Migrant Workers' Varieties of Arabic in Hijaz, Madinah: Pidgin or Interlanguage Varieties?

Abrar Bazerbay

Thesis submitted for the degree of Doctor of Philosophy (Integrated)

School of English Literature, Language and Linguistics

Newcastle University

January 2024

Abstract

This thesis investigates whether the Arabic L2 varieties spoken by 30 Madinah-based migrant workers, who have Bengali, Hindi-Urdu or Tagalog as their L1, are pidgin or interlanguage varieties. To answer this research question, I first examine the verbal and nominal agreements in the speech of three migrant workers. I then investigate in detail the impact of a series of internal and external factors on the use of the definite article /ʔal-/ 'the', the coordinating conjunction marker /wa/ 'and', the production of /f/ in the speech of all 30 migrant workers in the corpus. I hypothesise that the migrant workers' Arabic L2 can be considered a pidgin variety if the following apply: they have reduced verbal and nominal systems; they typically delete the morphosyntactic features and substitute /f/ with other variants; contextual predictors negligibly impact their L2 production; and their Arabic L2 is simplified compared to Hijazi Arabic (HA), the local variety of Arabic. However, if their use of the morphological features follows the usage patterns of L1 HA speakers and their use of morphosyntactic and phological features is significantly conditioned by contextual factors, the migrant workers' Arabic L2 can be considered interlanguage varieties.

Data were collected by way of one-hour semi-structured Zoom interviews and questionnaires. The questionnaires included a series of demographic information, questions concerning the participants' use and acquisition of Arabic and attitudinal information. The qualitative analysis of targeted morphological features reveals that the Arabic speech of the three migrants lacks both subject-verb agreement and noun-adjective-agreement. This result possibly supports the pidgin hypothesis. The quantitative analysis shows that while the examined linguistic features are employed categorically in L1 HA, they are used variably in the migrants' L2 varieties. Mixed-effect logistic regression models show that the migrant workers in my sample overwhelmingly delete the morphosyntactic features, with most of the independent variables not significantly affecting their use. The variations across various linguistic variables and throughout various language groups suggest that the L2 Arabic spoken by the migrants represent an interlanguage. With the phonological feature /f/, the impact of migrants' L1 becomes apparent, and the use of this variable is significantly affected by the independent variables included in the modelling. Hence, these findings suggest that the migrants' Arabic L2s are interlanguage varieties.

The analysis of morphological features and the analysis of morphosyntactic and phonological features point towards opposite conclusions. I tentatively propose that the Arabic varieties of the migrant workers studied in this thesis are located on a continuum, where the higher end may be closer to an interlanguage.

Acknowledgments

(الحمدلله الذي هدانا لهذا وماكنا لنهتدي لولا ان هدانا الله)

‘Praise to Allah, who has guided us to this; and we would never have been guided if Allah had not guided us’ (Quran 7: 43). First and foremost, I am grateful to Almighty God for granting me strength, patience and health throughout the duration of my work.

I am immensely grateful to my supervisors Dr. Heike Pichler and Dr. Daniel Duncan for reviewing my drafts and providing me with the meticulous feedback I needed. Being under their supervision was an honour and a tremendous privilege. Without their guidance, constructive feedback and constant support and encouragement, this work would not have been completed. I wish to express my sincere gratitude to my thesis examiners, Prof. Anthony Grant (Edge Hill University) and Dr. Niamh Kelly (Newcastle University) for agreeing to examine this thesis and for their insightful comments throughout the viva voce examination. The time they dedicated is extremely appreciated.

I am incredibly thankful to my mother, sisters and brothers, who prayed for me, believed in me and supported me throughout my journey. My gratitude also goes to my beloved son, Ammar and daughter, Layal. They have been my companions on this journey and the strongest motivation for me to keep working.

In conclusion, I would like to express my gratitude to my sponsors in Saudi Arabia: the Ministry of Education and Taibah University, as well as the Saudi Cultural Bureau in London for their support and for funding my scholarship.

Table of Contents

List of Tables

List of Figures

List of Abbreviation

1	First Person
2	Second Person
3	Third Person
AIC	Akaike Information Criterion
ADJ	Adjective
AUX	Auxiliary
BHA	Bedouin Hijazi Arabic
CLF	Classifier
CONJ	Coordinating Conjunction Marker
COP	Copula
CPH	Critical Period Hypothesis
DEF	Definite Article
DP	Determiner phrase
F	Feminine
FOC	Focus
GA	Gulf Arabic
GPA	Gulf Pidgin Arabic
HA	Hijazi Arabic
IMP	Imperative
L1	First Language
L2	Second Language
LoR	Length of Residence
M	Masculine
NEG	Negative, Negation
NP	Noun phrase
PIS	Participant Information Sheet
PL	Plural
POSS	Possessive
PROG	Progressive

PRS Present

PST Past

S Subject

SG Singular

UHA Urban Hijazi Arabic

Chapter 1: Introduction

Saudi Arabia is a contact situation due to migration, attracting a considerable amount of migrant workers to undertake low-status work. This thesis focuses on the L2 Arabic varieties of migrant workers, who use Bengali, Hindi-Urdu and Tagalog as their L1, in Madinah (also spelled Medina), Hijaz. It will test whether their L2 Arabic varieties are an interlanguage or a pidgin.

Previous research has shown that social factors, such as the amount of using L1/L2, age and length of residence (LoR) significantly influence interlanguage production while minimally influencing pidgin production (Albaqawi 2020; Almoaily 2012; Drummond 2010; Piske *et al.* 2001). I operationalize these findings to distinguish two possibilities: if L2 Arabic production of the migrant workers, including the definite article /ʔal-/ 'the', the coordinating conjunction marker /wa/ 'and', and the production of /f/, is conditioned by internal and external factors, we can infer that their varieties are interlanguage. However, if migrants mostly delete the morphosyntactic features, substitute the phonological variable with other variants and the contextual factors have a negligible impact on their use of native Arabic features, we can predict that their varieties are pidgin.

This chapter provides a background of the migrant workers in Madinah; Hijaz Saudi Arabia, the rationale, significance and aims of the study, research questions of the study together with the structure of this thesis.

1.1 Migrant Workers in Madinah; Hijaz; Saudi Arabia

Since the oil boom in 1938, Saudi Arabia has attracted a high number of migrant workers from Western and poorer neighbouring Arab countries as well as different South Asian countries (De Bel-Air 2018). Migrant workers in general work in the oil companies and in different industries such as the construction sector, air conditioner technicians, tailors and various other fields (Bakir 2010). The need for migrant workers continues nowadays particularly because young native Saudis are less likely to perform manual labour, either on account of the low pay or the social stigma of such work (Alhazmi 2020; Almoaily 2012).

One of the main cities in the Hijaz Region of Saudi Arabia is Madinah, a city with a population of 2,137,983 in 2022, and which is visited by thousands of Muslims each year for religious purposes. Thanks to the Prophet Mohammed's Mosque there is a stable tourism economy, attracting many migrant workers to work in tourism and associated industries supporting it. According to 2019 data otained from the General Authority for Statistics in Saudi Arabia, there are 131,034 local workers and 237,843 migrant workers in Madinah. The latest Saudi census reveals that among the top ten migrant nationalities in 2022 are Bangladeshi, Indian and Filipino. For this reason, I focus on these particular migrants' use of Arabic as an additional language. These workers generally tend to leave the Hijaz region of Saudi Arabia, eventually. Although their work visa is valid for two years and it can be renewed several times, they do not have a pathway to Saudi citizenship; hence, they remain permanently migrant workers. Prior to arriving in Saudi Arabia, most of these workers do not have any knowledge of Arabic, as reported in my study. Also, there is no common language among these worker groups and as they need to communicate with each other and with L1 Hijazi Arabic speakers, particularly at work, they adopt an L2 Arabic variety.

In popular culture, there is a perception that the Arabic varieties of the migrant workers are different from the variety spoken by the L1 Hijazi Arabic speakers. For instance, Alshraidah (2011) describes the migrants' varieties in Alriyadh Newspaper as consisting of simplified grammar and limited vocabulary, giving fuel to the idea that their varieties are pidgins. However, there is another, non-pidgin outcome of language contact that should be considered as well: whether the migrant workers' Arabic varieties are in fact an interlanguage. My study sets out to clarify whether the Arabic varieties spoken by migrant workers in Madinah constitute pidgin or interlanguage.

1.2 Rational, Significance and Aims of the Study

There is a general paucity of studies investigating the Arabic varieties of migrant workers. In Saudi Arabia and the Gulf area, previous studies on the varieties of migrant workers are mostly restricted to qualitative descriptive work like Albaqawi (2016), Al-Zubeiry (2015), Bakir (2010), Hobrom (1996), Næss (2008) and Salem (2013). It is my understanding that only the study completed by Aljutaily (2018) applied quantitative methods. However, the studies carried out by Albaqawi (2020), Almoaily (2012) and Alghamdi (2014) conducted qualitative and quantitative research.

My study also employs a mixed methods approach. While the aforementioned descriptive studies did include migrant workers with varying lengths of residence and linguistic backgrounds, the potential impact of L1 and LoR has not been examined in these studies. Beyond analysing the effect of L1, LoR, amount of Arabic and L1 used and age on migrant workers' L2 production, as done by Albaqawi (2020), Aljutaily (2018) and Almoaily (2012), I will also test the impact of linguistic constraints and other contextual predictors, such as motivation, informal exposure, formal instruction, migration identity, educational level in L1, proficiency in Arabic and language anxiety.

Albaqawi (2020) and Almoaily (2012) discovered that selected social variables minimally impact migrant workers' use of morphosyntactic features. This negligible effect means the independent variables influence a small number of the linguistic variables. The relevant findings from these studies are presented in chapters 5 and 6 of the analysis. While the participants in Almoaily's study are males and the participants in Albaqawi's study are females, she subsequently compared her result with the result obtained by Almoaily study to assess any sex effects. Both studies were conducted in the Riyadh region in general. My study is the first study that includes both male and female migrant workers in Madinah, thus enabling me to test the impact of speaker sex on migrants' language choices. My research is also different from these two studies in the sense that it investigates not only the morphosyntactic features but also the phonological feature. Similarly, it is different from Aljutaily's (2018) research, which focused on certain marked consonants—rare sounds—by examining an unmarked sound, a common and shared phoneme across languages. In addition, in general, my study is distinguished from the studies completed by Albaqawi (2020), Aljutaily (2018) and Almoaily (2012) by way of testing the impact of much larger range of independent variables on migrants' L2 Arabic usage.

As a recommendation for future studies, Almoaily suggested examining how the substrate languages may influence the phonological system of the migrant workers' L2 Arabic variety. Moreover, he proposed comparing the productions of males with females to identify any sex effects on variation. My research builds upon Almoaily's study and attempts to fill in these gaps by examining the phonological variable /f/ besides the definite article and the coordinating conjunction. It will also include sex as an independent variable alongside other internal and

external factors, so as to discover if the L2 varieties of the migrant workers are interlanguage and influenced by their L1 or a pidgin. Additionally, he suggests testing other substrate languages which are linguistically divergent from the Indo-Aryan or Dravidian languages. For instance, Malayalam with Indonesian or Tagalog. While Albaqawi (2020) includes Tagalog in her study, she has not examined its effect as a substrate language on the production of /f/ which my study examines. Also, Aljutaily (2018) indicated that the realisation of Arabic marked consonants might be influenced by the participants' motivation which his study did not examine, although it is suggested for future research. Similarly, Albaqawi and Oakes (2019) believe that there are other factors that may influence the participants' choice between the variants of the linguistic features that they examined in their study, such as motivation. My study will examine motivation as an independent variable that may impact the migrants' use of linguistic variables under investigation. Similarly, as far as I am aware, most studies examining migrant workers discussed the migrants' Arabic variety from the perspective of pidgin language. However, I will investigate this variety from the L2 acquisition perspective too. Specifically, it considers the possibility that their varieties can be interlanguage. Therefore, my study attempts to fill these gaps in the literature and contribute to the field of sociolinguistics in general and Arabic sociolinguistics in particular. It will provide new insights for scholars of language variation. Additionally, it will add to the existing knowledge of pidgin languages and L2 acquisition.

Crucially, the motivation for examining the verbal and nominal systems and studying the impact of a series of internal and external factors on the use of the definite article /ʔal-/ 'the', the coordinating conjunction marker /wa/ 'and', and the production of /f/ is to answer important questions about categorising the migrant workers' varieties. The main aim of this study is to investigate whether the L2 Arabic varieties of migrant workers who have Bengali, Hindi-Urdu or Tagalog as their L1 in Madinah, Saudi Arabia, are pidgin or interlanguage varieties.

1.3 Research Questions

My research seeks to address the following questions:

1. Do migrant workers use features of HA categorically or variably? If variably, what are the linguistic predictors conditioning the variation?

2. Do migrant workers with different heritage language backgrounds, i.e., Bengali, Hindi-Urdu and Tagalog, share the same underlying grammar of variation regarding the selected linguistic features?
3. Do migrant workers with different heritage language backgrounds use their own distinct system of Arabic which might be affected by their L1? If yes, what are these distinguishing characteristics?
4. Do social factors, including sex, amount of Arabic and L1 used, motivation, length of residence, informal exposure, formal instruction, migration identity, age, educational level in L1, proficiency in Arabic and language anxiety, influence the migrant workers' variable use of selected HA features?

Based on the literature review in the interlanguage (see 2.5.1 - 2.5.12) and pidgin language (see 5.1, 6.1 & 7.1), I will interpret the evidence that my study uncovers as follows:

1. If the results reveal that there is variation in the Arabic spoken by migrant workers with different heritage language backgrounds and that these are impacted by characteristics of their respective heritage languages, we have reason to assume that their variety might be an interlanguage.
2. If we discover that social factors affect the migrants' use of HA features, then we can predict that their Arabic variety is an interlanguage.
3. If the results show that there is a grammatical simplification in the variety of the migrant workers, then we will assume that their variety might be a pidgin language.
4. If we find that social factors do not influence the migrants' use of HA features, then we can expect that their Arabic variety is a pidgin language.

1.4 Organisation of the Thesis

This thesis is composed of seven chapters and organised as follows:

In Chapter 2, language transfer, different forms of transfer and theories pertaining to the acquisition of a second language are discussed. This is due to the fact that the interlanguage may be categorised as an outcome of the Arabic varieties that are employed by migrant workers in

Madinah. In light of the possibility that the L1 of the migrants may have an effect on the HA variables that they produce, it is of the utmost importance that this information be considered. In addition to providing the context for the internal and external predictors of variation, Chapter 2 includes a variety of research that supports the hypothesis that contextual factors have a significant influence on the use of L2 features. Hence, if the target variables utilised by the migrant workers in my research are significantly affected by contextual factors, their L2 Arabic might be better categorised as interlanguage varieties. Furthermore, investigating relevant research concerning pidgins is crucial to this chapter as it represents an additional potential result of the varieties exhibited by migrant workers in Madinah. In particular, it examines the stages of pidgin evolution and the general linguistic characteristics of pidgin. This extensive analysis not only enhances our comprehension of the evolution of pidgin languages but also offers valuable insights into the possible attributes displayed by the Arabic varieties being studied. The relevance of the literature that has been reviewed to the overarching research question demonstrates its significance. The simplification and reduction of linguistic features that were noted in previous research on migrant workers in the Arabian Gulf countries provided a foundation for predicting the results that would be obtained in my study. Migrant workers may delete morphosyntactic variables, replace the phonological variable /f/ with alternative variants, and discover that contextual predictors have a minimal impact on the utilisation of HA features. These are a few of the consequences that are expected to occur. It is also possible to categorise the Arabic varieties that are employed by migrant workers in Madinah as pidgin, taking into consideration all of these linguistic traits as a whole.

Chapter 3 describes my research methods. It establishes the fundamental phases concerning the research process, containing the ethical considerations that are specific to participants who are vulnerable. My study has adhered to ethical principles and protected the rights of migrant workers. Chapter 3 will also explain the procedures implemented to foster a rapport with the participants, thereby facilitating transparent dialogues concerning their thoughts and experiences. To guarantee that migrant workers are exposed to the same linguistic input and to assess the effect of migrants' L1 on their use of HA features, the choice of participants was conducted in accordance with specific inclusion and exclusion criteria. Additionally, Chapter 3 reveals that the two strategies related to participant recruitment that supported the research of vulnerable and difficult-to-reach groups are judgment sampling and snowball sampling. Additionally, the chapter will demonstrate

that sociolinguistic interviews and quantitative questionnaires were utilised to acquire data for my study. The first method made it possible to collect the tokens that were related to the variables that were being investigated, while the second method played a role in the classification of people according to the social factors that may have an impact on their use of HA features. Owing to the COVID-19 epidemic, the interviews were conducted via Zoom. The interviews were transcribed using ELAN and the tokens were extracted with AntConc. Statistical analysis is the last section of the chapter.

Chapter 4 illustrates how the oldest speaker in each language group - Bengali, Hindi-Urdu and Tagalog – uses the three HA morphological features: subject-verb agreement in gender, number and person; noun-adjective agreement in gender, number and definiteness; and number marking on nouns. In Chapter 4, some GPA studies that examined the verbal and nominal agreements in the speech of the migrant workers are presented. If my descriptive analysis exhibits similar results, we can predict that the Arabic varieties of the migrant workers would be better described as a pidgin. These results specifically include simplification in the use of HA morphological features, for example, a lack of verbal and nominal agreements. Chapter 4 also describes the utilisation of verbal and nominal agreement in HA and the L1s of the migrant workers. Based on how the migrants' L1 is different, their use of the HA morphological features might be influenced. This may possibly be a sign that the Arabic varieties they speak are interlanguage. The qualitative analysis in Chapter 4 reveals that the three migrant workers apply a simplified variety of Arabic characterised by a lack of morphological agreement.

