived in your current county in England?
(county = larger geographical region, eg. South Yorkshire, Greater London, Kent, etc.)

Please Select...

If you have lived in any other counties in England for periods of a year or longer, please list these below:

How do you feel about Brexit?
(4 = neutral/no opinion)

Completely disapprove	1	2	3	4	5	6	7	**Completely approve**

How religious do you consider yourself to be?

Not at all religious	1	2	3	4	5	6	7	**Extremely religious**

Linguistic background

If you know any languages other than English, please list these below:

If you have ever studied **Linguistics (as a science, as opposed to language learning)** please provide information about this below:

Tick below if you consider yourself to have any **Specific Learning Difficulties or disabilities** that might affect your language skills
(eg. dyslexia, ADHD, dysgraphia, etc.)

☐ Tick here if you have any difficulties/disabilities that might affect your language skills

Demographic background

How old are you?

Select age bracket

What is your gender?

Please Select...

What is your sexuality?

Please Select...

What is your ethnicity?

Please Select...

Socioeconomic background

What is your current UK postcode?
(This will only be used to determine the community demographics in your area of residence. Your precise address will not be traceable from this information)

What is the highest education qualification you have completed?

Please Select...

What is your current official job title?

(If you are currently unemployed, answer with the last job role you held. If you are a student or have never been employed, please state)

Note: If you are unsure of your official job title, you can search for it using the 'Occupation Title' field here *(link opens in new tab)*

Optional: **Enter your email if you are interested in being contacted for future research**
(Even if you do provide your email, this will be entirely non-committal)

user@example.co.uk

9.5.2 Demographic Information Form (Polish-born)

About you

In order to further inform this research, please answer the following questions

How old were you when you first moved to England?

Select age bracket

What were your main reasons for moving to England?
(Select up to 3)

- [ ] For study/education
- [ ] For work/employment
- [ ] To join family/friends/partner already living in England
- [ ] Moved with parent(s) as a child
- [ ] Due to cultural differences between Poland and England
- [ ] Other (please specify)

How long (to the nearest year) have you lived in England?

Please Select...

How long (to the nearest year) have you lived in your current county in England?
(county = larger geographical region, eg. South Yorkshire, Greater London, Kent, etc.)

Please Select...

If you have lived in any other counties in England for periods of a year or longer, please list these below:

How do you feel about Brexit?
(4 = neutral/no opinion)

Completely disapprove	1	2	3	4	5	6	7	**Completely approve**

How religious do you consider yourself to be?

Not religious at all	1	2	3	4	5	6	7	**Extremely religious**

Linguistic background

Roughly how many years ago did you begin learning the English language?

Please Select...

What formal English language tuition have you been exposed to?
(tick all that apply)

- [ ] English language lessons as part of the school curriculum
- [ ] English language courses at a language school
- [ ] Self-motivated formal English language study (eg. through textbooks/online courses)
- [ ] I have had no formal English instruction
- [ ] Other (please specify)

If you know any languages other than Polish and English, please list these below:

If you have ever studied **Linguistics (as a science, as opposed to language learning)** please provide information about this below:

Tick below if you consider yourself to have any **Specific Learning Difficulties or disabilities** that might affect your language skills
(eg. dyslexia, ADHD, dysgraphia, etc.)

- [ ] Tick here if you have any difficulties/disabilities that might affect your language skills

Demographic background

How old are you?

Select age bracket

What is your gender?

Please Select...

What is your sexuality?

Please Select...

What is your ethnicity?

Please Select...

Socioeconomic background

What is your current UK postcode?
(This will only be used to determine the community demographics in your area of residence. Your precise address will not be traceable from this information)

What is the highest education qualification you have completed?

Please Select...

What is your current official job title?

(If you are currently unemployed, answer with the last job role you held. If you are a student or have never been employed, please state)

Note: If you are unsure of your official job title, you can search for it using the 'Occupation Title' field here *(link opens in new tab)*

English Speaking Test

The final part of this study is the **Versant English Speaking Test**, which you will complete at a later time.

This will take **15 minutes** and is done electronically. It will involve simple speaking and reading exercises.

You will not be judged on how 'well' you know English. The results will only be used to give a general idea of your English proficiency.

Benefits for you

- This is a widely accredited test. You will get an official Score Report which you can show potential employers, and you can add your Versant test score to your CV/job application. The report equates your Versant score to other common language testing scales (eg. GSE and CEFR)
- You will get feedback about different aspects of your English language use - Pronunciation, Fluency, Vocabulary, and Sentence mastery.
- You may find it personally interesting to learn how your English language use is scored (and you could share this study with Polish-born friends to compare scores!)
- You will get access to this test for free (it normally costs $34)

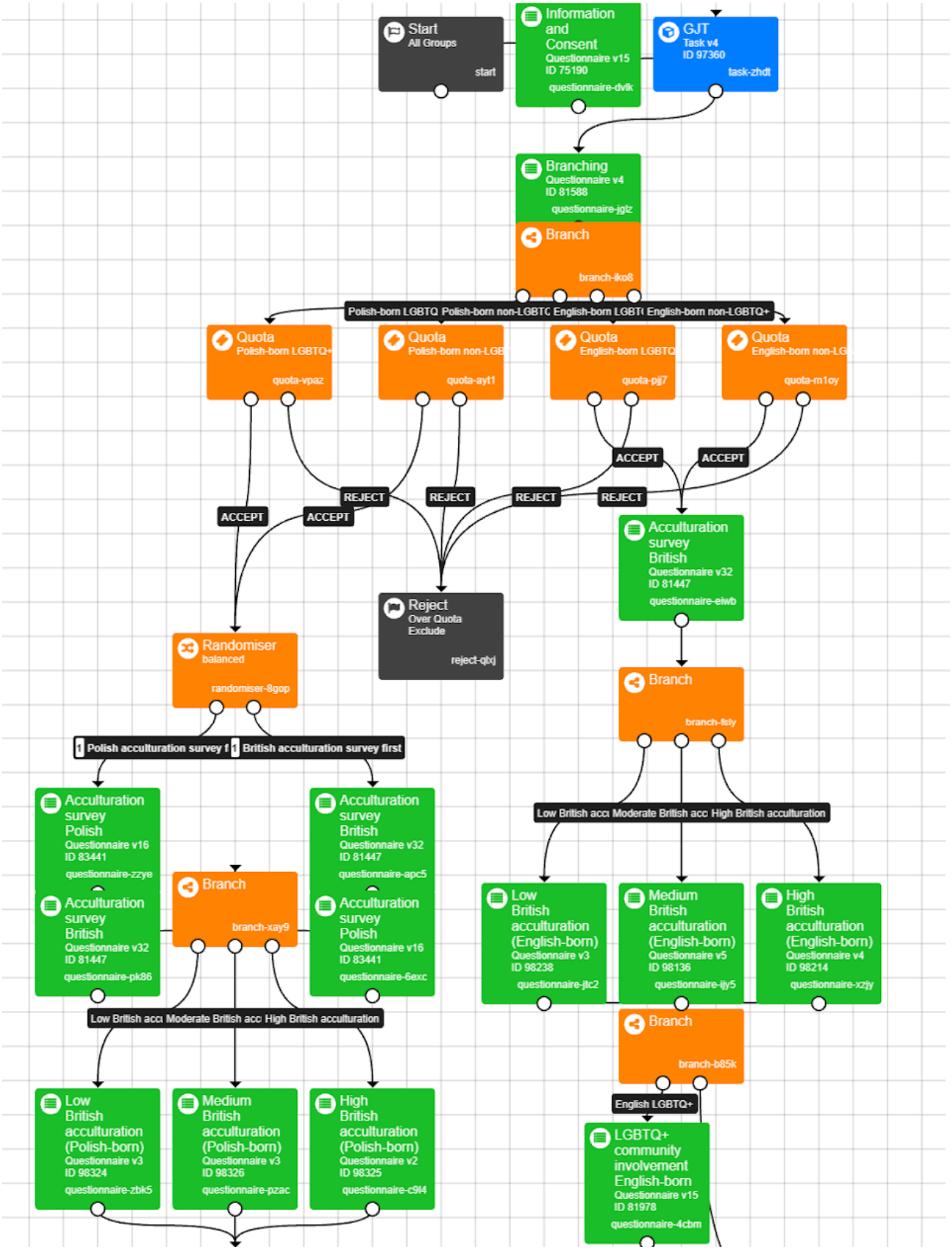
Completing the test

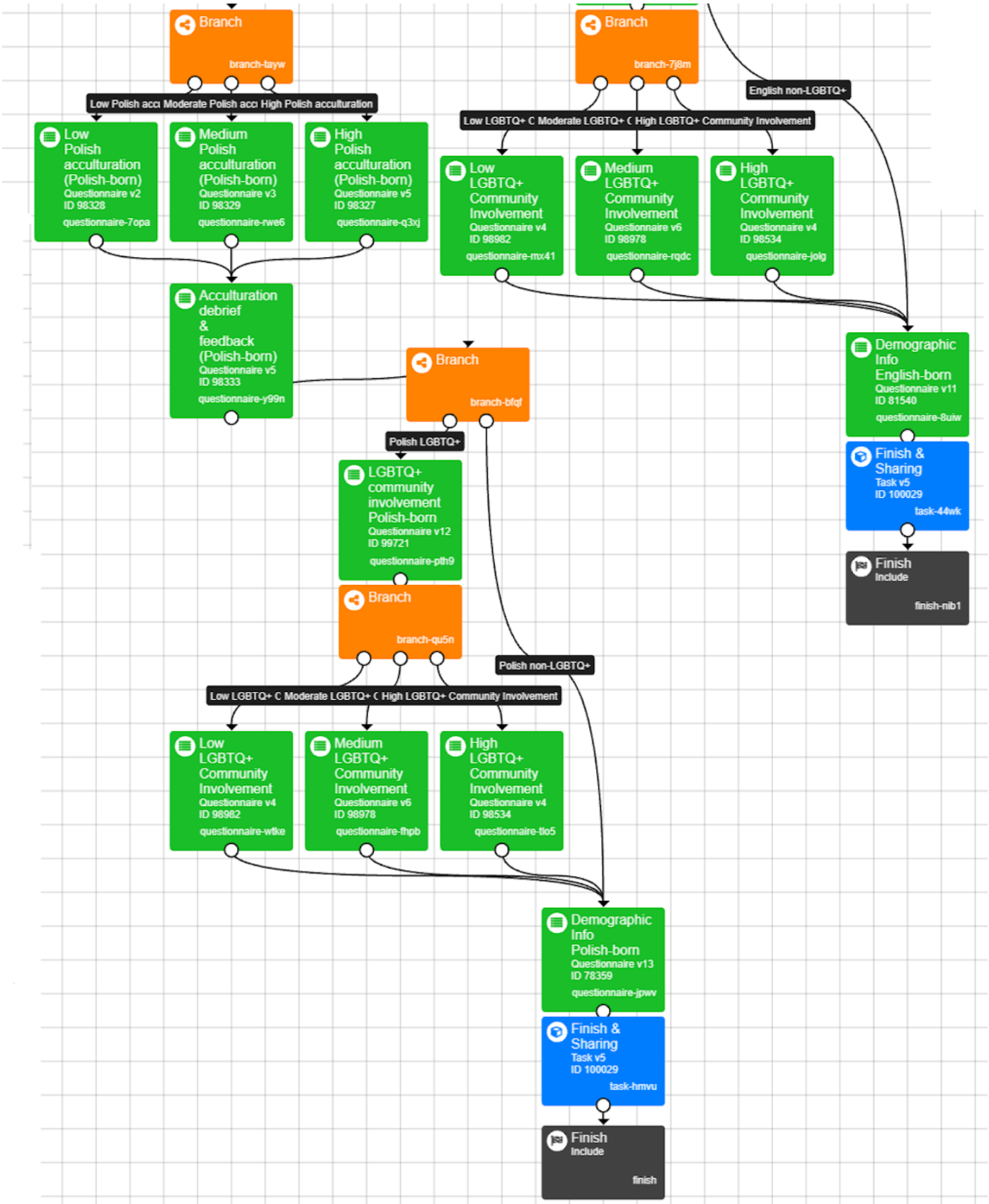
Please provide your email address and, within 24 hours, you will receive instructions for completing the test. You will complete the test in your own time.

user@example.co.uk

☐ (Optional) Tick here if you are also interested in being contacted regarding future follow-up research. This will be totally non-commital.