Chapters 5, 6 and 7 include the results of my examination of how migrant workers in Madinah use three specific language elements: the definite article, the coordinating conjunction marker and the phoneme /f/. Each chapter provides a concise overview of the context that shapes the selection of each HA feature. This selection is prompted by the results of numerous quantitative and qualitative research conducted on the Arabic varieties spoken by the migrants in the Gulf region. In the same way, each chapter describes the utilisation of a particular linguistic variable in HA and the L1s of the migrant workers. Depending on how the migrants' L1 is different, how they use the L2 Arabic variables may be affected. This may possibly be a sign that the Arabic varieties they speak are interlanguage. Throughout each chapter, both the general distribution of each linguistic

feature as well as the quantitative data analysis of that variable were disclosed. Regarding the definite article and the coordinating conjunction marker, migrant workers in Madinah mostly delete these variables. Although there is limited evidence exhibiting the significant effect of the contextual variables on the use of these two variables by migrant workers, variations across different language groups and linguistic variables appear to be more effective in supporting the interlanguage hypothesis in comparison to the null results which are in favour of the pidgin hypothesis. Concerning the use of /f/ by migrant workers who have Bengali, Hindi-Urdu and Tagalog as their L1 in Madinah, each language group exhibits different patterns of use for /f/ when speaking L2 Arabic. The independent variables mostly have a significant effect on the use of /f/ by the migrants, suggesting that the Arabic varieties of the migrants are interlanguage.

In Chapter 8, the results are discussed in relation to the research questions, the relevant literature review and a conclusion. Chapter 8 explores the evidence for interlanguage status and also considers the evidence that might support pidgin status. The conflicting evidence is assessed and the impact of social predictors across various linguistic variables and speaker groups is examined. In addition, Chapter 8 discusses the implications of the research on GPA studies and contact languages in general. It also illustrates the opposite trends in the results of the morphological features on the one hand and the morphosyntactic and phonological features on the other. The conclusions that can be drawn from the findings of this thesis are that the Arabic varieties spoken by the migrants in Madinah are located on a continuum, where the higher end may be closer to an interlanguag.

Chapter 2: Interlanguage and Pidgin

2.1 Introduction

This thesis investigates the L2 Arabic varieties spoken by migrant workers in Madinah, with a particular focus on individuals whose L1 is Bengali, Hindi-Urdu or Tagalog. Determining if their varieties are pidgin or interlanguage is the fundamental purpose.

The significant impact that contextual predictors have on the use of L2 features is the motivation for the decision to concentrate on interlanguage. This chapter begins with a definition of the concept of 'interlanguage,' because the varieties of migrant workers in Madinah might be categorised as such. To gain a better understanding of the possible influence that the two forms of language transfer have on the use of HA features by migrants, this chapter studies both. In this chapter, we examine theories pertaining to L2 acquisition that establish a theoretical framework to understand the possible effects of transfer on the migrants' Arabic varieties. The chapter considers both internal and external factors when addressing language variation. Similarly, the chapter provides a number of studies that confirm how these factors significantly impact the production of L2 variables. If these factors have a significant effect on how migrants employ HA features, it might be better to refer to their L2 Arabic varieties as an interlanguage.

As a pidgin language could be the other outcome that describes the L2 Arabic varieties of migrant workers in Madinah, the chapter focuses on specific topics related to pidgin. It provides clarification on the stages of pidgin language development, which improves our general comprehension of the evolution of pidgin languages. By presenting the general linguistic features linked to pidgin languages, this chapter offers significant perspectives on the potential characteristics that may exist in the Arabic varieties being researched. To emphasise the relevance of the literature review to my study, it concludes with a review of previous research on migrant workers' varieties in the Arabian Gulf countries. These studies assert that the linguistic structure of migrant workers' varieties, compared to Arabic, is characterised by simplicity and reduction. The anticipated outcomes, for instance the deletion of morphosyntactic features and the substitution of the phonological variable /f/ with other variants, along with the negligible impact of contextual predictors on the use of L2 features by the migrant workers, may imply that their Arabic varieties are pidgin.

2.2 Interlanguage and Language Transfer

The concept of an interlanguage was introduced by Selinker (1972) who asserted that when language learners are at any stage of development, they develop a structured linguistic system that is independent of their L1 and L2 (Romaine 1988: 210). Bayley (2007: 134) defines interlanguage as a system 'which share[s] features of the learner's first language and the target language but [is] fully explained by neither'. According to Matras (2009: 74), this phenomenon pertains to the peculiar utilisation of particular linguistic structures by a single learner in the target language. The use of this idiosyncratic language may alter in various circumstances, including formal and informal settings, and undergo modifications as the learning process develops (Matras 2009: 74). Selinker (1974: 37) outlined five principal cognitive processes that can shape and influence the interlanguage system, namely: language transfer, transfer of training, strategies of L2 language learning, strategies of L2 communication and overgeneralisation. This section will focus on language transfer because I hypothesise that if migrant workers' Arabic varieties are interlanguage, language transfer will exhibit a significant effect on their variable use of the L2 features.

Even when adult learners are extremely successful in acquiring a second language (L2) or additional language, they are often distinguishable from people who speak it as their first language (L1) by their 'accent' (Matras 2009: 72). To analyse this phenomenon systematically, Weinreich (1953) argued that adult learners use the knowledge of their L1 to process the L2, at least initially. Given that they can express more complex ideas in their L1, this can frequently serve as a base that they can build on when they acquire the L2. The L1 structure can be useful to the learner in understanding the L2 corresponding structure. Romaine (1988: 206) states that in regard to the theory of transfer, there is a propensity for the interlanguage development to be shaped by the features of the learner's L1. Trask (2000) and Aikhenvald (2002) indicated that language transfer occurs when an individual unintentionally carries over linguistic features from his or her mother tongue to L2.

To refer to cross-linguistic influence, researchers use several labels, such as linguistic interference, language transfer, native language influence and the role of the mother tongue (Odlin 2003: 436). In the literature pertaining to L2 acquisition, it is consistently termed transfer or interference. However, in the studies of pidgin, creole and other contact situations, it is called substrate influence

(Alshammari 2018: 100). Although transfer and interference are used interchangeably in L2 studies, Odlin (1989: 26) believes that the term transfer does not generally imply interference and that it only refers to negative transfer. Doughty and Long (2003: 439) are in agreement with Odlin and asserted that 'transfer should not be equated with interference'. In my study, the term 'L1 influence' is employed as a catch-all term encompassing all the aforementioned phenomena. Language transfer can impact all the subsystems of the language (Odlin 2003: 437; Velupillai 2015: 144). Velupillai (2015: 145) reports instances of L1 German speakers adding the inflectional marker -s to the English word *hair* to make it plural. This is in accordance with the pattern that is employed in German. This is in contrast to the English language, where the plural form of *hair* does not have an inflected form, as seen in example (1) rather than example (2).

(1) * I have washed my hair**s**
I have wash.PST my hair.**PL**
'I have washed my hair'
(Velupillai 2015: 145)

(2) I have washed my hair
I have wash.PST my hair.SG
'I have washed my hair'
(Velupillai 2015: 145)

My research will determine if language transfer has any impact on morphosyntactic and phonological features in the varieties of migrant workers who speak Bengali, Hindi-Urdu and Tagalog as their L1 in Madinah. I am analysing transfer at these two levels because, as far as I am aware, previous quantitative studies on Arabic migrant workers' varieties have focused only on one of them, either morphosyntactic features or phonological features (see Albaqawi & Oakes 2019; Albaqawi 2020; Aljutaily 2018; Almoaily 2012). Although I am only analysing a small number of variables, my approach may help me to evaluate whether the migrants' L1s influence features at both of these levels in L2 or just one of them.

Traditionally, in the learning process, transfer and interference are considered negative outcomes and in the acquisition of the target language structure, they are failed manifestations (Matras 2009: 74). But because they do not lead to misunderstanding and breakdown in communication, they can be considered as enabling factors that assist language learners to use their repertoire of linguistic forms and maintain communication.

Transfer theory asserts that learners who speak different L1s will learn the same L2 in different ways because of the distinct influence of transfer from their respective L1, which can be positive transfer or negative transfer (Romaine 1988: 207). When the L1 and L2 are comparable, learners' L1 knowledge would facilitate learning the L2 (Romaine 1988: 207). By contrast, when the L1 and L2 differ, learners' knowledge of the L1 would conflict with the L2. Both these types of transfer will be discussed in the following paragraphs because if the L1 influence of the migrants in my study impacts their variable use of HA features, we may categorise their Arabic varieties as interlanguage.

2.2.1 Positive Transfer

According to Odlin (1989), positive transfer may occur when there are similarities between the learner's L1 and the target language. This is because the structures from the learner's L1 assist with the acquisition of the target language structure. For instance, Oller and Redding (1971) discovered that L1 speakers of German, French, Italian, Portuguese and Hebrew performed better in the use of the English articles than L1 speakers of Korean, Thai, Japanese, Persian and Chinese. The significant difference between the two groups might suggest positive transfer, as the L1 of the former group has articles that are equivalent to the ones in English, but the latter group do not. The similarities between languages, for instance, in the vowel systems, syntactic structures, vocabulary and writing systems can reduce the time it takes to acquire the target language and facilitate its acquisition (Odlin 1989: 36). Concerning my research, a specific group of migrants may employ the characteristics of the target L2 (HA) if their L1 contains phonemes and structures that are similar to those of the HA. This may well be be an indication of a positive transfer from their L1.

2.2.2 Negative Transfer

This form of transfer involves deviation from the target language's norms due to the impact of the L1, which possesses distinct features (Odlin 1989: 36). Alhamadi (2015: 162) discovered that Saudi students in Saudi Arabia who learn English as their L2, had difficulty achieving native-like preposition usage. When the English preposition had no direct Arabic equivalent, errors seem to be evident. For instance, although saying example (3) below without the preposition 'with' is permissible in Arabic, it is not in Standard English. This negative transfer occurs because students literally transferred the Arabic prepositional structure into English.

(3) * I do not agree Ø the concept
I do NEG agree Ø DEF concept
'I do not agree **with** the concept'
(Alhamadi 2015: 162)

Odlin (1989: 36) provided evidence that while negative transfer is commonly associated with production errors, additional methods exist by which the performance of a second language learner may deviate from that of native speakers. If a learner believes that a specific structure in the target language is very different from their L1, they might avoid using this structure. This is known as underproduction (Odlin 1989: 37). For instance, compared to Spanish and Persian students who learned English as their L2 and whose L1s have a similar relative clause structure to that observed in English, L1 speakers of Japanese and Chinese tended to use fewer relative clauses and therefore they produced fewer errors (Schachter 1974). In contrast, occasionally, as a consequence of the underproduction of one structure, the overproduction of another form might appear (Odlin 1989: 37). For instance, Japanese learners of English may overproduce simple sentences to avoid using relative clauses and by doing this they might violate the expectations of English written prose.

Substitutions and calques are types of production errors that might arise from language transfer. The first type includes using the forms of the L1 in the target language (Odlin 1989: 37). For example, a Swedish learner of English writes the word *bort* 'away' in the following sentence.

(4) * I must go **bort**
I must go away
'I must go away'
(Ringbom 1986: 157)

Alternatively, calques refers to errors that closely resemble the structure of the L1 (Odlin 1989: 37). Andersen (1979) examined English possessive constructions in the writing of Spanish students and discovered that as opposed to 'Carmen's porch', which is the more natural phrase in English, they use 'the porch of Carmen'. This refers to the impact of the Spanish word order *el balcon de Carmen*.

2.3 Theories about Second Language Acquisition

There are three possibilities as regards the degree of transfer from L1 to L2: no transfer, partial transfer and full transfer. It appears that transfer is more prevalent in the initial stage of acquisition rather than the final stage (Romaine 1988: 210). This section will illustrate these theories.

2.3.1 No Transfer

From this perspective, L1 grammar does not shape L2 grammar (White 2000: 135). Instead, it is assumed that universal grammar shapes the initial stage of L2 acquisition. Thus, the initial stages of L1 and L2 acquisition are the same. In this position, it is predicted that in the L2 final stage, the L2 learners' grammar is similar to that of the of L1 speaker. In other words, the ultimate attainment of L2 learners ought to resemble that of native speakers (White 2000: 136 & 144). Although several advocates of this position, for instance, Flynn and Martohardjono (1994) and Epstein, Flynn and Martohardjono (1996; 1998) 'exclude the L1 grammar properties from the interlanguage representation, they nevertheless assume some role for the L1, without specifying or clarifying its precise status' (White 2000: 135). Therefore, this position is inconsistent.

2.3.2 Full Transfer

White (1989) was the first researcher to introduce the earliest version of this position by investigating the pro-drop parameter. Her work shows that, learners of L2 start with the L1 parameter values and subsequently, they reset them to the L2 data. Schwartz (1998) and Schwartz

and Sprouse (1994; 1996) extended this notion by means of 'full transfer hypothesis' and assert that while the starting point for the acquisition of L1 and L2 are different, they are the same in relation to the involvement of the universal grammar (White 2000: 136). Thus, the universal grammar and the L1 'are complementary sources of knowledge that guide interlanguage development' (Montrul 2000: 232). Moreover, the L1 grammar is believed to shape L2 grammar at the initial stage of the acquisition process (White 2000: 136). According to this hypothesis, all L1 properties such as lexical projections, feature values, parameter settings, as well as functional structures transfer to the L2. For instance, both lexical (in this example, VP) and functional (NegP) projections are headed differently in Turkish and English. Whereas Turkish is verb final and Neg final, English is verb initial and Neg initial. In a case study of a single individual, Haznedar (1997) established that the English word order production of Erdem — a speaker they analysed — is virtually 100% head-final throughout the first three months of recording, indicating the transfer of headedness from his L1 Turkish. Erdem switched both the VP and NegP's headedness to English word order patterns in the fourth month, yielding consistently head-initial utterances. It is not necessarily predicted that L2 learners will converge on the L2 grammar in the final state, for the reason that their L1 grammar acts as a filter that might prevent them from noticing specific aspects of L2 input. Hence, this leads learners to fossilise 'at a point short of native-like competence' (White 2000: 137).

2.3.3 Partial Transfer

In this position, the initial stage of the L2 learning draws on not only the L1 properties but also universal grammar, concurrently (White 2000: 137). Only part of the L1 grammar is utilised in the L2 grammar, which contrasts with Full transfer. In the literature, there is debate about what parts are transferred from L1 and what parts are not transferred (Sabourin, Stowe & De Haan 2006). White (2000: 138) provides an overview of this discussion. In relation to the acquisition of syntax and according to the 'Minimal Trees Hypothesis', Vainikka and Young-Scholten (1994; 1996a; 1996b) argue that initially, lexical categories, such as NP and VP are transferred only from L1, although functional categories such as CP, DP and IP are assumed to be completely absent. In response to the L2 input, functional categories are developed gradually in a similar way to the acquisition in the L1. It is expected that learners of L2 may converge on the grammar of the target language in the L2 final stage. Schwartz and Sprouse (1996) argue against Vainikka and Young-

Scholten (1994; 1996a; 1996b) and claim that the 'Minimal Trees Hypothesis' fails to provide an explanation for ' S V Adv O' data for native speakers of French who acquire English. Schwartz and Sprouse provide evidence based on the adverb placement studies of White (1991a; 1991b; 1992). While the correct order in English is 'S Adv V O' for certain kinds of adverbs, it is 'S V Adv O' in French. Learners tend to produce 'S V Adv O' in English. This suggests a possible transfer of the French verb movement into English grammar given that the verb is positioned before the adverb, which deviates from the conventional English word order. This indicates that transfer has an effect on the functional domain of French because English does not raise the verb to a functional head; nevertheless, French does. Eubank (1994) proposes that not only lexical categories but also functional categories are transferred from L1 to L2 in the initial L2 stage. Conversely, functional features that encode agreement and tense are initially unspecified, inert or valueless. This is a concept that is investigated in the context of partial transfer. Within the framework of their 'Minimal Trees' methodology, Vainikka and Young-Scholten (1994; 1996a; 1996b) shed light on this phenomenon. Furthermore, the concept of 'Valueless Features' is explored by Eubank (1994). It is predicted that eventually, functional categories might be specified as regards the L2 feature values. White (1996) claims that not only are L1 lexical categories and functional categories transferred in the initial L2 learning stage where possible, but also feature values. Nonetheless, there will be cases where the L1 grammar cannot shape the L2 initial theory, such as the acquisition of English speakers of the French clitics.

2.4 Language Variation

Language variation is an essential element in L2. It is possible for any language to be variable, but L2 acquisition is particularly so (Regan 2013a: 272). Variation in L2 acquisition can be rule-governed and 'systematic', and it can also be free 'non-systematic' (Romaine 2003: 410). L2 variation starts off as free but becomes rule-governed over time. Ellis (1992) gives the example of a 11-year-old Portuguese learner of English who uses this instance of non-systematic variation: he used pre-verbal negation such as 'No look my card' and *don't*+ V such as 'Don't look my card' in random fashion. At the beginning of the acquisition of negation, he generalised the use of pre-verbal negation. Subsequently, when *don't*+ V entered his repertoire, he used it with *no* in free variation. This situation continued until other forms like *won't* and *can't* were added to his repertoire. At that point, his variation became systematic. Another illustration of rule-governed

variation is found in Ellis's (1992) work, where a learner added the suffix -s to present-tense verbs with third-person singular pronouns like in ‘he eats turkey’ but not with third-person singular nouns like in ‘John eat turkey’.

Variationist sociolinguists believe that variation is systematic and structured rather than random (Labov 1994: 2001), and variationist sociolinguists seek to identify and describe structured patterns of variation as well as how they are learned and employed. Numerous studies of language variation and change conducted thereafter have confirmed this finding, such as the study by Mackenzie (2019). And there are many works that explain how and to what extent non-native-speaking migrants acquire native speakers’ linguistic variation patterns. See, among others, Adamson and Regan (2001), Drummond (2010), Mougeon and Rehner (2001) and Uritescu *et al*. (2004). These researchers show that migrants can acquire native speakers’ variant repertoires but the constraints that govern their choice of variants may different from those of L1 speakers .