9.6 Gorilla Experiment Design




Branch
branch-tayw
Low Polish acci
Moderate Polish acci
High Polish acculturation
Low Polish acculturation (Polish-born)
Questionnaire v2
ID 98328
questionnaire-7opa
Medium Polish acculturation (Polish-born)
Questionnaire v3
ID 98329
questionnaire-rwe6
High Polish acculturation (Polish-born)
Questionnaire v5
ID 98327
questionnaire-q3xj
Acculturation debrief & feedback (Polish-born)
Questionnaire v5
ID 98333
questionnaire-y99n
Branch
branch-7j8m
English non-LGBTQ+
Low LGBTQ+ C
Moderate LGBTQ+ C
High LGBTQ+ Community Involvement
Low LGBTQ+ Community Involvement
Questionnaire v4
ID 98982
questionnaire-mx41
Medium LGBTQ+ Community Involvement
Questionnaire v6
ID 98978
questionnaire-rqdc
High LGBTQ+ Community Involvement
Questionnaire v4
ID 98534
questionnaire-joig
Demographic Info English-born
Questionnaire v11
ID 81540
questionnaire-8uiw
Finish & Sharing
Task v5
ID 100029
task-44wk
Finish
Include
finish-nib1
Branch
branch-bfqf
Polish LGBTQ+
LGBTQ+ community involvement Polish-born
Questionnaire v12
ID 99721
questionnaire-pth9
Branch
branch-qu5n
Polish non-LGBTQ+
Low LGBTQ+ C
Moderate LGBTQ+ C
High LGBTQ+ Community Involvement
Low LGBTQ+ Community Involvement
Questionnaire v4
ID 98982
questionnaire-wtke
Medium LGBTQ+ Community Involvement
Questionnaire v6
ID 98978
questionnaire-fhpb
High LGBTQ+ Community Involvement
Questionnaire v4
ID 98534
questionnaire-tlo5
Demographic Info Polish-born
Questionnaire v13
ID 78359
questionnaire-jpwv
Finish & Sharing
Task v5
ID 100029
task-hmvu
Finish
Include
finish

9.7 Ethics & Informed Consent

9.7.1 Ethical approval & committee feedback

Downloaded: 14/09/2020
Approved: 27/09/2019

Nathaniel Dziura
Registration number: 180249257
School of English
Programme: English Language and Linguistics

Dear Nathaniel

PROJECT TITLE: Variation and acculturation: Morphosyntax and Queer Polish speakers of British English
APPLICATION: Reference Number 027601

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 27/09/2019 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 027601 (form submission date: 22/09/2019); (expected project end date: 01/12/2021).
- Participant information sheet 1070654 version 4 (22/09/2019).
- Participant information sheet 1070653 version 4 (22/09/2019).
- Participant information sheet 1070695 version 2 (22/09/2019).
- Participant consent form 1070659 version 4 (22/09/2019).
- Participant consent form 1070658 version 4 (22/09/2019).
- Participant consent form 1070689 version 2 (22/09/2019).

The following optional amendments were suggested:

1. Explain the purpose of the Versant test, as well as what it is testing, what kind of scores it produces, and how to interpret them. What is the actual purpose for which the test was developed? Does it match the purpose for which you are using this test in your study? What aspects of language is it testing? Is it targeted to L2 learners specifically, or L1 speakers as well? How are the scores to be interpreted: i.e., does it provide a simple ranking of students or more detailed info about their level matched to the CEFR (Common European Framework of Reference). I am not suggesting that you should not employ this test, but you need to know the answers to all these questions before you administer it in order to understand what it can and cannot tell you about the participants' proficiency in English. 2. Consider using only one proficiency test across both groups of participants (i.e., focus group and questionnaire). If not, address the differences between them appropriately in the analysis without attempting to compare the results. Using a Cloze test to measure English proficiency in the participants taking the questionnaire is a good idea. However, be careful with the analysis of the two sets of data. The proficiency in the Focus Group participants and the Questionnaire participants will not be able to be compared if you use two very different proficiency tests. If you would like to make such comparisons, you would have to employ the same proficiency test in both groups. 3. Pseudo/anonymisation (how done and when each is used) process should be clarified. If the term 'pseudoanonymised' is used on the consent form, participants may want to know what it entails. Provide some explanation in the information sheet. What is anonymised and what pseudoanonymised, and how is this done (e.g. assigning alphanumerical codes?). 4. If audio/video are used 'primarily' for transcription and analysis, what else are they used for? 5. Consent forms: Adam Piette given as Head of School, not Joe Bray. Also, you wrote that "In the first focus group meeting, group members will be introduced to each other and it will be made clear what identities all participants have in common (eg. that they all share a Polish background, minority sexuality/gender identity, Sheffield residency, etc.)." Will they will be made aware of this before they accept to take place and sign the consent form? 6. You mention here the potential effect that Brexit and the current British anti-immigrant sentiment can have on the speakers' level of acculturation. Maybe you could think about how you will take that into account: e.g., interviewing Polish-born L2 English speakers who arrived to the UK before the referendum and after the referendum (I know the number of participants is an issue and it might not be possible, though.)

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Emma Bradley
Ethics Administrator
School of English

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy:

https://www.sheffield.ac.uk/rs/ethicsandintegrity/ethicspolicy/approval-procedure

- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

9.7.2 Information Sheet

05/05/2020

The University Of Sheffield.

Participant Information Sheet

1. Research Project Title:

Language use by Polish-born speakers of British-English

2. Introduction:

You are invited to take part in an online survey as part of a research project concerning language and society. Before you decide whether or not to take part, it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully and consider whether you would like to be involved.

Please email Nathaniel Dziura (nathaniel.dziura@sheffield.ac.uk) if there is anything that is not clear or if you would like some more information.

Thank you for reading this.

3. What is the project's purpose?

This research is being conducted as part of a PhD project in English Language and Linguistics. The project will conclude by January 2023.

This research aims to contribute towards the understanding of the use of English language variation by speakers whose first language is not English.

4. Why have you been chosen?

You are able to take part in this study if you are:

- A **Polish-born adult** living in England who is a native speaker of Polish and also
  - **identifies within the LGBTQ+ label**
  - **does not identify within the LGBTQ+ label**

or:

- An **English-born adult** living in England who is a native speaker of English and also
  - **identifies within the LGBTQ+ label**
  - **does not identify within the LGBTQ+ label**

Note: *LGBTQ+ stands for* Lesbian, Gay, Bisexual, Transgender, Queer, *but is also an umbrella term inclusive of many other minority sexualities and gender identities.*

The participant criteria have been selected as exploring perceptions of linguistic variation in people of different social and linguistic backgrounds can greatly contribute to our understanding of how language can be used as a social tool. As an English-born person with a bilingual and LGBTQ+ background, I believe in the importance

of exploring the language use of the above-outlined groups, and doing so will be crucial to answering the research questions of this study.