Tarone and Parrish (1988: 36 - 37) agreed with Littlewood's (1981) view that there are three factors that impact systematic variation in interlanguage varieties: linguistic context, social/external context and psycholinguistic context (Chung 2011: 182). The next section will focus on internal and external variation. I will argue that if these predictors affect the migrant workers’ use of HA features, we may conclude that their Arabic varieties are interlanguage.

2.5 Internal and External Variation

Internal and external variation are two distinct types of variation. In internal variation, the realisation of a linguistic variable is affected by linguistic factors (Romaine 2003: 411). For instance, the following segment has a significant effect on TD-deletion in British English (Baranowski & Turton 2020: 19). While a following vowel prevents deletion, a following consonant encourages it. Moreover, a following pause indicates intermediate rates. Analysing the effect on observed patters of migrant workers’ linguistic variation of internal predictors is essential to my study: if I discover that they have a significant effect on migrant workers’ variable use of HA features, this may indicate that their Arabic varieties are interlanguage.

Migrants can acquire some internal variation from the L1 speakers, whilst they can also produce their own constraints different from those of L1 speakers. For instance, Edwards (2011) examined how seven native Mandarin Chinese speakers in the US acquired the /t, d/ deletion patterns in word-final consonant clusters of English. The results confirm that, compared to the native speakers, the L2 speakers had not yet demonstrated a target-like use of /-t, -d/ deletion patterns across all constraints. Research on American English native speakers reveals that the strongest constraints that govern the /t, d/ deletion are grammatical conditioning; word category, the following phonological environment and the preceding phonological environment (see Labov 1967: 1989). The findings of Edwards indicate that while grammatical conditioning is not a significant factor, following phonological environment and preceding phonological environment play significant roles in influencing the patterns of /t, d/ deletion. This result supports the findings in the study by Bayley (1996) which was conducted in the US with Mandarin Chinese speakers who had been exposed to American English. Adamson and Regan (1991) discovered that although Cambodian and Vietnamese immigrants who are L2 speakers of English acquire some similar language internal constraints to American English speakers, they also develop a few different constraints with respect to variation. (-ing)-variation is conditioned by grammatical factors of the word that contains (ing) in the speech of the native speakers and non-native speakers, such as progressive and gerund. The variant [in] is favoured in their speech when the word ending in (ing) is a progressive and periphrastic future. However, this variant is also applied in preposition and words that can function as nominals, for example in *nothing* and *something*, in the speech of the non-native speakers, while native speakers rarely use [in] in these two-word categories. It is obvious from these two studies that when there is internally-conditioned variation in the native speakers' variety, immigrants may produce linguistic constraints partly similar to the natives but also develop new ones. Thus, my study will test if linguistic factors have an impact on the migrant workers' variable use of HA features.

With externally conditioned variation, variation between two or more variants is constrained by social or external factors for instance, ethnicity and region, where, for example, -t/ d deletion is less frequent in the speech of whites than in the speech of African Americans. There are many other external variables that may impact the speech of L2 learners, for example age, sex and style. The following subsections are in relation to the predictors that are examined in my research.

This literature review may help me discover the status varieties of migrant workers in Madinah, determining whether they are an interlanguage or a pidgin. Each of the following predictors are individual-level factors. Thus, if they correlate with the production of L2 features, it indicates that individual differences are more relevant than group norms.

2.5.1 First Language

One of the social factors that may shape migrants' L2 variety is their L1. Mougeon, Rehner and Nadasdi (2004: 426) refer to the importance of this variable in their study. Their results indicate that participants' different language backgrounds, which in their case are Spanish, Italian, English and non-Romance language, influence the variations they exhibit in French. Participants who speak Spanish and Italian employed five French variables, specifically the negative particle *ne*, *travail*, *alors, seulement* and *nous* more frequently than the rest of the participants because they had related equivalents in their L1s. English participants used only one variant the most because it has a similar equivalent in English: the restrictive adverb. There are also other researchers who have investigated the effect of L1 on the participants' L2 production. A number of these studies showed L1 background has an impact such as Mougeon and Rehner (2001), while others like Naro (1978) and Manessy (1977) discovered that L1 did not show any influence.

2.5.2 Sex

In the study of linguistic variation, the sex of the speaker is one of the most important social variables (Eckert 1989). [1] Labov (1990; 2001) presented three principles, which he terms 'the

[1] It is worth noting that researchers differentiate between sex and gender. Although the former is a physiological trait that distinguishes males from females, the latter is considered not only a social but also a dynamic construct (Butler 1988; Cheshire 2002; Eckert and McConnell-Ginet 1992). It is something that speakers perform or do in social situations. However, it is not something static that people have and it can neither be analysed without interaction with other components of social identity nor isolated from them (Eckert and McConnel-Ginet 1992: 433). Ellis (1994: 202) mentioned that the term 'gender' is preferred by several sociolinguists because the social construction of males and females is what gender emphasises. In variationist literature, both concepts; sex and gender, can be found (Cheshire 2002: 423). While occasionally they are used to differentiate biological features from social factors (see for instance, Chambers 1995), on other occasions they are used in a haphazard manner. In my study, the term sex will be used.

> It is difficult to keep the two concepts apart… [and] [c]urrent thinking in the humanities accept, in any case, that the dichotomy between sex and gender cannot be maintained, seeing the body and biological processes as part of cultural histories
> (Cheshire 2002: 423).

gender paradox' (2001: 292). In his work, Labov (1990: 205-206; 2001: 263) clarified the effect of gender on three different categories of changes which are 'stable sociolinguistic variables and change from above or change from below'.

While these principles were established on the basis of the analyses of L1 varieties, they might be relevant to the analysis of L2 variation. Gender pattern is part of structured variation. Therefore, if I encounter these group norms, this may indicate that I have found an interlanguage. The principles will be illustrated in the following paragraphs.

First, stable sociolinguistic variables exhibit stability; they do not undergo change in progress. Decades of research have provided evidence in support of a sex-effect for these variables that Labov (2001: 266) summarizes as Principle 2: 'for stable sociolinguistic variables, women show a lower rate of stigmatised variants and a higher rate of prestige variants than men'. The alternation between [n] and [ŋ] among males and females for /ing/ which is one of the most widely studied variable in English reveals that females produce the prestige form [ŋ] more than males. This appears in many studies, for instance the studies conducted by Fischer (1958) in New England, Labov (1966a) in New York, Cofer (1972) in Philadelphia, Trudgill (1974) in Norwich and Woods (1979) in Ottawa.

Second, Labov (2001: 274) defines change from above as change above the level of conscious awareness. In this type of change, people are usually aware of the standard (prestigious) variant and its non-standard (stigmatised) co-variant and they can more or less consciously alterate between them. The new linguistic prestige is introduced from outside the speech community. This particular type of change is socially motivated (Labov 2001; Tagliamonte 2012). Regarding this kind of change, Labov (2001: 274) refers to Principle 3 which asserts that 'in linguistic change from above, women adopt prestige forms at a higher rate than men'. For instance, Gal (1978) conducted a study in a bilingual community in Oberwart, Austria. Previously, the town was located in Hungary. Language shift is clearly evident in Gal's study of German and Hungarian bilingualism in the direction of use of German only. While Hungarian became more and more associated with the life of the peasant, German was a high variety. Young, female peasants consciously favoured using German more than the young men and older people. They rebelled

against their low status and preferred to make their position equal to the working status of the German-speaking population. Their marriage choices were increasingly toward people who spoke German and eventually, German was taking over.

The third type of change is change from below: i.e., below the level of conscious awareness. It is primarily motivated by internal factors (Labov 2001). Regarding this sort of change, Labov (2001: 292) states Principle 4 which is 'in linguistic change from below, women use higher frequencies of innovative forms than men do'. For instance, Wolf and Jiménez (1979) conducted a study in Buenos Aires to investigate the devoicing of /ʒ/. They discovered that there is a change in Spanish and a strong shift toward the devoiced variant by younger females.

Ellis (1994: 202) believed that Labov's (1990) principles suggest that in learning an L2, females might do better than males because they tend to exclude linguistic forms from their interlanguage that deviate from the norms of the target language. Likewise, they tend to be more open to emerging linguistic features in regard to L2 input. This does not necessarily imply a specific level of difficulty for men and women in learning group norms. Nevertheless, this demonstrates how flexible and willing language learners are, particularly women, to accept and use new features of language as they progress in their language acquisition. Thompson (1991) discovered that sex has a significant effect on pronunciation accuracy. For example, Russian women who are L2 learners of English received higher ratings in L2 pronunciation than men. Similarly, Adamson and Regan (1991) established that Cambodian and Vietnamese females who speak English as their L2 use [ŋ], which is the prestigious form of /ing/, more often than males. Major (2004) also noticed that male L2 speakers of English whose L1 are Japanese and Spanish use the forms of less prestigious or casual speech more than females. These studies suggest that the language patterns of males in an L2 context might be influenced by whoever they are in contact with.

2.5.3 Amount of Arabic and L1 Used

One of the factors that may affect migrant workers' use of HA features is the amount of Arabic they use and the amount of use of their L1. The significance of this variable is evident in various L2 studies. For instance, Flege, Yeni-Komshian and Liu (1999) compared two groups of native Korean speakers who live in the US. While the first group used Korean infrequently and English

often, the other group rarely used English and Korean often. The English pronunciation of the first group was significantly better than the other group. Similar results were found in the studies of Flege, Frieda and Nozawa (1997) and Piske, MacKay and Flege (2001) with regard to frequent L1 use and its negative effect on L2 pronunciation accuracy. Additionally, Major (2014) discovered that average L1 use is negatively correlated with the participant's English proficiency in L2.

It might be natural to expect that as L2 use increases, L2 production and performance might improve (Drummond 2010). However, there are studies that suggest otherwise. Flege and Fletcher (1992) established no significant correlation between the frequency of English use as an L2 and the degree of L2 accent among native Spanish speakers in the United States. Thompson (1991) observed a simple correlation between how much native Russian speakers used English and their L2 accent. However, the multiple regression analysis failed to establish a statistically significant correlation.

2.5.4 Length of Residence

Length of residence (LoR) measures the number of years that people spend in a society where its predominant language is L2 (Piske, MacKay & Flege 2001). A number of studies have established that LoR significantly affects the use of L2 features. For example, Drummond (2011) investigated variation in /t/ among 40 Polish adults who learnt English as L2 and have been living in Manchester for varying lengths of time. He discovered that the glottal replacement is extremely low at (2.9%) for speakers whose LoR is 0-24 months compared to those whose LoR is 25-48 months and who have a mean glottal replacement of (12.7%). Regan (2013b) found that LoR was a significant predictor of the acquisition of *ne* deletion in French by Polish migrants. To be more specific, those who had lived in France for a longer time were more likely to eliminate *ne* in comparison to those who had been there for shorter periods. Flege, Bohn and Jang (1997) observed that non-native adults whose average LoR in the US was 7.3 years, produced the English vowel /**ɪ**/ more accurately than those whose average LoR was 0.7 years.

There are also a few studies that show LoR does not have an influence on L2 performance. Flege (1988) noted that there was no significant difference in foreign accents among Chinese adults whose LoR in the US was 1.1 and 5.5 years on average. Flege (1993) also did not observe

significant differences in English phonetic segment production between Chinese adults who stayed in the US for between 1.2 and 5.1 years. Moreover, neither did Flege, Munro and Skelton (1992) observe any significant differences among native Spanish adults with an average LoR of 0.4 and 9.0 years, or among native Mandarin adults whose average LoR was from 0.9 and 5.5 years. Studies on L2 morphosyntax have likewise been unable to demonstrate that L2 performance improves with increasing LoR.

Flege and Liu (2001) believe that the amount of L2 input is not an essential determinant of performance in the L2 and can therefore be a possible reason as to why LoR has not been determined to affect L2 performance. Given that the majority of the participants in these studies were late bilinguals, a different hypothesis is that L2 learning is restricted by a critical or sensitive period (Patkowski 1980; Scovel 1988). Flege and Liu also indicated that most immigrants who are adults and work at home might have little contact with locally-born native speaker individuals, and those who work outside their homes may or may not have regular contact with locally-born native speaker individuals: only 'immigrants who receive a substantial amount of native-speaker input' will acquire a LoR effect (Flege 2009: 7). Flege further explains why LoR plays a role in some studies but not in others. Flege cites different findings as evidence, such as: Flege *et al.* (2006), in which the difference in pronunciation between Korean children who lived in North America for 3 years and those who stayed for 5 years was not significant; Winitz, Gillespie and Starcev (1995) mentioned the case of a 7-year-old Polish boy who achieved native-like pronunciation. Due to his living environment, the native-speaker input that the Polish boy received was more than that of Korean children.

All of the above L2 studies demonstrate that the input that the L2 learners receive can be the reason behind the effect of LoR on their L2 production. Although I acknowledge that factors are related to each other, the complexity of the situation and that measuring LoR alone without also measuring exposure to the L1 fails to provide us with sufficient information, it is still essential to discover if LoR has a possible effect on the migrant workers' use of HA features.

2.5.5 Motivation

Motivation can be considered an essential factor that influences migrant workers' variety. It has been tested in many studies of L2 acquisition, such as Dörnyei (1990; 2001; 2003), Drummond (2011), Gardner (1985b), Oyama (1976), Purcell and Suter (1980), Suter (1976) and Thompson (1991). Abu-Rabia and Kehat (2004) established that in two of their case studies, those who speak Russian and English as their L1 were successful in acquiring Hebrew without an L2 accent. Despite the fact they were exposed to the L2 after puberty, they indicated that it was important for them to sound like native speakers. Abu-Rabia and Kehat also implied that the success in L2 pronunciation competence might be due to the participants' motivation. In contrast, other studies found that the motivational variable does not play a role in L2 pronunciation, such as Oyama (1976) and Thompson (1991).

Birdsong (2007) believed that if individuals have a strong motivation to sound like L1 speakers, this will not necessarily guarantee their success, as the studies of Bongaerts *et al*. (1997) and Moyer (1999) suggest. Both of them recruited highly motivated participants in their research to find out if late learners can achieve native-like pronunciation. In Bongaerts *et al*'s (1997) study, 13 British English native speakers rated the speech samples of Dutch learners of English who read six sentences three times. The raters used a five-point scale of judgment; 1 denotes that the accent is definitely non-native whereas 5 means the accent is definitely native. Bongaert *et al.* found that only five out of 11 participants received native-like ratings. In Moyer's (1999) study, the raters comprised four German native speakers who listened to four speech samples of each of the participants who were 24 English learners of German. The samples included reading a list of 24 terms, a list of eight sentences and a text paragraph. Moreover, the participants were provided with five topics to talk about for several minutes. A six-point scale of judgment was used, where 1= definitely native and 6= definitely non-native. The findings revealed that only one out of 24 subjects achieved a native-like rating. The participants in the study by Bongaert *et al.* (1997) were rated to be more successful in the pronunciation of L2 than the participants who participated in the study completed by Moyer (1999). Piske, MacKay and Flege (2001: 202) believe that the difference in the results obtained by these studies is due to age of arrival and the amount of L2 experience, which are difficult to assess post-hoc for the two studies.

There are two types of motivation that can influence the migrant workers' production. Seright (1985) suggested that if the desire of L2 learners is only to function and communicate in the L2, they will often fall far short of native-like usage, while if their goal is to be native-like, they will succeed to a greater degree in L2 attainment. Certain researchers refer to these two types as respectively as instrumental and integrative motivation. It is important to mention that they cannot always be disambiguated (Brown 2000) and many studies do not seek to differentiate them (Drummond 2010). The latter refers to the motivation of an individual who has a desire to be more like the L2 community members and feel a sense of belonging to them, whereas the former refers to an individual's desire to be proficient in a language for practical or utilitarian purposes (Krashen 1981; Thompson 1991), such as getting a better job, translation work, meeting school or university graduation requirements, achieving higher social status, reading technical material and requesting higher pay based on language ability (Hudson 2000). In my research, these types of motivation are combined to reveal if motivation in general has an impact on migrant workers' production of HA features. My study does not set out to distinguish between integrative and instrumental motivation.

Major (2014: 24) suggests that to some extent, the success of L2 learners can be the result of the degree of motivation that they possess; it includes their attitude toward the target language and the learning environment. Whereas positive attitudes yield positive behaviour, feelings and thoughts toward an object, negative attitudes produce the opposite - a negative response (Pratkains, Breckler & Greenwald 1989: 75). Drummond (2011: 125 & 221) ascertained that use of the Manchester STRUT vowel is significantly influenced by migrants' attitude towards Manchester. The shift towards the local variant of STRUT was higher among those who have a positive attitudes towards Manchester. Al-Musnad (2018) conducted a study in Riyadh, Saudi Arabia involving 109 female migrant workers who are nurses, with the aim of investigating their motivation and attitudes toward Arabic and its native speakers. To collect the data, Al-Musnad used a questionnaire based on Gardener's Attitude/Motivation Test Battery which is a study tool created to evaluate the primary affective factors implicated in L2 learning (Gardner 1985a). The results revealed that the participants were highly motivated as regards both types of motivation and moreover, that their integrative motivation was higher than the instrumental motivation. Regarding their attitude, while the participants had a more or less neutral attitude toward the native speakers, they had positive

attitudes toward learning Arabic. In contrast to Al-Musnad's study, my study will focus on testing the effect of the motivation and the attitude of migrant workers on their production of HA features.

2.5.6 Informal Exposure

Unlike formal learning, informal exposure is unplanned and does not comprise learning objectives (Werquin 2007: 24). Moreover, it might be spontaneous and may occur in any place and at any time (Eaton 2010: 17). Learners can have contact with L2 through various sources e.g. through technological advances, such as social-networking websites, digital games, internet-TV and mobile applications (Lyrigkou 2019: 237). Informal exposure is one of the factors that may help migrant workers in my study to acquire HA. Seeing as Arabic is a diglossic language, migrants may be exposed to different Arabic varieties. The three dependent variables under investigation in my study exist in the three varieties of Arabic, each exhibiting specific differences. The three types mentioned are classical Arabic (specifically the variety found in the Holy Quran), Modern Standard Arabic (a formal variety often used in media) and HA (an informal variety that is spoken by the inhabitants who live in Hijaz.