Participants for this study will be self-selected, and it is likely that several hundred participants will be involved.

5. Do you have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you may download this information sheet to keep by navigating to 'File → Download as...' at the top left of this page. By completing the Consent Form within the survey, you will be consenting to all aspects of the study.

You may withdraw from the study at any time before 01/12/2020 - even after your participation in the study is over - without any negative consequences. If you wish to do so, please email nathaniel.dziura@sheffield.ac.uk. You do not have to give a reason for your withdrawal and, in the event of this, data pertaining to you will not be used.

6. What will happen if you take part? What do you have to do?

*Your participation in this study will involve completing an online survey. This will take around **15 minutes** and will consist of a short **language exercise** followed by an **interactive quiz** about your social and cultural background.*

- *In the language exercise, you will evaluate different English sentences. You will not be judged for your responses to these sentences.*
- *The cultural background quiz covers your language use, attitudes & identity, and lifestyle.*
- *LGBTQ+ participants will answer a few additional questions about how they interact with the LGBTQ+ community.*
- *You will be given feedback outlining what your quiz results suggest about your cultural background, and will have the chance to comment on how accurate you think they are.*
- *You will also provide some demographic information about your age, and socioeconomic and linguistic background.*

- *After completing this survey, **Polish-born participants** will additionally, at a later time, be required to complete a **15 minute** online Versant English speaking test (free of charge). This will involve simple speaking and reading exercises. Test results will be used to provide a rough indication of English proficiency, but participants will not be assessed in terms of the 'correctness' of answers.*

7. What are the possible disadvantages and risks of taking part?

There is no risk of physical harm associated with taking part in this study. Nothing you say will be scrutinised and broad assumptions and stereotypes about the groups involved will be avoided. To minimise discomfort, sensitive questions (eg. regarding how you label your sexuality) will be optional.
*Participation in the online survey will take around **15 minutes** for English-born participants, and around **35 minutes** for Polish-born participants (because they will also complete the Versant English speaking test).*

8. What are the possible benefits of taking part?

Whilst there are no immediate benefits for people participating in this project, this work aims to broaden the

research that has been carried out on language and society, and on minority groups living in the UK. The results of this study will contribute to a better understanding of the complexities of language use.

*Additionally, **Polish-born participants**, having completed the Versant English Speaking Test, will obtain an **official Score Report** for this widely accredited test, which can be used towards future CVs or job applications. The Score Report shows how your Versant score relates to other common language testing scales (eg. GSE and CEFR). Access to the test will be provided for free for participants of this study (normally $34).*

9. Will your taking part in this project be kept confidential?

Information collected from you during the course of the research will be pseudonymised; You will be allocated an alphanumeric participant ID number for purposes of data processing. Your survey data may be quoted in publications, reports, web pages, and other research outputs resulting from this study, but you will not be identifiable from any of these research materials. No identifiable personal data will be included when analysing and presenting data. Within the survey, you will be asked for your postcode, job type, and education level, however, this will only be used to determine your socioeconomic background, and will not be redistributed or published in research materials.

Polish-born participants will have to provide an email address in order to receive instructions for the Versant English speaking test. Additionally, all participants will have the option to provide their email address if they would like to be contacted regarding follow-up research. This is entirely non-obligatory and does not mean you are committing to any future involvement. Email addresses will only be used for the purposes outlined above; Only the lead researcher will have access to identifiable personal data (eg. email addresses) and this will not be redistributed.

10. What is the legal basis for processing your personal data?

According to data protection legislation, I am required to inform you that the legal basis I am applying in order to process your personal data is that 'processing is necessary for the performance of a task carried out in the public interest' (Article 6(1)(e)). Further information can be found in the University's Privacy Notice https://www.sheffield.ac.uk/govern/data-protection/privacy/general.'

11. What will happen to the data collected, and the results of the research project?

The results of this study will be used to inform my PhD study, which should be completed by January 2023. If you would like to obtain a copy of my thesis once it has been assessed, please contact me by email.

All identifiable personal data that is collected (eg. email addresses) will be stored on a password-protected University Filestore and will not be redistributed. All of your identifiable personal data will be destroyed within 5 years of the study's completion.

Your pseudonymised data may be used in my future research, the results of which may appear in publications, reports, web pages, etc. Due to the nature of this research, it is very likely that other researchers may find the data collected to be useful in answering future research questions. Therefore, it is possible that your pseudonymised data may be shared with other authorised data processors. I will ensure that any data collected from you is untraceable back to you before allowing others to see it.

12. Who is organising and funding the research?

This research is part of a PhD study in English Language and Linguistics within the School of English at the

University of Sheffield. The degree is funded by a White Rose College of Arts and Humanities studentship.

13. Who is the Data Controller?

The University of Sheffield will act as the Data Controller for this study. This means that the University is responsible for looking after your information and using it properly. For this reason, you will be asked on the Consent Form to assign the copyright you hold in any materials generated as part of this project to The University of Sheffield.

14. Who has ethically reviewed the project?

This project has been ethically approved via The Department of Arts and Humanities' ethics review procedure through The University of Sheffield's Research Ethics Committee.

15. What if something goes wrong or you wish to complain about the research?

If you have any concerns at all, please contact me, the lead researcher, Nathaniel Dziura (nathaniel.dziura@sheffield.ac.uk) or my supervisors, Prof Emma Moore (e.moore@sheffield.ac.uk) and Dr Robyn Orfitelli (r.orfitelli@sheffield.ac.uk).

Please do not hesitate to contact my supervisors if you have any complaints regarding unfair treatment during the study or in the event that something serious occurs during or after your participation in the study. This will be treated very seriously. Should you feel that your complaint has not been handled to your satisfaction by my supervisors, then please contact the Head of the School of English, Professor Jane Hodson (j.hodson@sheffield.ac.uk) who will then escalate your complaint through the appropriate channels.

If the complaint relates to how your personal data has been handled, information about how to raise a complaint can be found in the University's Privacy Notice: https://www.sheffield.ac.uk/govern/data-protection/privacy/general.

16. Contact for further information

Lead researcher:

Nathaniel Dziura - nathaniel.dziura@sheffield.ac.uk

Project supervisors:

Prof Emma Moore - e.moore@sheffield.ac.uk

Tel.: +44 (0)114 222 0232

Address:

Room 5.03
Jessop West
1 Upper Hanover Street
Sheffield
S3 7RA

Dr Robyn Orfitelli - r.orfitelli@sheffield.ac.uk

05/05/2020

Tel.: +44 (0)114 222 0224

Address:

Room 3.14
Jessop West
1 Upper Hanover Street
Sheffield
S3 7RA

Thank you very much for your time!