Kusyk and Sockett (2012) and Sockett (2014) conducted studies on French university students whose L2 is English and investigated the influence of watching online-TV on their comprehension and production of vocabulary. These studies are based on a corpus of fifty most common and frequent four-word chunks that are heard in five popular American television series. Both studies produced similar findings. There was a significant language difference between frequent viewers who reported watching English TV shows online more than once a week and occasional viewers on both the vocabulary knowledge scale; this scale, adapted from Paribakht and Wesche (1993), relates to measures of meaning recognition and on the use of chunks in written production. Araujo and Dinis da Costa (2013) and Costa and Albergaria-Almeida (2015) used the European Survey on Language Competences, assessing the reading, writing and listening skills of secondary school students in fourteen countries in Europe. The examination aimed to determine the proficiency of English as a foreign language, according to the Common European Framework of Reference (CEFR). Students at the Pre-A1 level were compared to those at the Basic User level (which consists of CEFR levels A1 and A2), and those at the Independent User level (which consists of levels B1 and B2) were compared to those at the Basic User level. They determined that exposure

to the media in the target language has a strong impact on English language proficiency. In general, watching films in a foreign language, especially without subtitles, has a positive effect on language proficiency.

An additional aspect of informal exposure that may affect L2 speakers is listening to songs in the L2. Schwarz (2013) conducted a study of 74 Austrian students who do not use English as their L1. The students' ages ranged between 13 and 15. Schwarz examined whether or not acquiring incidental vocabulary from English pop songs can occur. The results confirmed that the participants' performance on vocabulary assessments conducted before and after listening to songs differed statistically significantly. The mean score of the post-test was higher than the mean score of the pre-test. Similarly, Pavia, Webb and Faez (2019) noted that listening to pop songs contributed to vocabulary learning. To my knowledge, no L2 studies have examined the impact of this predictor on acquiring native-like usage patterns in phonology and morpho-syntax. As watching TV and listening to songs in the target language are useful vehicles of L2 acquisition in the previous mentioned studies, I predict that migrant workers who watch more Arabic TV, listen to more Arabic radio and also listen to the Quran will use more HA features. I also predict that as the participants are exposed to different types of informal exposure, their acquisition of HA features will increase.

2.5.7 Formal Instruction

Formal instruction is another independent variable that might play a role in L2 acquisition. Krashen *et al.* (1978) conducted a study of 116 students who were learning English as their L2 in the US and examined the effect of the amount of formal English study on their English proficiency. They found that proficiency in English is significantly related to individuals' years of formal instruction. Their finding replicates the results of Krashen and Seliger (1976) and Krashen, Seliger and Hartnett (1974) who also reported that more formal instruction is linked with greater English proficiency.

However, several studies report that formal instruction does not have an impact on L2 proficiency. To illustrate this, Patkowski (1980) investigated the English syntactic proficiency of 67 migrants in the United States who speak different L1s, including Spanish, Chinese, Polish, Bengali, Turkish,

among others. He calculated the hours of English formal instruction that each participant received and found that the impact of this variable on the English syntactic proficiency is negligible. In Manchester, Drummond (2010: 218) discovered that formal English instruction did not significantly affect Polish migrants' variable use of these linguistic variables: -ing, /t/, 'h' dropping and STRUT. This could be due to the fact that they had received the majority of their tuition in Poland before moving to the UK. The LoR results indicate that the influence of pre-arrival tuition is less significant than the effect of time spent in the United Kingdom in terms of acquiring local variation. The one aspect of formal instruction that would have been beneficial for the acquisition of local patterns of language variation is if the lessons held in Manchester had been conducted by a teacher who spoke with a local accent. As a result of the restricted data (only a few of the participants had taken lessons in the United Kingdom), it was impossible to ascertain the accents of the instructors. Consequently, it is still doubtful whether or not the instructor's accent might potentially have an effect on the L2 learners' acquisition. I will investigate the impact of this predictor 'Arabic formal learning' on the use of HA features and whether it occurs in the migrants' home country or in Hijaz.

To the best of my knowledge, there is no specific institution that teaches Arabic to the migrants who do not speak Arabic as their L1 in Madinah except the institution of Islamic University. However, there are some Quranic circles for both native and non-native speakers of Arabic, but migrant workers usually do not attend these circles because they do not have time. Moreover, even if they have time, women migrant workers specifically are not allowed to leave their accommodation except to go shopping once or twice a month, as the female participants reported. Therefore, it is rarely expected that migrant workers will have received Arabic by means of formal instruction since coming to Saudi Arabia. The effect of the formal instruction in my research might be more related to what migrant workers learned about Arabic or the Quran in their countries of origin than in Saudi Arabia. Hence, my study will determine whether formal instruction plays a part in the use of the HA features. Although it appears unlikely that migrant workers were provided with formal instruction in Hijaz, it remains important to investigate the influence of this variable on their use of HA features. While the expectation is low, exploring the influence of formal instruction can provide valuable insights into language variation among migrant workers.

2.5.8 Migration Identity

According to Le Page and Tabouret-Keller (1985: 181), how individuals use language is determined not only by how they perceive themselves but also how they want others to see them. It is possible to think of inter-individual linguistic differences as 'acts of identity'. According to Piller (2002), identity is a social variable that is frequently linked to nationality, race, gender, ethnicity and migration (see also for instance, Drummond 2012; Drummond & Schleef 2016).

In relation to language, the term identity is defined by Kroskrity (1999: 111) as 'the linguistic construction of membership in one or more social groups or categories'. The importance of this variable in acquiring new variants is clear for native speakers of a given language. For example, Ito and Preston (1998) examined the spread of the Northern Cities Vowel Shift to a non-urban part of northern Michigan. They discovered that the level of language acquisition is related to the individuals' sense of loyalty towards their local rural area and their feeling of social identity. We can extend this idea to L2 acquisition: migrant workers' acquisition of HA features can be viewed as reflecting a growing sense of identification with Hijazi culture. However, lack of HA acquisition might indicate a determination to sustain their L1 identity and resistance to the Hijazi culture.

Some individuals in L2 tend to change their identity and choose the native-like form to converge with the locals while others avoid the acquisition of the native-like form deliberately to diverge from the locals and enhance their L1 identity. Marx (2002: 273) described the fourth stage out of six in her personal experience, specifically detailing the phase related to her time living in Germany; the country where she acquired L2 and subsequently, upon returning to Canada; the country of her L1. She described this phase as German identity construction and English attrition. During that time, she appropriated the L2 accent and considered it a big achievement when she could 'trick' people and make them believe that she was actually an L1 speaker of German. Gatbonton, Trofimovich and Magid (2005) examined the correlation between the affiliation of ethnic groups and pronunciation accuracy in L2, based on data from two separate studies. The results in general showed that 'the more learners sound like the speakers of their target language, the less they are perceived by their peers to be loyal to their own group' (Gatbonton, Trofimovich & Magid 2005: 504).

Drummond (2012) believed that one of the ways that identity can be interpreted and measured is by means of asking participants about their future plans. He investigated /ing/ variation among Polish migrants in Manchester and discovered that their future plans have a significance effect on the speech of these non-native speakers. Participants who intended to settle in the UK or did not have any plans to return to Poland produced the four local variants of /ing/ which are [ɪŋ], [in], [ɪŋk] and [ɪŋg] (however, with slightly different degrees), whereas others who planned to return to the country of their heritage language, and thus arguably feel allegiance towards Poland and their culture and a stronger sense of identity,were less likely to use the [in] variant than participants who intended to settle in the UK or did not have any plans to return to Poland. Moreover, in the analysis that comprises only [ɪŋk] and [ɪŋg], participants who planned to return to the country of their heritage language were more likely to use [ɪŋk]. Similarly, Elliott (2018: 248) discovered that the vowel productions of Slovak immigrants who preferred to stay in Scotland and those who are uncertain about whether to remain in Scotland or return to Slovakia were highly monophthongal for FACE and GOAT vowels and therefore, similar to the native speakers in Edinburgh. However, immigrants who intended to return to Slovakia had more diphthongal pronunciations.

The tools that Drummond and Elliott used to measure allegiances could be used as a proxy for migration identity of the participants and show how this may influence their production of HA features.

2.5.9 Age

One independent variable that can affect language acquisition is age. According to the 'Critical Period Hypothesis' (CPH) proposed by Penfield and Roberts in 1959 and then refined by Lenneberg in 1967, its effect is such: mastering a language after the age of 12 is impossible because the biological/neurological period of development ends around this age due to cerebral plasticity changes. Initially, this hypothesis was related to L1 acquisition; it proposed that the critical period expands from the age of two approximately until the end of puberty, which is the age of 14 approximately, beyond which acquiring linguistic proficiency is impossible.

The notion of 'sensitive period' has been proposed in relation to the acquisition of L2 to indicate that the critical period is neither unexpected nor absolute criterion and after that period acquiring

L2 is impossible (Long 1990: 10). However, it is a gradual process where the degree of L2 attainment is ultimately variable. Specifically, acquiring the L2 is possible although not to the extent of achieving native-like competence. In the literature pertaining to L2, the concepts of 'critical period' and 'sensitive period' are frequently used interchangeably (Piske, MacKay & Flege 2001: 196).

Various scientists have argued that those who start learning L2 in adulthood cannot achieve native-like attainment, as the hypothesis of maturational constraints denotes (Long 1990). For instance, Scovel (1988) alleges that achieving pronunciation similar to that of native speakers in a foreign language is impossible after a certain critical period. Conversely, Flege (1981) mentioned that the age of the adult speakers might not prevent them from acquiring the L2 successfully. Furthermore, acquiring a new language with the fluency of a native speaker is not lost over time; rather, it continues to exist. Bongaerts, Planken, and Schils's (1995) results do not support the claim made by Scovel because they determined, through accent-rating studies, that the pronunciation of several native speakers of Dutch, who were late L2 learners, was indistinguishable from that of the native English control group. Likewise, Birdsong (1992), using the same type of study, examined the syntax of 20 English native speakers who learned French after puberty. He compared them with 20 French native speakers and found that several of the learners were linguistically indistinguishable from the control group.

Age of arrival in a place where the L2 is spoken also affects L2 acquisition. Patkowski (1980) examined the syntactic proficiency of 67 adult immigrants in the United States. Two judges were trained to assign syntactic ratings to written transcripts of the participants' recorded oral interviews. Patkowski's study tested the effect of age of arrival. The predictor levels were <15 and >15 years old at the time of arrival, and he included in his analysis other variables, such as the exposure to English in formal and informal settings. The results indicated that whereas age at arrival was the only predictor that had a significant effect on syntactic proficiency, the other independent variables had little influence. Learners who were prepubescent performed better than learners who were postpubescent. Johnson and Newport (1989) conducted a study of 46 adult immigrants who spoke Chinese or Korean as their L1 and learned English as their L2. Their age of arrival in the United States ranged from 3 to 39. The participants' knowledge of 12 grammatical

rule categories that are related to syntax and morphology was examined using a grammaticality judgment task. The task included 276 sentences. Of these sentences, 140 were ungrammatical and 136 grammatical. The findings revealed that there is a strong relationship between the age at arrival of the migrants and their command of English syntax and morphology. A positive correlation exists between the age at which people commence English language acquisition and their achievement of higher test scores. This relationship persisted even when considering factors such as the amount and timing of their formal instruction in English. Consequently, individuals who arrived earlier demonstrated an advantage for those over those who arrived late.

Age may not always be an obstacle in regard to acquiring an L2. This pertains to not only tutored learners but also untutored students. Loup *et al.* (1994) examined an untutored learner, Julie, a native speaker of English. When she was 21 years old, she married an Egyptian national and emigrated to Cairo. Nine days after her arrival to Cairo, her spouse was called to military service so she had to communicate in Arabic. Julie had never received formal Arabic instruction; therefore, she could not read or write in Arabic. After almost two and a half years, Julie was fluent in Egyptian Arabic. She had lived in Egypt for 26 years when the study was conducted.

Julie was tested in three domains, specifically speech production, recognising accents and syntactic rules, and her performance was compared with native speakers and a non-native Arabic speaker called Laura. While Julie learned Arabic in a natural setting, Laura learned it in a formal setting and was regarded as very proficient. Egyptian language experts assessed Julie and Laura. These 13 experts were teachers of Arabic as a foreign language who listened to the spontaneous speech of Julie and Laura when they shared their favourite recipe. Each expert had to rate Julie and Laura as either native or near-native Arabic speaker. Regarding accent identification, Julie and Laura were asked to discriminate various regional Arabic dialects, such as Palestinian, Sudanese, Libyan, Kuwaiti, Syrian and two Egyptian dialects. While two Arabic native speakers achieved 85% accuracy, Julie and Laura achieved 100%, revealing that their perceptive abilities were excellent. Moreover, they were asked to differentiate between the Cairene variety and other Egyptian dialects. This task was more difficult than the previous one because the phonological differnces between the Egyptian dialects are very subtle and not all the native speakers can differentiate them

correctly. The performance of Julie and Laura on this task was not identical. Julie's responses were closer to the responses of the native Egyptian Arabic speakers. Loup *et al.* (1994) ascertained that it is evident from the two measures that both women achieved similar performance levels and the difference between them was insignificant.

Age may not negatively impact an individual's L2 acquisition if there are strong motivations for them to acquire the L2. One reason for Julie's success might be that she accepted feedback regarding her errors, manipulated the language grammatical structure consciously, focused on the morphological variations and reviewed the entries in her copybook regularly (Loup *et al.* 1994: 92). However, as noted by Loup *et al.* (1994: 92), learners of L2 primarily do not achieve the proficiency of native-like speakers even with conscious awareness. Hence, there are additional factors. Julie reported a number of inherited traits, such as twinning, allergies and left-handedness, that have been linked to language learning aptitude (Obler 1989). Likewise, Loup *et al.* (1994) think that her aptitude concerning the acquisition of L2 might be related to the rapid progress in L1 acquisition. This supports Skehan (1986), as Julie reported that when she was two years old, she could imitate people on the radio.

To sum up, the previous examples suggest that some people – both tutored and untutored - acquire (near)-native competence in an L2 even if they begin to learn it later in their lives. Although most learners do not attain native-like speech after the age of puberty, there are exceptions. This dissertation includes age as a fixed factor.

2.5.10 Educational Level in L1

Migrants' level of education in their L1 might play a significant role in their acquisition of L2. Alderson (2005: 133 & 164) notes that there is a significant difference between L2 learners of English; those with a university degree perform better in language proficiency than those with secondary or primary degrees. Moreover, Faltis and Coulter (2008) compare the ability of acquiring English between adult immigrant students who have previous L1 literacy and academic experience**,** and those who have never been in school or have little formal education. Their study established that the former group acquired the target language more quickly than the latter group. In Al Kendi's (2021) study, L1 educational level was examined as a continuous variable. 22 female

migrant domestic helpers provided the entire number of years that they had spent at school. The study shows that their lenght of schooling played varying roles as regards the participants' phonological performance. Although it did not have an influence on the production accuracy of their consonants, it had a significant effect on their phonological sensitivity and the production of coda consonant clusters.

Bigelow and Tarone (2004) claim that some research found evidence of a relationship between indivduals' literacy in L1 and their literacy in L2. For instance, the studies performed by Bialystok (2002) and Cummins (1991) suggest that the literacy in L1 assists learners to be literate in their L2. Collier (1989: 527) noted through his review of the literature on L2 that those who have never been in school during their L1 acquisition may spend 7 to 10 years attempting to attain the average performance level of the native speakers of L2 and learn L2 literacy (Bigelow and Tarone 2004: 690). Furthermore, catching up with native-speaking peers appears to be unachievable by some of those who are illiterate in L1. In contrast, it took less time for those who are literate in their L1 to acquire similar L2 literacy skills. In relation to linguistic competence, L1 literacy impacts L2 linguistic processing. Bigelow and Tarone (2004: 693) suggest that when the ability of more literate learners is compared to the ability of less literate learners in distinguishing between linguistic structures in oral L2 input and one's own output, it appears that the ability of the latter group is reduced. This refers to the 'Noticing hypothesis' developed by Schmidt (1990) which states that L2 learners must consciously notice a linguistic form to acquire it.

2.5.11 Proficiency in Arabic

Proficiency in L2 might impact acquiring variation in the use of the L2 linguistic features. Ulbrich (2021) investigated whether or not phonetic accommodation varies with the proficiency level of female native Spanish speakers who are learning German as their L2. The study's primary focus is on the recognition of a process that occurs in German but not in Spanish - the final voicing contrast neutralisation in plosives. Significant differences were ascertained between speakers with high and low proficiency in terms of closure time, post-release voicing and burst release duration. Compared to low proficiency speakers, accommodation effects were more clearly visible in the data of speakers with a high proficiency level. Drummond (2010) discovered that level of English is a statistically significant condition in the analysis of h-dropping, (ing)-variation and glottal variation

in /t/. The relevance of this independent variable demonstrates that the likelihood of using local variants increases along with proficiency. Proficient English speakers use /h/ droppings, alveolar variants of (ing) and glottal variants /t/. Similarly, Young (1991) identified that the production of the plural -s inflectional morpheme is significantly constrained by the proficiency of Chinese native speakers who are learning English as their L2, with greater proficiency favouring the production of -s over low proficiency. In my study, the level of Arabic proficiency will be taken into account, with a focus on speaking ability. At the time of the interview, participants will be asked to rate their own level of Arabic. Similar to the effect of this predictor in the above-mentioned studies, proficiency may possibly have a significant effect in my study too.

2.5.12 Language anxiety

Language anxiety is one of the predictors that can have a significant effect on L2 acquisition. MacIntyre and Gardner (1994: 284) defined it as 'the feeling of tension and apprehension specifically associated with second language contexts, including speaking, listening and learning'. MacIntyre (1998: 27) imagined it as "the worry and negative emotional reaction aroused when learning or using a second language'. According to MacIntyre and Gregersen (2012), language anxiety is aroused when learners have to perform in the L2 and it includes various negative emotions, such as fear and worry. The possible effect of language anxiety on L2 performance is extensively recognised. Nonetheless, the connection between language anxiety and the selection of variants of a language might not be obvious at the outset. My study will examine how this variable influences the use of HA features.

One of the fundamental theories in L2 acquisition is the 'affective filter' hypothesis. Dulay and Burt devised this particular theory in 1977. Krashen (1982) subsequently developed it and indicated that affective factors, such as language anxiety, self-confidence and motivation, can significantly impact language acquisitionin that they can disrupt acquisition and block input. Therefore, when anxiety levels reach a specific threshold, acquiring language at that point might become impossible. Absorbing comprehensible input can be easier in an environment with low anxiety. Various researchers in L2 agree that language anxiety might be the most significant affective factor that influences the learning process (Arnold & Brown 1999; Horwitz, Horwitz & Cope 1986; MacIntyre 1998).

Language anxiety was determined to be debilitating in different L2 contexts and in various target languages. Bosmans and Hurda (2016) hypothesised that there is a link between language anxiety and French phonological attainment in distance language learning settings. Low levels of language anxiety assist the participants to have high scores in phonology. MacIntyre and Gardner (1989) discovered that there is a significant negative correlation between language anxiety and performance on not only learning but also the production of vocabulary in French. Phillips (1992) found that there is a significant negative correlation between language anxiety and the skill of speaking French. Abu-Ghararah (1999) discovered that there is a harmonious negative correlation between language anxiety and the participants' achievement in English. Aida (1994) found that language anxiety has a significant impact on performance in Japanese.

In conclusion, my study will examine the impact of the above-mentioned predictors to discover the Arabic varieties spoken by Bengali, Hindi and Filipino migrant workers in Madinah. If the independent variables have significant effects on the migrants' use of HA features, this may suggest that their Arabic varieties are interlanguage. In contrast, if the factors have little effect on their use of HA features, I expect that their varieties are pidgin. We cannot decide that the migrants' variety is a pidgin by examining the result of one social factor independently without analysing the effect of other factors. This is for the reason that social factors may not operate independently but are related to each other due to the social reality which is not only nuanced but also dynamic and complex. Hence, predicting social variation via isolated social variables can be misleading (Albirini 2016: 188).

2.6 Interlanguage vs. Pidgin

There are various similarities between pidgin and interlanguage varieties, such as both might be learned as a L2. However, certain scholars differentiate between them, for example, Flick and Gilbert (1976). Their distinctions between pidgin and interlanguage are related to the social context in which each variety emerges.

Flick and Gilbert (1976) argue that pidgin is a group phenomenon. Pidgin speakers typically form a close-knit community and communicate using the pidgin. Schumann (1978) believed the debate between a group phenomenon and an individual phenomenon is only acceptable if tertiary

hybridisation[2] is the sole way for pidginisation to occur. This argument, however, loses significance if we accept that secondary hybridisation can also contribute to legitimate pidginisation although it does not result in a true pidgin. A pidgin language is less affected by the target language's corrective influence and develops more independently (Flick & Gilbert 1976). Flick and Gilbert (1976) along with Meisel (1977) contend that the social distance and lack of contact between the superstrate and the substrate language groups are great. Meisel (1977) also argues that pidgin speakers adhere to a language "norm" that is relatively stable. Put differently, there are concepts pertaining to grammaticality that are intuitively shared by the community of pidgin speakers.

On the other hand, Flick and Gilbert (1976) argue that interlanguage is an individual phenomenon. Consequently, during intragroup communication, L2 learners typically avoid utilizing the target language. In addition, the interlanguage of the L2 learner is constantly corrected and developed until it approaches the target language closely. Schumann (1978) believed it is simply not true that all L2 learners' interlanguages eventually approach the target language very closely, as shown by Alberto in Schumann's (1976) study. Alberto is a 33-year-old Costa Rican acquiring English in an untutored situation. His language acquisition was studied longitudinally for 10 months. The results prove that there has not been much progress in the use of auxiliaries, interrogatives and negatives. According to Flick and Gilbert (1976) and Meisel (1977), in an interlanguage situation, social distance and a lack of contact between the superstrate and the substrate language groups can be meaningful, but it may not be as applicable as in a pidgin situation. Interlanguage learners do not share the same stable norm (Meisel 1977). This sort of speech undoubtedly differs between people and is frequently significantly variable within people.

In my study, to establish whether migrant workers' Arabic L2 in Madinah an interlanguage or a pidgin, I will analyse the impact of contextual predictors on the migrants' use of L2 HA features.

[2] According to Whinnom (1971), tertiary hybridisation is based on the assumption that unless the lexifier impact has disappeared from the contact situation, a stable pidgin cannot appear. In order for stable pidgins to emerge, Whinnom also claimed that there must be at least three languages in contact. Meisel (1975) suggested that only this type results in a true pidgin. In secondary hybridisation, Whinnom (1971) suggests that the access of the L2 learners to the lexifier language is not completely cut off.

2.7 Fossilisation and Pidgin Language

Typically, the concept of interlanguage presupposes that there is some continuation of the acquisition process. However, language learning is frequently characterised by learners' failure to achieve full target language proficiency even though they can manage to sustain effective and successful communication in it (Matras 2009). Regardless of the learner's amount of instructions or age, fossilisation was introduced to capture a permanent adoption phenomenon as regards interlanguage features which are idiosyncratic in the L2 of the learner. Selinker (1972) defined it as the production of a cognitive mechanism which is a non-target-like-end-state.

Interlanguage theory suggests that pidgins appeared as imperfect versions influenced by a model language. In this specific case, Arabic serves as the model language in my study, representing the lexifier (Velupillai 2015: 139). According to this theory, the aim of non-native speakers of a target language is to acquire it. However, due to the contact situation conditions, they are unable to master it fully. The result of their unsuccessful attempts to match the target is a pidgin, which in this sense is a version of the lexifier with a very restricted lexicon and simplified grammar (Sebba 1997: 79). Thus, the pidgin would represent a stage of fossilisation in the process of the interlanguage development in L2 acquisition (Velupillai 2015: 140).

Other approaches to pidgins consider them to be:

> examples of partially targeted or non-targeted second language learning, developed from simpler to more complex systems as communicative requirements become more demanding. Pidgin languages by definition have no native speakers, they are social rather than individual solutions. Hence, they are characterised by norms of acceptability.
> (Mühlhäusler 1986: 5)

Pidgins also arise 'to fulfil certain restricted communication needs among people who have no common language' (Todd 1990: 2). Holm (1988: 5) also defines pidgins as 'a reduced language that results from extended contact between groups of people with no language in common'. My study will investigate the Arabic varieties of migrant workers, which might be a pidgin language. Their L2 varieties could be fossilized, leading to the production of features distinct from those of the L1 speakers. These features may be simplified and resemble characteristics typically associated with pidgin languages. Even if these workers have limited contact with local L1 speakers of Arabic

and use Arabic in restricted contexts, they may learn it informally from L1 Arabic-speaking friends. My study will help to clarify the status of the Arabic varieties of the migrants.

2.8 The Stages of Pidgin Development

When pidgin languages develop, they go through several phases which are characterised by varying levels of complexity and stability to serve their purpose of facilitating communication. The current stage of pidgin is defined by both linguistic and social features (Siegel 2008). However, Velupillai (2015) believes that there are no precise boundaries among these stages. The following are the three main stages that pidgin languages move through. However, it is important to mention that extended pidgin did not go through all the previous stages of development. As my study predicts that the Arabic varieties of the migrants could be described as a pidgin, the following reveals the language development of a pidgin.

2.8.1 Jargon

Linguists use various terms to refer to Jargon, such as pre-pidgin, unstable pidgin and incipient pidgin (Bakker 2008: 132). It is essentially the earliest phase of pidgin development and is also the least structured stage (Sebba 1997: 102). Moreover, this type of pidgin can disappear if communication among two groups stops (Singh 2000). Siegel (2008: 2) describes it as attempts by an individual to interact with people in a contact situation by using lexical items and phrases that they have learned from other languages which are predominantly from the lexifier language.

This phase of pidgin development is characterised by different linguistic features compared to the following two stages of pidgin. Immense instability in vocabulary and grammar and the lack of stable norms is the most crucial characteristic concerning jargon (Sebba 1997). The use of non-verbal communication and gesture is frequently applied in this stage (Ansaldo & Matthews 2004). Velupillai (2015) notes that it is a highly variable and unsystematic form of L2 that is used in interethnic communication (Bakker 2008: 131). Moreover, this variety is significantly influenced by the grammar of the L1s (Field 2004; Sebba 1997). According to Sebba (1997), jargon is employed in social situations that are the most restricted. However, jargon might also evolve into a stable pidgin if people continue using it in their contact (Mesthrie 2008).

2.8.2 Stable Pidgin

The next stage is known as 'stable pidgin' in Mühlhäusler (1997) words, whilst Bakker (2008) terms it 'pidgin'. When jargon persists in communication, it is the means for not only communication among the dominant and subordinate groups, but also among different subordinate groups in the contact situation. This communication web results in the emergence of a stabilised pidgin (Sebba 1997). In this stage, the speakers' target is not the lexifier language but the pidgin. Sebba (1997: 105) is of the opinion that this phase of pidgin development, characterised by its own rules for syntax, pronunciation, as well as vocabulary, may arise from an unstable jargon. Nonetheless, Mühlhäusler (1997) suggests that the least stable elements in this context are pronunciation and phonology. The influence of the substrate languages in this phase diminishes and the variability that exists in a jargon is reduced (Mühlhäusler 1986: 176). Furthermore, although derivational morphology exists in this stage, it is rather limited (Sebba 1997). Pidgin has 'no irregularity, high transparency, no morphological categories, and no syntagmatic redundancy' (Trudgill 2009: 174).

2.8.3 Extended Pidgin

Expanded pidgin, also called pidgin creole, can develop when people who use stable pidgin expand its functions into new domains of life (Sebba 1997). While the earlier stage could have been utilized for spoken conversation, extended pidgin might be used in new domains, extended pidgin might be used in new domains, that include spoken or written communication, such as scriptural translations. As such it may become the main language of the speech community. Additionally, it might become the L1 for particular community speakers, such as Tok Pisin for community members in Papua New Guinea (Velupillai 2015: 21).

The morphological level is possibly where a pidgin's expansion is most noticeable (Sebba 1997). This also has an impact on the lexicon and grammar. Lexical expansion occurs especially through derivational morphology, which is underdeveloped at the stable pidgin stage. Similarly, inflectional morphology tends to be more complex, even though it never gets as complex as in the superstrate or substrate languages (Sebba 1997).

Stages of Migrant Workers' Varieties of Arabic

My study focuses on the varieties employed by the migrant workers in Hijaz; Madinah. Various studies conducted in different regions across Saudi Arabia and in different Arabian Gulf countries have also investigated the varieties of migrants. None of these studies are relevant to the specific group of migrants that I am presently studying in my research. The literature presents opposing opinions concerning the level of development of GPA varieties, which are spoken by migrant workers in the Gulf region. Some researchers such as Alghamdi (2014), Alshammari (2018) and Næss (2008) suggest that they are incipient pidgins, whereas Almoaily (2008), Bakir (2010) and Hobrom (1996) describe them as pidgins. In responding to Almoaily (2012: 45) who maintains that GPA 'carries some typical creole features, such as the use of adverbs to mark for TMA […] and the use of reduplication as a word-formation process', these features were retested by Avram (2015) who concluded that this variety has only TMA markers, which is more characteristic of pidgin languages. Additionally, using a set of diagnostic features which are primarily taken from Mühlhausler (1997), Avram concludes that GPA is in the jargon stage. Avram used academic literature about GPA, general descriptions of GPA and 25 websites to determine the developmental stage of GPA by focusing on phonology, morphology and vocabulary. He established that GPA possesses the majority of the characteristics associated with jargon. For instance, in phonology, there is inter-speaker variation. There is also a small vocabulary size and a lack of prepositions, complementisers, conjunctions and tense and aspect markers, and the personal pronoun system is minimal. Besides these characteristics, Avram mentioned that there are several factors that hinder stabilisation. To illustrate this, in GPA, the L1 of the migrants contributes to inter-speaker variation. A further issue that makes stability challenging is related to the type of Arabic input. Bakir (2010) argues that the level of exposure to Gulf Arabic is dramatically different between domestic workers and oil or building workers. The former who regularly work in small groups or alone are more likely to be exposed to Gulf Arabic than the latter group who work in big groups and generally speak the same L1. The exact varieties of Gulf Arabic that migrant workers are exposed to vary as well. Finally, given that GPA is not a "target language," it is not passed along from speaker to speaker as such but is instead developed independently. Furthermore, various "fossilised" characteristics are used by various GPA speakers, which causes feature "fossilisation" to begin (Bakir 2010). In conclusion, based on the analysis conducted by Avram (2015), GPA is identified as being a part of the jargon stage. My study will investigate whether the Arabic varieties

of migrant workers in Madinah are pidgin and could be in a jargon stage similar to the Arabic varieties of the other migrant workers in different regions across Saudi Arabia and in different Arabian Gulf countries or if they are an interlanguage.

2.9 General Characteristics of Pidgin

One of the common features of pidgin languages is simplification, which concerns all aspects of grammar. Compared to source language grammars, pidgin grammars are typically less complex (Albaqawi 2020). This section presents the various general linguistic characteristics associated with pidgin languages.

2.9.1 Phonology

In pidgin languages, the syllable structure that is preferred is CV (Parkvall 2020: 266; Sebba 1997: 48). Vowels might be inserted to break up consonant clusters. For instance, *sitiret* 'straight' in Tok Pisin and *sikis* 'six' in Fanakalo. Pidgins tend not to be tonal languages. Cluster reduction may also occur. Consequently, the tonal distinctions of lexifiers are lost or simplified. For example, whereas Zulu, the superstrate language of Fanakalo, is a tone language, Fanakalo is not. Holm (2000: 138) indicated that in the pidginisation process, particularly at the phonology level, universals can play an essential role. On the one hand, there is a preference for common sounds that occur in the phonological system of most of the world's languages like /d/ and /m/. Conversely, marked sounds, which are rare and more difficult to pronounce, are avoided, such as /ð/ in Icelandic and English and /θ/ in Greek and English. This causes a reduction and a simplification in which speakers of pidgins typically adjust the consonant system to their L1 system (Holm 2000: 139; Sebba 1997: 47). For example, speakers of Tok Pisin substitute the labiodental /f/ such as in 'finish' to *pinis* with a /p/ (Holm 1989). Likewise, in Ghanian Pidgin English, Akan speakers might replace /v/ with /b/ or /f/ because it is absent in the Akan phonemic inventory (Huber 1999: 168). Moreover, the vowel systems in pidgins tend to be simplified (Sebba 1997: 46).

2.9.2 The Lexicon

Most pidgins have a smaller number of lexical items compared to their lexifier languages (Holm 2000: 108; Parkvall 2017: 17; Sebba 1997: 51). In general, pidgin languages have few synonyms. According to Hall (1953), a speaker of a typical language uses between 25,000 and 30,000 lexical

items. In contrast, there are only 1,500 lexical elements in Neomelanesian, a pidgin spoken in Papua New Guinea. The notion of 'maximum use of a minimum lexicon' was coined by Mühlhäusler (1986: 171). It refers to the use of a single lexical with multiple functions and meanings. Hence, to compensate for the restricted size of the lexicon, multifunctionality, by which the same word has many syntactic uses and functions as different parts of speech, is common in pidgins (Holm 2000: 108). For example, the lexeme *sik* 'sick' in Tok Pisin can function as not only an adjective but also a noun (Romaine 1988: 38). According to Silverstein (1972: 381), the lack of lexical specificity raises the information content of every word in the lexicon. The vast majority of Chinook jargon's lexicon is composed of grammatically and semantically ambiguous words. Similarly, polysemy, which means the same word has many meanings, is frequent in pidgins (Holm 2000: 108). For example, the Chinook Jargon word *muckamuck* has various English equivalents, including 'bite,' 'eat' and 'drink' (Bakker 1994: 36). In addition, grammatical function words, such as prepositions and pronouns, are unusual in pidgins (Sebba 1997: 45).

2.9.3 Morphology and Syntax

In pidgins, synthetic morphology is very rare (Sebba 1997: 46). Instead, there is a tendency toward analytical morphology (Parkvall 2017: 3; Schreier 2008: 37; Sebba 1997: 46). To illustrate this, while the noun system of Zulu, the lexifier language of Fanakalo, has a set of noun classes that includes ten different prefixes for singular and plural, there are only two such prefixes in Fanakalo. In addition, many pidgins do not have inflectional morphology at all which can result in the loss of grammatical categories such as tense or gender (Sebba 1997: 44). For instance, there is a lack of gender morpheme distinction in Tok Pisin (Schreier 2008: 37). Specifically, English object pronouns 'him, her and it' are neutralised and represented as *em.* Moreover, pidgins lack inflections that indicate plurality and agreement. The loss of agreement and plurality in Tok Pisin is shown in examples (5) and (6). The noun form remains unchanged and numeral items are used to indicate plurality.

(5) Sikspela man i kom
Six people are come.PROG
'Six people are coming'
(Schreier 2008: 37)

(6) Wanpela man i kom
A man is come.PROG
'A man is coming'
(Schreier 2008: 37).

In addition, most pidgins employ subject-verb-object (SVO) word order (Parkvall 2020: 266; Sebba 1997: 40). The negation marker is typically a separate word, its form is invariant, and, in most pidgins, it occurs before the verb (Schreier 2008: 42). For instance, in Melanesian Pidgin, it always precedes the verb, as illustrated in example (7) below (Sebba 1997: 39).