9.8 Model outputs

The following are truncated model outputs containing only the model terms shown to have significant effects on AJT Response. Original p values are given, along with Benjamini-Hochberg (BH) adjusted p values to control for False Discovery Rate (FDR). Rows that are greyed out were significant in the original model outputs, but no longer significant under BH adjustment. The outputs also show the coefficient estimates (in terms of ordered log-transformed odds, or 'logits'), proportional odds ratio values (exponentiated equivalents of the original log odds estimates), and other elements of the model output (standard errors and z values of the coefficient estimates). Confidence intervals are provided for the A larger margin between the 2.5% and 97.5% CI indicates lower precision of the OR calculation, whereas a small CI range indicates higher precision. ORs with a 2.5%-97.5% CI range higher than 10 (a threshold that has been arbitrarily selected based on the observed range of CI values across the data) are underlined in order to highlight that the confidence in the associated statistical power calculation is relatively lower than for other terms, because the margins between the CIs are higher. Model terms containing '.L' indicate a linear effect. Certain quadratic ('.Q') and cubic ('.C') effects are also present in some of the models, but these were not included in analysis.

9.8.1 Model 1 (*all participants*)

Model term	Model output				p value *(BH adjusted)*	PO Ratios	CIs for PO Ratios	
	Estimate (ordered log odds/ logits)	Std. Error	z value	p value			2.5%	97.5%
IMD.L	-0.52167	0.23042	-2.26403	0.024	0.044	0.59353	0.37784	0.93234
AJTConditionLD	-2.88477	0.46849	-6.15763	0.000	0.000	0.05587	0.02230	0.13994
AJTConditionNS-was	-4.57251	0.46815	-9.76719	0.000	0.000	0.01033	0.00413	0.02586
AJTConditionNS-were	-4.51699	0.45529	-9.92108	0.000	0.000	0.01092	0.00447	0.02666
AJTConditionNS-weren't	-3.77340	0.45267	-8.33589	0.000	0.000	0.02297	0.00946	0.05579
AJTConditionRD	-2.35683	0.47068	-5.00730	0.000	0.000	0.09472	0.03765	0.23828
AJTConditionTopicalisation	-3.48267	0.45585	-7.63995	0.000	0.000	0.03073	0.01257	0.07508
ageBand.L	0.72376	0.29437	2.45871	0.014	0.028	2.06217	1.15814	3.67189
ageBand.C	0.70712	0.26411	2.67736	0.007	0.015	2.02815	1.20861	3.40340
AJTConditionNS-was:birthCountryPolish-born	-1.12257	0.22646	-4.95709	0.000	0.000	0.32544	0.20879	0.50726
AJTConditionNS-were:birthCountryPolish-born	-0.95167	0.22168	-4.29294	0.000	0.000	0.38610	0.25003	0.59620
AJTConditionNS-weren't:birthCountryPolish-born	-1.66808	0.22215	-7.50883	0.000	0.000	0.18861	0.12203	0.29151
AJTConditionRD:birthCountryPolish-born	-1.06014	0.22812	-4.64720	0.000	0.000	0.34641	0.22152	0.54171
IMD.L:AJTConditionNS-was	0.57955	0.19818	2.92437	0.003	0.008	1.78524	1.21061	2.63263
IMD.L:AJTConditionNS-were	0.59301	0.19431	3.05181	0.002	0.005	1.80942	1.23635	2.64814
IMD.L:AJTConditionTopicalisation	0.44289	0.19666	2.25211	0.024	0.044	1.55721	1.05914	2.28950
IMD.Q:AJTConditionTopicalisation	-0.39402	0.18757	-2.10064	0.036	0.061	0.67434	0.46690	0.97396
AJTConditionLD:regionNorth	-0.80782	0.36558	-2.20968	0.027	0.047	0.44583	0.21776	0.91275
AJTConditionNS-was:regionNorth	-1.15672	0.36156	-3.19930	0.001	0.004	0.31452	0.15484	0.63885
AJTConditionNS-were:regionNorth	-0.92983	0.35418	-2.62527	0.009	0.018	0.39462	0.19711	0.79006
AJTConditionNS-weren't:regionNorth	-1.28595	0.35245	-3.64865	0.000	0.001	0.27639	0.13852	0.55147
AJTConditionRD:regionNorth	-1.10000	0.36816	-2.98786	0.003	0.007	0.33287	0.16177	0.68494
AJTConditionTopicalisation:regionNorth	-0.70514	0.35746	-1.97262	0.049	0.080	0.49404	0.24518	0.99548
AJTConditionNS-were:regionSouth	-0.65764	0.32912	-1.99817	0.046	0.076	0.51807	0.27179	0.98751
AJTConditionNS-weren't:regionSouth	-0.97262	0.32806	-2.96473	0.003	0.007	0.37809	0.19877	0.71919
AJTConditionRD:regionSouth	-0.78166	0.34320	-2.27754	0.023	0.044	0.45765	0.23356	0.89674
AJTConditionLD:LGBTQYes	0.54120	0.24000	2.25499	0.024	0.044	1.71806	1.07337	2.74997
AJTConditionNS-was:LGBTQYes	0.90469	0.23904	3.78471	0.000	0.000	2.47117	1.54679	3.94796
AJTConditionNS-were:LGBTQYes	0.78491	0.23407	3.35330	0.001	0.002	2.19220	1.38561	3.46833
AJTConditionNS-weren't:LGBTQYes	0.74935	0.23351	3.20900	0.001	0.004	2.11562	1.33866	3.34354
AJTConditionRD:LGBTQYes	0.93187	0.24052	3.87435	0.000	0.000	2.53926	1.58479	4.06856
AJTConditionNS-was:ageBand.L	-1.32888	0.25550	-5.20104	0.000	0.000	0.26477	0.16047	0.43688
AJTConditionNS-were:ageBand.L	-1.21244	0.25104	-4.82972	0.000	0.000	0.29747	0.18187	0.48655

AJTConditionNS-weren't:ageBand.L	-1.58539	0.25129	-6.30905	0.000	0.000	0.20487	0.12519	0.33525
AJTConditionTopicalisation:ageBand.L	-0.79216	0.25132	-3.15193	0.002	0.004	0.45287	0.27672	0.74114
AJTConditionLD:ageBand.C	-0.81586	0.23109	-3.53041	0.000	0.001	0.44226	0.28117	0.69564
AJTConditionNS-was:ageBand.C	-0.66688	0.23041	-2.89435	0.004	0.008	0.51331	0.32678	0.80631
AJTConditionNS-were:ageBand.C	-0.92059	0.22607	-4.07220	0.000	0.000	0.39828	0.25572	0.62033
AJTConditionNS-weren't:ageBand.C	-1.04777	0.22587	-4.63885	0.000	0.000	0.35072	0.22527	0.54603
AJTConditionRD:ageBand.C	-0.91757	0.23146	-3.96430	0.000	0.000	0.39949	0.25380	0.62881
AJTConditionTopicalisation:ageBand.C	-0.78223	0.22815	-3.42859	0.001	0.002	0.45738	0.29247	0.71529

9.8.2 Model 2a (*Polish-born participants only*)

Model term	Model output				p value *(BH adjusted)*	PO Ratios	CIs for PO Ratios	
	Estimate (ordered log odds/ logits)	Std. Error	z value	p value			2.5%	97.5%
AJTConditionNS-was	-5.23975	0.78650	-6.66215	0.000	0.000	0.00530	0.00114	0.02477
AJTConditionNS-were	-3.74237	0.76687	-4.88003	0.000	0.000	0.02370	0.00527	0.10653
AJTConditionNS-weren't	-4.50961	0.76815	-5.87071	0.000	0.000	0.01100	0.00244	0.04959
AJTConditionRD	-2.16853	0.79946	-2.71250	0.007	0.023	0.11435	0.02386	0.54792
AJTConditionTopicalisation	-1.61982	0.76138	-2.12747	0.033	0.080	0.19794	0.04451	0.88026
LGBTQYes	-0.97488	0.41468	-2.35091	0.019	0.054	0.37724	0.16735	0.85034
higherEdYes	1.50528	0.44683	3.36883	0.001	0.004	4.50543	1.87671	10.81622
learningEnglish_Grouped.Q	1.23343	0.59402	2.07642	0.038	0.087	3.43298	1.07163	10.99763
ageBand.L	-1.45966	0.54700	-2.66847	0.008	0.025	0.23232	0.07952	0.67872
AJTConditionLD:regionNorth	-1.63453	0.69673	-2.34601	0.019	0.054	0.19505	0.04978	0.76417
AJTConditionNS-was:regionNorth	-1.51576	0.67339	-2.25095	0.024	0.063	0.21964	0.05869	0.82205
AJTConditionNS-were:regionNorth	-2.13839	0.66043	-3.23786	0.001	0.006	0.11784	0.03230	0.43000
AJTConditionNS-weren't:regionNorth	-2.07936	0.65718	-3.16407	0.002	0.007	0.12501	0.03448	0.45325
AJTConditionLD:regionSouth	-1.42243	0.67149	-2.11832	0.034	0.080	0.24113	0.06467	0.89912
AJTConditionNS-were:regionSouth	-1.24853	0.62468	-1.99868	0.046	0.100	0.28693	0.08434	0.97610
AJTConditionNS-weren't:regionSouth	-1.51262	0.62222	-2.43102	0.015	0.047	0.22033	0.06508	0.74595
AJTConditionLD:LGBTQYes	1.14633	0.37778	3.03438	0.002	0.010	3.14661	1.50064	6.59796
AJTConditionNS-was:LGBTQYes	1.85238	0.37801	4.90039	0.000	0.000	6.37497	3.03891	13.37329
AJTConditionNS-were:LGBTQYes	1.79096	0.37131	4.82329	0.000	0.000	5.99518	2.89560	12.41267
AJTConditionNS-weren't:LGBTQYes	1.90398	0.37186	5.12016	0.000	0.000	6.71256	3.23863	13.91284
AJTConditionRD:LGBTQYes	1.45693	0.37610	3.87379	0.000	0.001	4.29275	2.05400	8.97165
AJTConditionLD:higherEdYes	-2.07696	0.40389	-5.14240	0.000	0.000	0.12531	0.05678	0.27655
AJTConditionNS-was:higherEdYes	-1.14212	0.40065	-2.85066	0.004	0.017	0.31914	0.14553	0.69987
AJTConditionNS-were:higherEdYes	-1.97822	0.38710	-5.11034	0.000	0.000	0.13832	0.06477	0.29537
AJTConditionNS-weren't:higherEdYes	-1.46444	0.39047	-3.75051	0.000	0.001	0.23121	0.10756	0.49701
AJTConditionRD:higherEdYes	-2.01399	0.39548	-5.09253	0.000	0.000	0.13346	0.06148	0.28971
AJTConditionTopicalisation:higherEdYes	-2.30915	0.38711	-5.96507	0.000	0.000	0.09935	0.04652	0.21216
AJTConditionLD:learningEnglish_Grouped.Q	-1.20463	0.55212	-2.18184	0.029	0.073	0.29980	0.10160	0.88470
AJTConditionNS-weren't:learningEnglish_Grouped.Q	-1.37361	0.55521	-2.47405	0.013	0.043	0.25319	0.08528	0.75170
AJTConditionRD:learningEnglish_Grouped.Q	-1.19155	0.55008	-2.16615	0.030	0.074	0.30375	0.10334	0.89278
AJTConditionTopicalisation:learningEnglish_Grouped.Q	-1.47255	0.54290	-2.71237	0.007	0.023	0.22934	0.07913	0.66466
AJTConditionNS-was:learningEnglish_Grouped.C	-1.21765	0.44642	-2.72759	0.006	0.023	0.29592	0.12336	0.70986

AJTConditionNS-were:learningEnglish_Grouped.C	-1.02695	0.43407	-2.36587	0.018	0.054	0.35810	0.15294	0.83846
AJTConditionTopicalisation:learningEnglish_Grouped^4	-0.72661	0.31494	-2.30715	0.021	0.057	0.48355	0.26083	0.89642
AJTConditionLD:ageBand.L	1.70773	0.49296	3.46425	0.001	0.003	5.51643	2.09919	14.49657
AJTConditionRD:ageBand.L	1.70460	0.49159	3.46750	0.001	0.003	5.49916	2.09823	14.41252
AJTConditionTopicalisation:ageBand.L	1.46075	0.47992	3.04371	0.002	0.010	4.30917	1.68222	11.03839
AJTConditionNS-were:ageBand.C	-0.70269	0.35848	-1.96022	0.050	0.107	0.49525	0.24530	0.99991
AJTConditionNS-weren't:ageBand.C	-0.72905	0.35811	-2.03584	0.042	0.094	0.48237	0.23909	0.97319
AJTConditionLD:lifestyleEnglish.L	0.89868	0.38466	2.33632	0.019	0.054	2.45636	1.15577	5.22050
AJTConditionNS-was:lifestyleEnglish.L	1.26664	0.38757	3.26814	0.001	0.005	3.54890	1.66032	7.58569
AJTConditionRD:lifestyleEnglish.L	1.18605	0.38642	3.06933	0.002	0.009	3.27412	1.53523	6.98257