(7) Yu no save wokabaut
You NEG know (how to) move
'You can't move'.
(Sebba 1997: 39).

Also, tense, aspect and modality (TAM) are evident from context (Sebba 1997: 42). The vital clues for TAM are occasionally provided by adverbials, such as perhaps, before, now, soon and tomorrow, conceivably (Parkvall 2017: 8; Sebba 1997: 42). TAM can also be marked externally to the verb. To explain this, while *mhlaumbe* 'perhaps' is probably an adverb in Fanakalo, it indicates modality in example (8).

(8) Mhlaumbe mina zo hamba
Perhaps I FUT go
'I may go'
(Sebba 1997: 112).

2.10 Previous Studies of Migrant Workers' Varieties in Arabian Gulf Countries

The varieties of Arabic used by migrant workers who are from different heritage language backgrounds, such as Javanese, Malayalam, Tagalog, Sinhala, Tamil, Hindi, Urdu and Bengali have been studied in six countries in the Gulf region including the United Arab Emirates, Saudi Arabia, Qatar, Bahrain, Oman and Kuwait (Albaqawi 2016: 263). Albaqawi (2016: 263) noticed

that different labels are suggested for these varieties. Næss (2008) who conducted his study in Buraimi, Oman and Al-Ein, UAE and Bakir (2010), who undertook his study in Qatar, referred to the migrants' variety as 'Gulf Pidgin Arabic'. In regard to Saudi Arabia, Abed (2017), Albaqawi and Oakes (2019), Alghamdi (2014) and Alshammari (2018) prefer to call it 'Gulf Pidgin Arabic' too. Likewise, the studies conducted by Almoaily (2008) and (2012) in Saudi Arabia, use 'Urdu Pidgin Arabic' and 'Gulf Pidgin Arabic' respectively. Albakrawi (2012) and Alshammari (2010) who collected their data in Saudi Arabia term it 'Saudi Pidgin Arabic'. In relation to the literature, generally, migrant workers' varieties in the Gulf region are frequently lumped together under the name 'Gulf Pidgin Arabic' (Albaqawi 2016: 263). Although Saudi Arabia is a Gulf country, only the eastern part of the country is close to other Gulf countries. Therefore, the Saudi Arabic varieties there are similar to the Arabic Gulf dialect in comparison to the Hijazi dialect. In the literature, although the study undertaken by Alghamdi (2014) focused on the migrants' variety and was conducted in the West of Saudi Arabia, which is located on the opposite side of the Arabian Gulf, the migrants' variety continues to be known as 'Gulf pidgin Arabic'. However, Al-Shurafa (2014) tends to be more specific and labels it 'Hejaz Gulf Pidgin Arabic'. Moreover, Al-Azraqi (2020) favours calling the migrants' variety 'Arabian Peninsula Pidgin Arabic', because she conducted her study in Abha in the Southwest of Saudi Arabia. She suggests this term for all other migrant workers' of Arabic spoken in the Arabian Peninsula. Moreover, it is evident that there is inconsistency in the literature regarding what the migrant workers' variety should be known as. Therefore, certain researchers have a tendency to refer to it as 'Gulf Pidgin Arabic'. Notably, scholars studying migrant L2 Arabic in Saudi Arabia have labelled it as a pidgin. The crucial aspect is not the name that it is given but rather the reasons behind classifying these varieties as pidgins.

There are few published studies related to Arabic-based pidgin. Versteegh (1984: 115) observes that '[It is] surprising that there is no mention at all of pidgins based on Arabic in any other part of the world, except in Africa'. For instance, there is not only Juba Arabic and the creole of KiNubi Arabic in the south of Sudan, Kenya and Uganda, but also other Arabic varieties spoken in Nigeria and Chad (Owens 1997).

The following paragraphs explore several GPA studies. As simplification, in general, appears in the speech of migrant workers in these studies, it might also manifest in the speech of migrant

workers in my study in the form of primarily deleting the morphosyntactic features and substituting the phonological variable /f/ with other variants. If my study determines this and the contextual variables demonstrate little or no impact on the production of HA features, the migrant workers' varieties can be classified as a pidgin. A number of studies conducted on GPA are problematic in terms of their methodology, such as Smart [3](1990) and Wiswall[4] (2002). For this reason, their results are unreliable, and they will not be considered further in this dissertation; the focus instead is on those that produced reliable results.

Hobrom (1996) conducted a study in the northwest of Saudi Arabia. He described the Arabic variety of seven Indian workers who speak Tamil, Malayalam and Kannada as their L1. Audio recording was used to record free conversations between the participants and two interviewers who speak Saudi Arabic. However, how many hours or minutes of recording is not obvious in his thesis. Hobrom examined specific phonological, syntactic and morphosyntactic features in the migrants' Arabic variety. He discussed three phonological processes: deletion, insertion and substitution; analyzed syntactic features, namely reposition, agreement problems and copula insertion; and explained various morphosyntactic features, including bound forms, free forms, types of verbs and inflectional affixes. Hobrom's results, in general, reveal that the participants simplified the complex Arabic elements. Hobrom classified the participants' number of phonological deviations in Arabic in relation to their L1 and length of stay. He ascertained that Kannada and Malayalam speakers had the fewest phonological deviations, respectively. Tamil speakers, on the other hand, made the greatest number of deviations. Moreover, he discovered that fewer deviations were made by the participants who spent more time in an L2 environment. Although Hobrom did not include quantitative analysis in his study, my research will do that.

Naess (2008) examined the status of the Arabic variety of Asian migrant workers in Buraimi, Oman and Al-Ein, UAE and whether it is a separate language variety or an individual attempt to

[3] The source material in Smart's (1990) study is primarily based on two Emirati newspapers; articles and humorous cartoon captions that were written by native Arab journalists who imitated the workers' language to make it entertaining.

[4] Wiswall (2002) tested some linguistic features in the Arabic variety that is used among the locals in the UAE and Indian workers to determine if the two groups speak it differently. Arabic native speakers were asked to translate eight sentences from English to colloquial Arabic, then imitate how migrant workers produce these sentences in their speech. Indian workers were also asked to translate similar sentences from their L1 into GPA during interviews conducted over the phone.

master Gulf Arabic. She described the phonology, negation, verb phrase, possession and the copula *fi* of GPA. Naess interviewed sixteen participants who she stratified by social class, specifically middle-class and lower-class. She stratified them by their sex. Additionally, the participants speak different L1s, such as Urdu, Sinhala, Malayalam, Javanese, Tamil, Tagalog and Javacano and worked in different jobs. Like Alshammari (2018: 52) and Almoaily (2012: 59), I am critical of the research Naess undertook. Although there are many advantages as regards her data, in particular because it is varied, she does not provide any quantitative analysis. There is a possibility that using quantitative analysis would have made Naess' claims much more reliable and assisted her to generalise her results to situations other than the one she studied. Naess ascertained that GPA speakers' language is mostly characterised by simplicity and reduction. She aimed to establish that the status of the migrants' variety that she investigated is not a collection of various people's efforts to acquire Gulf Arabic but a separate language variety.

Al-Shurafa (2014) conducted a study in Jeddah, Hijaz, Saudi Arabia of seven male migrant workers from Asia and Africa. None share the same L1 and, at the time of the study, they had been working in Hijaz for ten or more years. Given that the participants were not only friends but also live in the same neighbourhood, this facilitated their participation together in three tea gathering occasions and recording them for five hours. This data collection method might be considered an advantage as regards her study because she could be certain that the participants speak Arabic, without the researcher's Arabic variety impacting them during conversation. Al-Shurafa examined how migrants use the syntactic construction /fi:/ 'there is' and its negative counterpart /ma: fi:/ 'there isn't' when they speak Arabic. In general, she established that the structure of the migrants' Arabic variety is characterised by 'impoverished grammar' and its rules are simplified compared to the superstrate language, HA. Although when the participants use /fi:/ there is generally some degree of consistency, this is not always the case but depends on the participants' L1 typology. The Arabic variety employed by these migrants is not completely stabilised, however. Al-Shurafa argues that it is a simplified and pre-grammatical system which is like a pidgin language that is in the early stages of development. The degree of pidginisation substantiates the variations in the participants' speech. When the participants practice Arabic for a longer time, the variation becomes less and the consistency improves, which in turn, increases regularity. The participants not only use /fi:/ as a preposition and an existential element which are the basic functions, but also

to achieve interrogative, modality, negation and prediction. Moreover, when /ma: fi:/ is expected, it is not actually present. Two participants who speak Swahili and Tigrinya as their L1 did not use /fi:/ and /ma: fi:/ or barely used it. This instance supports the hypothesis that the migrants' L1 lacks this syntactic structure, which is common in the L1s of the Asian labour force, such as Urdu (Bakir 2010: 218).

Ten migrant workers, seven males and three females, participated in the study conducted by Alghamdi (2014). The males are employed in different low paid professions and the females work as housemaids. The L1s of the migrants are Bengali, Urdu, Indonesian and Sinhala. Their length of residency ranged from 18 months to 21 years. The variability of the participants is one of the advantages of Alghamdi's study, for the reason that this may enhance the study's external validity and generalizability. On the other hand, Albaqawi (2020: 89) believed that, as the females are housemaids who live with local families, these families may use Gulf Arabic (GA) when they communicate with them. Hence, the input that the recently arrived migrant workers are exposed to is Gulf Pidgin Arabic and the amount of Gulf Arabic input that they are exposed to may be limited.

Alghamdi interviewed five of his participants as a group and the other five individually. The data collected in his study included approximately one hour of recorded conversation between Alghamdi and each of the participants. Given that his study is of a small corpus, Alghamdi suggested a future study to run quantitative analysis on a diverse, inclusive and large corpus. Alghamdi hypothesised that if migrant workers use a high level of consistency in GPA linguistic features, this may be a good indicator of regularity in pidgin languages. He used quantitative methods to gauge the degree of consistency in the migrants' use of Arabic negation markers, the pronominal system and morphological inflections, not only concerning nouns but also in relation to verbs and sentence structures. He also used qualitative methods to describe some essential linguistic features within the migrants' Arabic lexicon system, morphology (inflectional affixation) and syntax within the nouns, pronouns and verbs. Generally, Alghamdi's findings reveal that there is an obvious reduction in complex linguistic components and an increased regularity in the use of simpler linguistic items instead of the more complicated ones.

Aljutaily (2018) completed his research in the Qassim region of Saudi Arabia, in a city called Buraidah. He examined the variation of seven marked consonants in Arabic, specifically / x, ɣ, sˁ, tˁ, ħ, ʕ, ɵ/ in the speech of 40 male migrant workers whose L1 is Urdu or Malayalam. To collect the data, short sociolinguistic interviews that ranged between ten to fifteen minutes with each participant were conducted in conjunction with a picture-naming task. The findings revealed that there is inter-speaker variation among the two groups across all the target consonants. There is also a correlation between the migrants' L1, Urdu and their production of uvular and dental fricatives. This influence is a result of the presence of uvular fricatives in Urdu grammar, as well as the influence that Arabic has had on Urdu to some extent. Age does not have the same effect on the realisation of Arabic consonants as other variables, such as the amount of exposure and length of residence. These two independent variables demonstrate a more significant effect than age.

Albaqawi and Oakes (2019), Albaqawi (2020) and Almoaily (2012) investigated the effect of the L1s and LoR on their participants' use of ten GA morphosyntactic features, such as copula, definiteness marker, conjunction markers, object and possessive pronouns, as well as agreements in the verb and noun phrase. Almoaily (2012) conducted his study on sixteen male migrant workers in Riyadh and Alkharj, two cities in central Saudi Arabia. These participants speak Malayalam, Punjabi and Bengali as their L1. Conversely, in the other two studies, 72 female migrant workers working in Riyadh were examined. The L1s of the participants are Bengali, Sinhala, Punjabi, Tagalog, Malayalam and Sunda. L1 and LoR reveal a negligible impact in Almoaily's study. In other words, they affect only a small number of the dependent variables, the conjunction markers. Similarly, both of the predictors appear to have a minimal impact on the females' production of the morphosyntactic features in the other two studies. Regarding the impact of the migrants' L1, significant variation to the GA system was discovered in nominal agreement and conjunction markers. LoR in the Gulf significantly impacts only the definite article and the conjunction markers. Moreover, Albaqawi (2020) included gender as an independent variable in her study. She compared her result with the males' result in Almoaily's (2012) study. Albaqawi discovered that there is a significant difference between genders based on length of residence in the use of only one feature, namely conjunction markers.

Finally, Alhazmi (2020) conducted his study in the province of Madinah. He compared the Arabic speech of six male migrant workers who work and live in the city centre with the speech of four males who work and live in villages around Madinah. He interviewed each person for approximately an hour and a half. The former group came from Bangladesh, India and Pakistan and their length of residence ranged between 15 to 30 years. The latter are from Bangladesh and Pakistan and their length of residence ranged between 5 to 10 years. Alhazmi discovered that there is a statistically significant difference between the two groups in the use of pronominal subjects, such as overt or null, verb inflection and the predicational function of the existential which are /fi:/ and /ma: fi:/. He concludes that the language systems of the two groups are distinct. Whereas urban workers tend to use pidgin, rural workers tend to exhibit a higher level of proficiency in Bedouin Hijazi Arabic.

The studies mentioned above illustrate that the linguistic structure of the varieties of the migrant workers is characterised by simplification and reduction in relation to syntax, phonology, lexicon and morphology compared to the superstrate language, Arabic. If I discover simplification in the speech of the migrants with regard to the features that I examine, then I expect that the Arabic varieties of the migrants might be pidgin.

2.11 Conclusion

The focus of this chapter was on reviewing the literature on interlanguage and pidgin as they are two outcomes that might describe the Arabic varieties of migrant workers in Madinah. The chapter also demonstrated how certain variables significantly influence the speech of L2 speakers. The internal and external predictors I have reviewed here - L1, sex, amount of Arabic and L1 used, motivation, length of residence, informal exposure, formal instruction, the identity of the migrant, age, educational level in their L1, proficiency and language anxiety - could also impact the speech of the migrant workers in my study. If I establish that these contextual factors have significant effects on migrant workers' variable use of HA features, this may indicate that their Arabic varieties are interlanguage. In contrast, if their varieties do not mirror HA, the predictors do not point toward HA being learned, and they have little effect on their use of HA features, then I expect that their varieties are pidgin. The next chapter sets out my methodology.

Chapter 3: Methodology

3.1 Introduction

This methodology chapter provides an overview of the substrate languages that migrant workers in Madinah speak, such as Bengali, Hindi-Urdu and Tagalog. This is relevant because my research focuses on how participants' L1 impacts their use of Arabic L2 features. As I examine vulnerable populations, the chapter presents the procedures followed to protect their rights and maintain ethical issues. The chapter will also describe the strategies employed to create trust with my participants since these processes made it possible for them to have open and sincere discussions relating to their opinions and life experiences. This chapter explains that the selection of the participants in my research was carried out according to certain inclusion and exclusion criteria. The purpose of this is to ensure that they speak a certain L1 and are exposed to the same input.

In this chapter, the methods that are used to recruit the participants are also discussed. These include judgement sampling and the friend-of-a-friend method (snowball sampling), which are influenced by my research objectives. These approaches assisted me in researching difficult-to-reach people. To gather significant data, the chapter demonstrates the use of sociolinguistic interviews as well as quantitative questionnaires. The former enabled me to collect the tokens that corresponded to the investigated variables. The latter helped to classify people according to social variables that may impact their use of HA features. The chapter also explains how the interviews were conducted online via Zoom to overcome certain obstacles posed by the COVID-19 pandemic. Finally, the chapter will explain the data analysis process.

3.2 Non-migrant Social Groups and Dialects in Hijaz

In this section, the two distinct varieties of Arabic spoken in Hijaz – each associated with a different social group - are introduced. I have not been able to establish which of these social groups the migrant workers in my study interact with the most and which of their two distinct varieties they are exposed to the most. Hence, the analysis of my study focuses on the features shared by both social groups and found in both of their varieties.

In Saudi Arabia, there are three main dialects (Omar 1975: V). These comprise: the Najdi dialect, which is spoken in Riyadh, the capital city of Saudi Arabia, around Riyadh and in part of the northern centre of the country; the Shargi dialect, which is used in the eastern region; and the Hijazi dialect which is spoken on the West coast and includes the two holy cities of Makkah and Madinah, (see Figure 1) where the data for my study were collected. While the Najdi dialect is described as conservative, close to modern standard Arabic, and is the dialect used by the Saudi royal family, throughout the country, Hijazi is used by the government and for commercial reasons. Moreover, across the Arabian Peninsula, the Hijazi dialect has become the most widely understood (Omar 1975: V).

Figure 1: Hijaz Region, Saudi Arabia (Source: Google Maps).

The urban centres of Hijaz, i.e., Makkah, Madinah and Jeddah are home to two main social groups (Al-Jehani 1990). The first group is the Bedouin Hijazi. This group includes those of tribal descent who originated in the Arabian Peninsula and settled in the interior of the Hijaz region (Al-Jehani 1990; Nahedh 1989). Historical records reveal that the Bedouin tribes in Hijaz originally worked as farmers and herders, as they primarily inhabited the rural areas (Al-Jehani 1985). The country's economy accelerated after oil was found in the 1930s. Owing to this, people began to live in urban

areas. Several large tribes of the Bedouin Hijazi, for instance the Quraish, Juhaina, Otaiba and Harb, inhabit the Hijaz region (Rutter 1931). The last one is the dominant tribe in the region and the most sizeable.

Settlers or newcomers to the Hijaz region represent the second social group (Al-Jehani 1990; Nahedh 1989). They mostly came for religious purposes. This group is termed the Urban Hijazi. It includes people from different ethnic groups throughout the Muslims world, such as Turkey, Iraq, Egypt, Tunisia and Indonesia (Alansari 1970). Most people of this background have generational ties to the region and have Saudi nationality due to a 1953 Ministry of Interior Affairs declaration that all residents and their families who had stayed in Saudi Arabia for a long period of time or were born in the country would be granted Saudi nationality[5] if they wished. Urban Hijazi Arabic speakers have created a dialect that is mutually comprehensible and indicates their identity as newcomers (Al-Jehani 1985).