9.8.3 Model 2b (*Polish-born participants with Versant/CEFR only*)

Model term	Model output				p value (*BH adjusted*)	PO Ratios	CIs for PO Ratios	
	Estimate (ordered log odds/ logits)	Std. Error	z value	p value			2.5%	97.5%
AJTConditionNS-was	-5.97447	0.95791	-6.23701	0.000	0.000	0.00254	0.00039	0.01662
AJTConditionNS-were	-4.01386	0.92596	-4.33483	0.000	0.000	0.01806	0.00294	0.11091
AJTConditionNS-weren't	-4.69292	0.92393	-5.07931	0.000	0.000	0.00916	0.00150	0.05602
AJTConditionRD	-2.01518	0.96032	-2.09845	0.036	0.100	0.13330	0.02030	0.87548
LGBTQYes	-1.44173	0.60526	-2.38200	0.017	0.057	0.23652	0.07222	0.77457
CEFR.L	2.98407	0.99176	3.00887	0.003	0.012	19.76817	2.82997	138.08680
AJTConditionLD:regionNorth	-3.17638	0.87131	-3.64550	0.000	0.003	0.04174	0.00757	0.23024
AJTConditionNS-was:regionNorth	-2.25740	0.85253	-2.64787	0.008	0.032	0.10462	0.01968	0.55629
AJTConditionNS-were:regionNorth	-2.68924	0.84224	-3.19296	0.001	0.007	0.06793	0.01304	0.35399
AJTConditionNS-weren't:regionNorth	-2.73499	0.83728	-3.26651	0.001	0.007	0.06489	0.01258	0.33489
AJTConditionRD:regionNorth	-2.56554	0.87032	-2.94782	0.003	0.014	0.07688	0.01396	0.42327
AJTConditionTopicalisation:regionNorth	-2.66727	0.84003	-3.17520	0.001	0.007	0.06944	0.01338	0.36029
AJTConditionLD:regionSouth	-1.45584	0.73500	-1.98074	0.048	0.120	0.23320	0.05522	0.98485
AJTConditionNS-weren't:regionSouth	-1.41849	0.69028	-2.05495	0.040	0.105	0.24208	0.06257	0.93654
AJTConditionLD:higherEdYes	-1.40427	0.64288	-2.18433	0.029	0.085	0.24555	0.06965	0.86568
AJTConditionNS-were:higherEdYes	-1.78475	0.61300	-2.91148	0.004	0.015	0.16784	0.05048	0.55806
AJTConditionRD:higherEdYes	-1.72307	0.63492	-2.71386	0.007	0.027	0.17852	0.05143	0.61961
AJTConditionLD:LGBTQYes	1.91272	0.56857	3.36407	0.001	0.006	6.77147	2.22184	20.63726
AJTConditionNS-was:LGBTQYes	2.00295	0.56680	3.53378	0.000	0.004	7.41091	2.44011	22.50788
AJTConditionNS-were:LGBTQYes	1.83013	0.56080	3.26341	0.001	0.007	6.23469	2.07710	18.71419
AJTConditionNS-weren't:LGBTQYes	1.97181	0.56001	3.52100	0.000	0.004	7.18368	2.39697	21.52940
AJTConditionRD:LGBTQYes	2.03185	0.56758	3.57988	0.000	0.004	7.62817	2.50784	23.20281
AJTConditionTopicalisation:LGBTQYes	1.14982	0.55726	2.06336	0.039	0.105	3.15763	1.05931	9.41237
AJTConditionRD:ageBand.L	1.15388	0.54056	2.13462	0.033	0.094	3.17048	1.09901	9.14637
AJTConditionLD:ageBand.Q	-1.03199	0.51630	-1.99882	0.046	0.117	0.35630	0.12952	0.98014
AJTConditionRD:ageBand.Q	-1.16158	0.51554	-2.25312	0.024	0.076	0.31299	0.11395	0.85973
AJTConditionLD:CEFR.L	-2.13523	0.89305	-2.39096	0.017	0.057	0.11822	0.02054	0.68052
AJTConditionNS-was:CEFR.L	-3.12818	0.91600	-3.41505	0.001	0.005	0.04380	0.00727	0.26372
AJTConditionNS-were:CEFR.L	-2.26475	0.86707	-2.61196	0.009	0.034	0.10386	0.01898	0.56817
AJTConditionNS-weren't:CEFR.L	-2.81314	0.86983	-3.23415	0.001	0.007	0.06002	0.01091	0.33011
AJTConditionRD:CEFR.L	-2.80904	0.87939	-3.19430	0.001	0.007	0.06026	0.01075	0.33775
AJTConditionTopicalisation:CEFR.L	-4.27074	0.87451	-4.88358	0.000	0.000	0.01397	0.00252	0.07756

AJTConditionNS-was:CEFR.C	-2.00105	0.84830	-2.35889	0.018	0.059	0.13519	0.02564	0.71290
AJTConditionNS-was:lifestyleEnglish.L	1.49517	0.57712	2.59075	0.010	0.035	4.46010	1.43913	13.82259
AJTConditionRD:lifestyleEnglish.L	1.47191	0.58026	2.53664	0.011	0.040	4.35756	1.39741	13.58819