Due to these two distinct social groups who live together in the same region, city and location, two dialects are spoken (Al-Jehani 1990; Al-Jehani 1985; Nahedh 1989). There is inconsistency in regard to labelling these dialects. Alzaidi (2014: 73) indicated that there are two principal dialects in Hijaz, namely Urban and Bedouin. The former is spoken in the cities, whereas the latter is spoken in the countryside. Alqahtani and Sanderson (2019) also use the terms Urban and Rural. Conversely, Al-Mozainy (1981), who conducted his study of vowel alternations in the city of Madinah, referred to the dialect that he studied as Bedouin Hijazi Arabic. In research on the social and stylistic stratification of Arabic in Makkah, Al-Jehani (1985: 33&34) defined the two forms of Arabic that he investigated, as the sedentary variety, which he referred to as Al-Hadur and the nomadic variety, which he termed Al-Badu. Given that numerous Bedouin people have moved from the villages to the cities, it is established that Alhazmi (2018) identified them as Urban Bedouin Hijazi and Hadari Hijazi. Omar (1975) and Sieny (1972) described the features of the variety of the sedentary, Al-Hadur or Hadari Hijazi people and referred to it as Urban Hijazi Arabic. Therefore, in my study, the dialect that is used by the newcomers in Hijaz regardless of their origin or the country they originate from, is described as Urban Hijazi Arabic (UHA). And

[5] The situation is now more stringent and the requirements are much more extensive.

similar to Al-Mozainy (1981), the dialect that is spoken by the Bedouin is termed Bedouin Hijazi Arabic (BHA).

UHA and BHA are mutually intelligible; however, they do exhibit some differences, particularly in phonology. Al-Jehani (1985: 84) suggests that BHA speakers produce fricative dental sounds, such as /θ, ð and ðˤ/, whereas UHA speakers produce /θ/ as [s, t], /ð/ as [z, d] and /ðˤ/ as [z, dˤ], i.e., BHA dental fricatives are realised as stops and sibilants in UHA. As Bedouin dialects have many features that resemble the features of Classical Arabic which are lost in Urban dialects, Bedouin dialects are considered more conservative than the other (Rosenhouse 2006: 259). While the former is considered 'pure', the latter is considered to be more 'corrupt' by the general public (Miller 2007: 4).

When we compare HA with other main dialects such as Najdi Arabic, it is clear that HA has received little interest in the academic literature. In contrast to the wider literature on other varieties, there are only two books that have concentrated on UHA (Sieny 1972; Omar 1975). In terms of the phonological characteristics of UHA, Abaalkhail (1998) and Jarrah (1993) conducted research that predominantly deals with these aspects. They investigated vowel change and syllabification.

To the best of my knowledge, there is no research that provides a comprehensive definition of BHA. However, researchers who focused on Bedouin Hijazi mostly describe a specific tribal dialect. This suggests that tribes may differ linguistically. Despite using the term 'Bedouin Hijazi' to characterise the dialect he outlines in his thesis, Al-Mozainy (1981: 14) only examined samples from the Harb tribes. Given that the Harb tribe is one of the largest tribes in Hijaz, I take the features of the Harb tribes' Hijaz variety described in Il-Hazmy (1975) to be features of BHA.

3.3 Substrate Languages of the Migrant Workers

The section provides a brief introduction to Bengali, Hindi-Urdu and Tagalog. These are heritage languages in South Asia spoken by three migrant groups whose L2 Arabic is studied in my thesis (see Figure 2). These languages are presented here because I am testing their impact as the L1 of

the migrant workers in Madinah on their use of Arabic L2 features. I will introduce features of the substrate languages that are of relevance to my investigation in 4.2, 5.2 and 6.2.

Figure 2: South Asia showing India, Bangladesh and the Philippines. And KSA (Source: Google Maps).

3.3.1 Bengali

Bengali, also known as Bangla, is an Indo-Aryan language (Katzner & Miller 2002: 15). It has official status in Bangladesh and the state of West Bengal, India, while also being spoken in other countries, such as the UK, US, Nepal and Singapore (Gordon 2005). Based on Ethnologue (2017), approximately 267,653,920 people globally speak Bengali as their native language, half of which live in Bangladesh. Bengali is among the top ten spoken languages in the world (David 2015). It is a diglossic language and comprises two types or styles, namely *Sadhu bhasa* and *Colti bhasa* (Klaiman 2009: 434). *Sadhu*, which is a high variety, makes extensive use of Sanskrit terms and numerous archaic word forms (Nasrin & van der Wurff 2009: 6). *Colti bhasa* is the variety that is used every day in informal discourse and largely in speech (Klaiman 2009: 434). Recently, it has been being used in written expression and situations relating to formal discourse. For the purpose of my research and in keeping with Almoaily (2012), I am interested in *Colti bhasa* because, based on its characteristics and its similarity to the spoken language of the Bengali participants, I

hypothesise that it may have a more significant influence on their production of HA. This is in contrast to *Sadhu bhasa*, which, being a high form, may have less direct impact on their colloquial language use.

3.3.2 Hindi and Urdu

Hindi is an Indo-Aryan language which is a member of the Indo-Iranian branch of the Indo-European family of languages (Kachru 2009: 399). In India, Hindi has a special status and it is spoken by the largest population in this country. It is the official language besides English of the Indian Union and in eleven state governments including Delhi (Koul 2008). Additionally, in other states where Hindi is not spoken as the first language, it is taught as a second language. Thus, in India, more than three hundred million people speak Hindi as their first or second language (Kachru 2009: 399). Urdu is an Indo-Aryan language that is closely related to Hindi (Kachru 2009: 399). It is the official language of Pakistan. Additionally, in a few Indian states, such as in Kashmir, Jammu and Utter Pradesh, it is the state language. Twenty-three million people in India and nearly eight million people in Pakistan speak Urdu as their mother tongue.

Although there are minor differences between Hindi and Urdu in phonology, morphology and syntax (Kachru 2009: 400), linguists typically use the term Hindi-Urdu to examine the shared elements in these two languages (Islam 2011: 9). The common form of Hindi and Urdu is known as Hindustani which is essentially a mix of Hindi and Urdu (Koul 2008: 1). During the struggle for freedom, Hindustani was adopted by the Indian National Congress and Mahatma Gandhi as a national identity symbol, although it never became a language of high culture and literature (see Bhatia 1987).

3.3.3 Tagalog

Tagalog or Filipino is an Austronesian language. Approximately 80 million (90%) of the population of the Philippines speak it as their first or second language (Schachter & Reid 2009: 833). In 1521, the Spanish arrived in the Philippines and Tagalog speakers migrated to the south of Luzon Island, located in the north of the Philippines. In Manila and surrounding regions, Tagalog has become the main language in recent years. Over time, it has spread across the entire country and is now the second language after English. In 1937, Tagalog became the national

language of the Philippines and nowadays, it is taught in schools and used in mass media. Tagalog is the native language of people who live in the largest city in the Philippines, Manila. Furthermore, in many cities that have mixed populations, it is used as the lingua franca (Schachter & Reid 2009: 833).

I have given a short description of Bengali, Hindi-Udru and Tagalog; the L1s of migrant workers in Madinah. My study will establish whether the Arabic L2 varieties of these migrants are pidgin or interlanguage varieties. I hypothesised that the migrant workers' Arabic L2 can be classified as a pidgin variety if the following conditions hold true: their Arabic L2 is simpler than HA; they mostly delete the morphosyntactic features; and replace /f/ with other variants; and contextual predictors have little effect on their L2 production. However, the migrant workers' Arabic L2 can be regarded as interlanguage varieties if their use of the target variables is significantly conditioned by contextual factors.

The following are the ethical issues encountered when collecting data from migrant workers who speak the above-mentioned languages as their L1.

3.4 Ethical Issues

This research has obtained ethical approval from the Faculty of Humanities and Social Sciences Ethics Committee at Newcastle University (01/10/2020 - 30/06/2021 & reference number 4600/2020).

Quest and Marco (2003: 1297) applied the term 'vulnerable' to refer to individuals characterised by 'social vulnerability.' This includes children, older people, homeless people, the unemployed people, in addition to ethnic groups and religious minorities, migrants, people with learning disabilities, prisoners, people who are educationally and economically disadvantaged, low status populations and refugees (Flaskerud & Winslow 1998; Moore & Miller 1999; Nyamathi 1998; Quest & Marco 2003; Russell 1999; Stone 2003; Weston 2004). Thus, the participants in my study are considered vulnerable minorities as they are migrants from different heritage language backgrounds working in Madinah.

Numerous ethical issues can arise when conducting research with vulnerable people, including migrant workers (Liamputtong 2006: 23). Van Liempt and Bilger (2018: 273) and Briman (2005: 157) suggest that migrants might experience difficulties in settling into communities in the country of arrival because they regularly find themselves discriminated against, isolated and excluded. Therefore, migrants might distrust their environment (Van Liempt & Bilger 2018: 273). Miranda *et al.* (1996) note that migrants and minorities are often reluctant to participate in studies and are unwilling to discuss their experiences (Dickson-Swift 2005). Cultural and language factors are also among the ethical issues in studying migrants because these factors might influence their understanding of what study participation involves (Dickson-Swift 2005). Additionally, regarding socio-economic status, education, rights, psychological position and culture, there are differences between the researcher and the immigrants (Van Liempt & Bilger 2018: 278; Van Liempt & Bilger 2012: 457). As a result of this unequal relationship that exists between them, there may be instances where migrants involuntarily participate in a study (Van Liempt & Bilger 2012: 457). Occasionally, migrants might find it hard to refuse to participate in a study, as a result of social pressure. For instance, Van Liempt and Bilger (2012) indicated that, after nearly an hour of interviewing, participants anxiously questioned the sufficiency of the information they had shared. This could be attributed to the power imbalance between the researcher and the interviewees, potentially influencing disclosure. The interviewees were somehow obliged to answer questions and compelled to share information that they prefer not to share. Such experiences or perceptions may raise ethical concerns (Daly 1992; Van Liempt & Bilger 2012).

Complying with due ethical process is critical and required for all research studies, including those involving vulnerable participants such as migrants and minorities (Liamputtong 2006). Even though the right to withdraw and decline is always available (Christians 2005; Sin 2005). This is particularly important when the participants are not L1 speakers of the host language or are otherwise vulnerable, for example, because they are members of an ethnic minority; such situations may exacerbate language issues such that eligible participants might not understand what they are consenting to, i.e, their consent is not in fact informed (Booth & Booth 1994). Barnard (2005: 14 -15) proposes that informed consent 'is much more elusive' when conducting research on vulnerable people. Discussion and negotiation are acceptable, but when a participant refuses, researchers must acknowledge the refusal. Wenger (2002: 265) maintains that it is never

appropriate for the interviewer to push too hard on the participant's part. Researchers should achieve a rapport, develop trust and make the participants feel comfortable about talking and sharing their life experiences. This is of significance because some participants might be suspicious about and sensitive to being subjects to the study (Berk & Adams 2001: 63). My study has taken into account all of these ethical considerations to guarantee that the participants' rights were protected and respected. Upholding ethical standards and establishing a foundation of trust with participants was the major objective since it would allow them to freely share their experiences and opinions. The following paragraphs explain how the ethical considerations in my study were addressed.

Certain procedures were required to address the main ethical challenges. Initially, the consent form[6] and the participant information sheet [7](PIS) were translated into the participants' L1- Bengali, Hindi, Urdu and Tagalog; this was done to ensure eligible participants' comprehension of the nature of their participation, i.e. ensure that their consent is in fact informed (see Appendices D and E for the English version of the consent form and participant information sheet, respectively). To gain the trust of the participants and encourage them to spend time with the interviewer, the general aim of the study was explained to them as conducting fieldwork on migrant workers' discourse and their experience of living in Saudi Arabia, without explicitly drawing attention to the features under study. When researchers introduce themselves to the participant via a referral of a friend, participants will not consider them strangers or outsiders as to when they introduce themselves as researchers only (Schilling-Estes 2007: 179). Therefore, the 'friend of friend' method for recruiting participants , was used in my study (see further Section 3.6).

In my research, all participants were informed that their participation is voluntary and the benefit that they will have is the contribution to our understanding of the human language. They were also notified that they had the right to withdraw from the study at any time without giving any reason and without any penalty. Additionally, they were informed that if they decide to withdraw, any

[6] For the consent form in the migrant workers' L1, refer to appendices H, L, P and T in the following: https://drive.google.com/file/d/1-_Qct2BBzkXCLg72IcreLEYCTPO-z9ZN/view?usp=sharing.

[7] For the PIS in the migrant workers' L1, refer to appendices I, M, Q and U in the following: https://drive.google.com/file/d/1-_Qct2BBzkXCLg72IcreLEYCTPO-z9ZN/view?usp=sharing.

data that they have provided up to that point will be omitted. They could also refuse to answer any question or stop the recording. The interviewees were informed that no one has access to the participants' research data except the researcher and her supervisors. Additionally, their personal information and their recordings were transcribed and stored anonymously.

3.5 Selecting Speakers

One issue that researchers encounter when they design a variationist study is how to select participants (Milroy & Gordon 2003: 26). Sankoff (1974: 21) suggests that researchers must decide what their 'sampling universe' is. According to Sankoff, a sampling universe is the societies or groups that the researcher would like to investigate and make generalizations about, such as the members of a reading group. The sampling universe in my study is migrant workers with certain heritage languages, namely Bengali, Hindi, Urdu and Tagalog, in Madinah. The following inclusion criteria were met by the participants under investigation:

- Their L1 is Bengali, Hindi-Urdu, or Tagalog.
- They are of Asian heritage and originate from Bangladesh, India or the Philippines.
- They are employed in low-paid service in Madinah, working as cleaners, waiters, grocery sellers, etc.
- They are literate in their L1.

The exclusion criteria for participation in the study are:

- Migrant workers who speak Arabic as L1.
- They have worked in Arabic cities or countries other than Madinah or the Hijaz region in general.
- They work as housemaids.

Below I justify my choice of these inclusion and exclusion criteria.

30 migrant workers were recruited (15 males, 15 females). To test the effect of participants' L1 on their production of their L2 Arabic variety, I had planned to stratify participants into three different heritage language backgrounds (Bengali, Hindi and Tagalog), with 10 speakers per heritage language group and evenly divided by speaker sex. However, not all of the participants in Madinah

who speak Hindi-Urdu as their L1 describe it as Hindi. A number may refer to their L1 as Hindi, while others may identify it as Urdu. I therefore included any Hindi-Urdu-speaking worker who speaks Hindi, Urdu, or both Hindi and Urdu as their L1 and who also meets my inclusion and exclusion criteria. Thus, the participants were stratified based on their L1, with 10 Bengali, 10 Hindi-Urdu and 10 Tagalog L1 speakers, with each group stratified by speaker sex. Participants had to be literate in their L1 to be able to answer the questionnaire written in their L1. My study does not include HA speakers as participants, due to time constraints and it is not one of the study's objectives.

Since I am focused on participants who were only familiar with HA dialect, migrants who had previously worked outside Madinah or the Hijaz Region were excluded from participation. This is to ensure that migrants are exposed to the same language input. Additionally, housemaids were not considered in the sample universe because I attempted to make the male and female samples comparable in terms of language input they receive and interactions they regularly participate in. Housemaids' language contact is mostly with L1 Arabic speakers because they live with them in the same house and are therefore likely less exposed to language input from non-native speakers of Arabic. For the reason that I hypothesise that the Arabic variety of migrant workers might be a pidgin if the contextual predictors exhibit no significant effect on their use of Arabic L2 features, I included in the speakers sample only those migrant workers who can be assumed to regularly speak Arabic with their employers as well as their co-workers who are non-native speakers of Arabic or mutual mother tongue.

In general, participants in my study work as cleaners, messengers, grocery salesmen, waiters and drivers. Female participants are mostly cleaners, whilst the messengers work in the Prophet Mohammad mosque, companies, and the university, as well as Saudi Airlines and various malls. To test the effect of length of residence, migrants vary in their length of stay. The number of participants who have lived in Hijaz for over 10 years is higher among L1 Bengali speakers than among L1 Hindi-Urdu and L1 Tagalog speakers. The participants' ages range from 26 to 54. Table 1 lists the participants, their L1, their sex, their age along with years spent in Hijaz.

Table 1: Participants' demographic information

ID	L1	Sex	Age	Years in Hijaz
BF1	Bengali	Female	29	7 years
BF2	Bengali	Female	34	3 years
BF3	Bengali	Female	26	4 years
BF4	Bengali	Female	30	4 years
BF5	Bengali	Female	32	7 years
BM1	Bengali	Male	26	3 years
BM2	Bengali	Male	42	18 years
BM3	Bengali	Male	44	18 years
BM4	Bengali	Male	41	16 years
BM5	Bengali	Male	44	18 years
HF1	Urdu	Female	48	4 years
HF2	Hindi	Female	36	4 years
HF3	Hindi & Urdu	Female	35	4 years
HF4	Urdu	Female	35	8 years
HF5	Urdu	Female	32	3 years
HM1	Urdu	Male	32	9 years
HM2	Urdu	Male	47	25 years
HM3	Hindi	Male	54	17 years
HM4	Hindi	Male	32	6 years
HM5	Urdu	Male	51	22 years
FF1	Tagalog	Female	34	9 years
FF2	Tagalog	Female	45	9 years
FF3	Tagalog	Female	33	4 years
FF4	Tagalog	Female	51	17 years
FF5	Tagalog	Female	29	4 years
FM1	Tagalog	Male	43	9 years
FM2	Tagalog	Male	48	13 years
FM3	Tagalog	Male	37	12 years

FM4	Tagalog	Male	36	9 years
FM5	Tagalog	Male	34	7 years

After defining the sampling universe, I decided on the sample size. Compared to the social sciences research, the sample size tends to be much smaller in linguistic studies (Buchstaller & Khattab 2013: 82; Milroy & Gordon 2003: 28). Many variationist studies such as the studies undertaken by Labov (1966a) and Trudgill (1974) were based on relatively small samples. Although in linguistic studies researchers often rely on fewer participants, they collect a large number of observations from each participant (Buchstaller & Khattab 2013: 82). Meyerhoff and Schleef (2010) contend that to make statistically valid generalizations based on the collected data, a minimum of five or six participants per cell is generally considered sufficient (Guy 1980). If this suggestion is applied for every single independent variable considered in my study, it will create a huge amount of data, too time-consuming to collect, process and analyse within the duration of my project. Therefore, the sample size is principally determined by practical reasons, specifically, how much data it is practical and feasible to analyse with the available resources such as time (Buchstaller & Khattab 2013: 84). Seeing as the sampling population in my study are migrants with three different heritage languages in Madinah and that I want to generalize across speaker sex, a of 30, with 5 speakers by L1/speaker sex cell, seems adequate. The sample was very unevenly distributed across some predictors: formal Arabic instruction in their country, the identity of the migrants and educational level in L1 (see Appendix A). This needs to be considered in the analysis regarding representativeness.