9.8.4 Model 3 (*LGBTQ+ participants only*)

Model term	Model output				p value *(BH adjusted)*	PO Ratios	CIs for PO Ratios	
	Estimate (ordered log odds/ logits)	Std. Error	z value	p value			2.5%	97.5%
AJTConditionLD	-1.93168	0.78684	-2.45499	0.014	0.042	0.14491	0.03100	0.67739
AJTConditionNS-was	-4.28381	0.76987	-5.56436	0.000	0.000	0.01379	0.00305	0.06236
AJTConditionNS-were	-3.17781	0.75431	-4.21289	0.000	0.000	0.04168	0.00950	0.18279
AJTConditionNS-weren't	-2.78297	0.75757	-3.67353	0.000	0.001	0.06186	0.01401	0.27304
AJTConditionTopicalisation	-2.80737	0.76322	-3.67832	0.000	0.001	0.06036	0.01352	0.26942
genderNotBinary	1.64839	0.61744	2.66972	0.008	0.027	5.19861	1.54997	17.43620
IMD.L	-0.74556	0.21531	-3.46274	0.001	0.003	0.47447	0.31113	0.72357
lifestyleEnglish.Q	0.87963	0.44384	1.98187	0.047	0.094	2.41002	1.00977	5.75199
regionNorth	1.34736	0.65466	2.05812	0.040	0.083	3.84725	1.06637	13.88019
ageBand.L	1.26020	0.49046	2.56944	0.010	0.035	3.52614	1.34841	9.22098
ageBand.C	0.99389	0.41273	2.40811	0.016	0.043	2.70173	1.20317	6.06677
AJTConditionNS-was:birthCountryPolish-born	-0.72386	0.32432	-2.23197	0.026	0.059	0.48488	0.25679	0.91556
AJTConditionNS-were:birthCountryPolish-born	-0.64018	0.31805	-2.01283	0.044	0.091	0.52720	0.28265	0.98333
AJTConditionNS-weren't:birthCountryPolish-born	-1.17969	0.32020	-3.68425	0.000	0.001	0.30737	0.16410	0.57573
AJTConditionRD:birthCountryPolish-born	-0.71447	0.33388	-2.13992	0.032	0.069	0.48945	0.25440	0.94169
AJTConditionLD:genderNotBinary	-1.10176	0.55593	-1.98183	0.047	0.094	0.33229	0.11176	0.98792
AJTConditionRD:genderNotBinary	-1.32279	0.54997	-2.40522	0.016	0.043	0.26639	0.09065	0.78280
AJTConditionTopicalisation:genderNotBinary	-1.73635	0.54251	-3.20060	0.001	0.007	0.17616	0.06083	0.51015
AJTConditionLD:lifestyleEnglish.L	1.04341	0.47178	2.21163	0.027	0.060	2.83888	1.12607	7.15699
AJTConditionNS-was:lifestyleEnglish.L	1.01110	0.45833	2.20605	0.027	0.060	2.74862	1.11939	6.74912
AJTConditionRD:lifestyleEnglish.L	2.12848	0.47864	4.44694	0.000	0.000	8.40205	3.28827	21.46857
AJTConditionNS-was:lifestyleEnglish.Q	-0.94710	0.38535	-2.45777	0.014	0.042	0.38787	0.18225	0.82545
AJTConditionTopicalisation:lifestyleEnglish.Q	-0.84652	0.37936	-2.23144	0.026	0.059	0.42891	0.20392	0.90214
AJTConditionNS-was:regionNorth	-1.73040	0.57006	-3.03547	0.002	0.011	0.17721	0.05798	0.54167
AJTConditionNS-were:regionNorth	-2.03410	0.56024	-3.63077	0.000	0.002	0.13080	0.04362	0.39217
AJTConditionNS-weren't:regionNorth	-2.19487	0.56289	-3.89928	0.000	0.001	0.11137	0.03695	0.33567
AJTConditionRD:regionNorth	-1.58053	0.59048	-2.67668	0.007	0.027	0.20587	0.06471	0.65494
AJTConditionTopicalisation:regionNorth	-1.71596	0.57049	-3.00787	0.003	0.012	0.17979	0.05877	0.55001
AJTConditionNS-were:regionSouth	-1.14579	0.48073	-2.38343	0.017	0.043	0.31797	0.12393	0.81581
AJTConditionNS-weren't:regionSouth	-1.17505	0.48513	-2.42215	0.015	0.043	0.30880	0.11933	0.79914
AJTConditionNS-was:ageBand.L	-1.94492	0.43467	-4.47450	0.000	0.000	0.14300	0.06100	0.33522
AJTConditionNS-were:ageBand.L	-1.80878	0.43427	-4.16506	0.000	0.000	0.16385	0.06995	0.38381
AJTConditionNS-weren't:ageBand.L	-2.26088	0.43236	-5.22919	0.000	0.000	0.10426	0.04468	0.24330
AJTConditionTopicalisation:ageBand.L	-1.67494	0.42822	-3.91144	0.000	0.001	0.18732	0.08093	0.43359
AJTConditionNS-was:ageBand.Q	-1.07028	0.39500	-2.70955	0.007	0.026	0.34291	0.15811	0.74372

AJTConditionNS-were:ageBand.Q	-1.47951	0.39191	-3.77518	0.000	0.001	0.22775	0.10565	0.49096
AJTConditionNS-weren't:ageBand.Q	-1.57761	0.39335	-4.01069	0.000	0.001	0.20647	0.09551	0.44635
AJTConditionRD:ageBand.Q	-1.03247	0.40183	-2.56943	0.010	0.035	0.35613	0.16202	0.78278
AJTConditionTopicalisation:ageBand.Q	-0.86364	0.38792	-2.22635	0.026	0.059	0.42162	0.19712	0.90182
AJTConditionNS-were:ageBand.C	-0.91379	0.35925	-2.54362	0.011	0.035	0.40100	0.19831	0.81085
AJTConditionNS-weren't:ageBand.C	-0.97412	0.35785	-2.72216	0.006	0.026	0.37753	0.18722	0.76128
AJTConditionRD:ageBand.C	-0.89348	0.37687	-2.37078	0.018	0.043	0.40923	0.19551	0.85656
AJTConditionTopicalisation:ageBand.C	-0.88667	0.36394	-2.43631	0.015	0.043	0.41203	0.20190	0.84083
AJTConditionLD:higherEdYes	-1.02676	0.43268	-2.37304	0.018	0.043	0.35817	0.15339	0.83633
AJTConditionRD:higherEdYes	-1.11635	0.43835	-2.54668	0.011	0.035	0.32747	0.13869	0.77322
AJTConditionLD:LGBTQCommunityInvolvement.L	-0.91460	0.30568	-2.99208	0.003	0.012	0.40068	0.22009	0.72943
AJTConditionRD:LGBTQCommunityInvolvement.L	-0.72909	0.30500	-2.39046	0.017	0.043	0.48235	0.26531	0.87695
AJTConditionTopicalisation:LGBTQCommunityInvolvement.L	-0.88639	0.29761	-2.97838	0.003	0.012	0.41214	0.23000	0.73854