3.6 Sampling Methods

In my study, the participants were recruited via judgment sampling and social network, or snowball sampling (Tagliamonte 2006; Milroy & Gordon 2003). These data collection approaches were motivated by the aims of the study.

Judgment sampling is used in many sociolinguistic studies, for instance, the study completed by Roberts (2015) and Wolfram and Fasold (1974). Researchers use this method when they identify the type of participants that are suitable for their studies in advance of participant recruitment and stratify the sample based on different factors of interest in the study, such as ethnicity, age, social

class and sex (Tagliamonte 2006). Given that my study aims to discover which type of Arabic varieties the migrants speak and if some social factors play a role in the use of HA features, there are certain inclusion and exclusion criteria and the participants were stratified based on their L1 and their sex. Hence, judment rather than random sampling was used.

To find the participants, I contacted several migrant workers that I know. However, these workers could not participate in my study because they were unable to meet my inclusion criteria. However, they referred me to friends. Snowball sampling was employed. Ruane (2005: 117) describes snowball sampling as a process that begins by contacting some members of the sampling population to participate in the study. Once they have finished participating, the researcher asks them to suggest other participants that meet the inclusion criteria. Naderifar, Goli and Ghaljaie (2017) suggest that when the target participants are hard to reach, snowball sampling is used. Buchstaller and Khattab (2013: 80) assert that if a researcher has been referred to the participant by their friend, they are less likely to refuse a request for an interview. As the migrant workers are vulnerable, it might not be easy to reach them without a referral, especially given the high number of people who refused to participate, possibly due to the power dynamic between me, i.e., my status as a native speaker of HA, holding Saudi nationality and not working in comparable job to that of the the potential participants. While several potential participants were reluctant and did not want to be recorded, others were too busy to participate. Many potential participants refused to participate without providing any reason for their negative response. Therefore, it was determined that using the snowball approach can mitigate recruitment challenges.

3.7 Data Collection Instruments

In my study, semi-structured sociolinguistic interviews and quantiative questionnaires were used as data collection instruments. This choice was motivated by the aims of the study.

3.7.1 Sociolinguistic Interviews

Labov (1981) suggests that one effective way to collect linguistic data and analyse it quantitatively is by using a face-to-face sociolinguistic interview. Researchers can obtain a large amount of speech in a short period of time by using this data collection technique (Schilling-Estes 2007: 172).

Despite the value of the sociolinguistic interview, it has a few limitations (Schilling-Estes 2007: 173). Compared to a spontaneous conversation, the sociolinguistic interview data have criticised for being less natural than naturally-occurring speech data. Many people are familiar with interviews that follow straightforward questions and answers like the ones on television. Typically, people are unaware of how spontaneous and conversational interviews can be. Consequently, interviewees might resist attempts of the interviewer to let them talk at length about chosen topics. Instead, interviewees may seek to provide brief or short responses. Similarly, by using sociolinguistic interviews, some speech low-frequency linguistic phenomena might be impossible to examine, for example, some syntactic variables such as relative clauses (Milroy & Gordon 2003: 63). The occurrence of this variable is infrequent in daily conversation and even more so in interviews, and this might create challenges for quantitative analysis.

Labov (1966b) suggested that using sociolinguistic interviews for data collection does not guarantee eliciting the vernacular because the interview setting can create a level of formality. The participant may be aware that their language is being observed and the presence of a recording device creates what Labov (1972) termed the 'observer's paradox'. Labov (1972: 209) stated that 'the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain these data by systematic observation.' In addition, Milroy (1987) maintains that one of the disadvantages of the sociolinguistic interview is the result of an unequal power dynamic between the interviewer and the interviewee, which negatively impacts the recording of natural speech. This power dynamic may be exacerbated in the context of my research. Because of their migrant status and language skills, migrant workers may feel that they are in a subordinate position to the interviewer. Nonetheless, the sociolinguistic interview continues to be a key instrument in variationist research (Schilling-Estes 2007: 174).

After describing the advantages and disadvantages pertaining to sociolinguistic interviews, the following paragraphs will describe the semi-structured sociolinguistic interview in my study, how it can be improved and how the effect of observer paradox can be reduced.

To collect vernacular langauge data from migrant workers, semi-structured sociolinguistic interviews were used in which participants were asked open-ended questions about their work, traditions, school days, etc., (adopted from Tagliamonte 2006). This method elicits large amounts of vernacular speech data for quantitative analyiss (Labov 1972). Before the data collection, I provided instructions to my male family member assistant on how to conduct a semi-structured sociolinguistic interview with the male participants, considering gender segregation in Saudi Arabia. Beforehand, I had provided him with guidance on how to interact with vulnerable participants and also addressed the ethical considerations that may arise during the data collection phase. I carried out this procedure to achieve maximum comparability between the groups of males and females.

To improve the sociolinguistic interview, Llamas (1999) argues that it is important to make it a pleasant event for the interviewees and make it as casual as possible with the aim of eliciting natural speech from the participants. Therefore, participants were allowed to discuss any topic they chose because the interviews were informal and resembled a chat as they were loosely structured. Likewise, other topics were developed naturally based on the participants' previous answers. Thus, this type of method allows not only free conversation, with minimal intervention by the interviewer, but also the collection of attitudinal and other information which are related to the daily life of the participants (see Labov 1972). To reduce the researcher's authority, Labov (1984) recommends that the amount of talk that is produced by the interviewers during the interview should be minimised. This can be achieved by making the questions as short as possible, which in turn will provide the interviewees with more time to talk. Consequently, I have attempted to keep the questions succinct and clear to give more time for the interviewees to speak. Additionally, to minimise dominance in the interview situation, interviewers should avoid relying significantly on a written interview schedule (Schilling-Estes 2007: 181). Therefore, a list of the interview questions was used as a guide only (see Appendix B).

The observer paradox can never be completely overcome. To mitigate its effect in my interviews and to record the participant's natural speech, similar to when they are not being observed or recorded, I not only used the sociolinguistic interview, but I also asked the participants questions related to storytelling which are included in Labov's interview schedule (see also Poplack 1989

and Tagliamonte 2006). Moreover, the first five minutes of each conversation were not recorded to assist the participant to settle into the situation (Drummond 2010: 87). During Almoaily's (2012) fieldwork trip, swine flu outbreak. Consequently, he found that asking participants about this topic assisted him not only in reducing the impact of the observer's paradox but also encouraged them to produce long turns. For instance, Almoaily (2012: 125) asked them if they 'are worried about swine flu'? Similar to swine flue, the outbreak of Covid-19 had an impact on my fieldwork trip. Therefore, I asked participants some questions about it, such as 'Do you know anyone who has been affected by or died from the virus? How did they get the virus? How did they recover?' During that time, COVID-19 was a prevalent and a topic of discussion. Therefore, I thought that asking participants about it might be beneficial, as it allowed them to talk more easily and possibly distracted them from being aware that they were being recorded, and in turn, may have reduced the effect of the observer's dilemma.

3.7.2 The Questionnaire

Questionnaires are a common method in the quantitative method (Dörnyei 2007). Ruane (2005: 123) argues that questionnaires can 'be an extremely efficient data collection tool'. They have many advantages for data collection, especially for accessing participants' opinions and attitudes. In general, we can acquire reliable information from this type of research instrument (Patten 2001). It is also an effective method (Cohen, Manion & Morrison 2007). Researchers can also use it to collect sensitive information (Colton & Covert, 2007: 11). This type of data collection tool has been used in sociolinguistic and L2 acquisition studies in conjunction with interview techniques. For example, the studies of Abed (2017), Drummond (2010) and Major (2014).

Designing a questionnaire is not easy (Newing 2011: 119-120). Researchers should pay attention to detail and meticulously consider the questionnaire. The questions should be carefully worded to ensure that the informants understand the questions. Moreover, they should use words that are appropriate and that the participants are culturally and socially familiar with. They also should avoid using unambiguous words because they risk have an impact on how participants respond to the questions. As it is a well-known fact that creating a reliable questionnaire is often a difficult and time-consuming task, Gillham (2002), Cornford and Smithson (1996) proposed a solution. Their recommendation is to use a questionnaire that has previously been used in other studies.

Therefore, prior to deciding which questionnaire should be applied in my research, many articles and studies that focus on similar social features as regards my research were reviewed and piloted in order to choose a questionnaire that was suitable for the aim of my study and for the targeted population of migrant workers.

In my study, the questionnaire (see Appendix C) was adapted from the study undertaken by Drummond (2010). The purpose of this questionnaire is to categoririse participants based on information they provided, specifically, on social factors assumed to impact thier L2 Arabic. The chosen questionnaire includes three sections. Section one gathers demographic information such as age, sex, L1, length of residence and job of the participants. It also includes some questions related to participants' exposure to Arabic such as watching TV and listening to the radio in Arabic, etc.

The second section is designed to collect information concerning their use and acquisition of Arabic:

- Self-evaluated for level of spoken Arabic.
- Amount of Arabic instruction.
- The amount of Arabic used and the participants' first language.
- Migration identity, future plan to settle in Saudi Arabia, returning to their countries, etc.

The third section accesses information about participants' attitude and motivation and includes the following five areas:

- Anxiety regarding Arabic.
- Attitudes towards Madinah, living there, its inhabitants and their variety.
- Instrument motivation.
- Integrative motivation.
- Distractor questions.

The answers to the questions in these three sections provide quantitative information about the social factors that help to determine what Arabic varieties the migrant workers speak (see Chapter 2). If these social variables show that they have a significant influence on the migrants' production

of HA features, then it is expected that their varieties are interlanguage. However, if the social variables reveal that they have a negligible impact on the migrants' production of HA features, then it is predicted that the migrants' varieties are pidgin languages.

In total, there are 18 statements in the third section. Dörnyei (2002: 34) reported that there is a general consensus among those who employ questionnaires that when multi-item scales are used, each main area that is addressed requires more than one question. Therefore, in my study, each area in section three consists of four questions except for the last one which comprises two distractor questions that are related to Arabic reading and writing ability. The purpose of adding these questions is to draw the participants' attention away from the actual focus which is their attitude and motivation. The layout of the questions in this section was influenced by the Attitude/Motivation Test Battery created by Gardner (1985a). A Likert scale that contains seven response options is provided under each question, with 1 meaning 'strongly disagree' and 7 meaning 'strongly agree'. The numbers between 1 and 7 remain unlabelled. Given that the answers are on the same scale, I can likely compare results across L1 groups.

The selected migrant workers in my study are literate in their L1 but may not be litereate in Arabic; therefore, the entire questionnaire[8] was translated into four different versions and presented to participants in thier L1. Although Arabic is the language under investigation, the research does not target participants' ability in Arabic reading. The questionnaire was used to gather informative answers rather than testing their capabilities in understanding Arabic texts. What is more, translating the questionnaire should prevent any misunderstanding that may take place if it is written in Arabic.

3.8 Procedures

My fieldwork was conducted between October 2020 and May 2021 in Madinah, which is located in the Hijaz region of Saudi Arabia (see Section 3.2). My research as well as the studies completed by Alhazmi (2020), Al-Shurafa (2014) and Alghamdi (2014) were conducted in the same region:

[8] For the questionnaire in the migrant workers' L1, refer to appendices G, K, O and S in the following: https://drive.google.com/file/d/1-_Qct2BBzkXCLg72IcreLEYCTPO-z9ZN/view?usp=sharing.

Hijaz. However, my study and Alhazmi's (2020) study were conducted in Madinah while their studies were carried out in Jeddah. My aims and variables are different to theirs (for more details see Section 2.10).

As data was collected during the Covid-19 pandemic and was dependent on Saudi government guidelines and the regulations in place in the Hijaz region at the time of data collection, the interviews were conducted online via Zoom, a cloud-based video communications App that allows virtual audio and video calling, for one hour. I only recorded the participants' voices and I did not use the video recording that Zoom automatically produces. In fact, I deleted it immediately to protect participants' identities. I did not save recordings to the Zoom cloud due to ethical reasons; instead I saved them to OneDrive. I also asked the participants to record the conversation with their mobile phones too, in case we experienced any difficulties with our recording. For ethical reasons, the participants were asked to send the recording to me immediately after the interview and then delete it immediately from their mobile phones. The format used in the audio files was WAV. While participants sent the recordings to me via email, they sent other documents to me via mail or email, such as the consent form after they signed it and the questionnaire after they had completed it, so as to safeguard ethics.

I conducted individual interviews with the female participants, while my assistant conducted the interviews with the males. Due to gender segregation in Saudi culture, I still required help from a male assistant. This was because it was challenging for me to find male volunteers on my own and talk to them, even though the interviews were conducted virtually. I provided him with a copy of the interview questions. Each interview was carried out at a time that the participants preferred. Additionally, I advised both the assistant and the participants to find a suitable place with little or no noise as that would enable a clear recording to be made.

Prior to starting an interview, when my assistant and I identified migrant workers who met the study inclusion criteria through a phone call, we introduced ourselves as 'a friend of Mr. X's', for example (Schilling-Estes 2007: 179). Afterwards, we clarified that I am a postgraduate student at Newcastle University and that the assistant is involved in the research. After being provided with

the general aim of the study, participants were also given a debriefing[9] sheet (see Appendix F for the English version) that elaborated on the study's objectives once they had completed the participation process. This was hoped to make them feel appreciated for participating in the study.

Participants who demonstrated their willingness to participate in my study and gave oral agreement were provided with the PIS and the consent form via email. They were asked to read each one and if they agreed to participate, sign and return them via email or mail. The questionnaire was sent to the participants via email at the end of the interview. It takes approximately 15-20 minutes to complete.

3.9 Variables of Interest

My study examined the use of six HA variables i.e., morphological, morphosyntactic and phonological features by migrant workers who speak Bengali, Hindi-Urdu and Tagalog as their L1 in Madinah. The selected characteristics for analysis in both UHA and BHA are comparable, as stated before (see Section 3.2). Additionally, there might be differences in the migrants' L1 (for further information, see Sections 4.2, 5.2, 6.2 and 7.2). These variations may potentially influence their production of HA characteristics. There is a possibility that their Arabic varieties constitute an interlanguage due to the significant impact of their L1 as well as other independent variables. In addition, the literature pertaining to pidgins suggests that these chosen features are environments in which we may expect to see simplification and L1 effects (see Albaqawi 2020; Almoaily 2012; Hobrom 1996; Naess 2008).

Morphological features

In UHA and BHA, the verb agrees with the subject in gender, number and person (Il-Hazmy 1975: 172; Omar 1975). Agreement patterns vary across tenses. Adjectives agree with the noun in gender, number and definiteness in UHA and BHA. HA nouns inflect for gender, number, definiteness and possessive markers. HA nouns consist of three 'number' forms: singular, dual and plural. With cardinal numbers, a plural inflectional morpheme is only used when the quantified

[9] For the debriefing sheet in the migrant workers' L1, refer to appendices J, N, R and V in the following: https://drive.google.com/file/d/1-_Qct2BBzkXCLg72IcreLEYCTPO-z9ZN/view?usp=sharing.

noun falls within the range of three to ten. When the number exceeds ten, the singular form of a noun is used. GPA studies have revealed that migrant workers have a reduced verbal and nominal system (Albaqawi 2020; Alghamdi 2014; Almoaily 2012; Alsusut 2022; Bakir 2010; Hobrom 1996). If the descriptive analysis in my study exhibits similar results, we can predict that the Arabic varieties of the migrant workers would be better described as a pidgin. These results specifically include simplification in the use of HA morphological features, for example a lack of verbal and nominal agreement.

Morphosyntactic features

- Definiteness: the definiteness marker in Arabic is the prefix *al-* 'the' (Il-Hazmy 1975; Omar 1975). In the GPA studies conducted by Albaqawi (2020) and Almoaily (2012), the results reveal that the contextual variables, for instance L1, LoR and gender, typically do not have a significant influence on the use of the definite article by migrant workers and they mostly omit the variable (see Section 5.1). If my study confirms similar findings to these studies, the Arabic varieties of migrant workers in Madinah may possibly be described as pidgin.
- Coordinating conjunction marker: there are many conjunction markers in HA like *laakin* and *bass* 'but', *willa* and *aw* 'or', and *wa* 'and'(Il-Hazmy 1975; Sieny 1972). My study will focus on *wa* only because it is among the common conjunction markers observed in the study undertaken by Almoaily (2012). The studies in GPA completed by Albaqawi 2020 and Almoaily (2012) show that the independent variables, for example L1, LoR and gender have a significant effect on the use of the conjunction markers (see Section 6.1). These findings are similar to the results in the literature on L2 acquisition (see 2.5.1, 2.5.2 & 2.5.4). Therefore, the significant effect of the contextual variables on the use of the coordinating conjunction marker by migrant workers in Madinah may denote that their Arabic varieties might possibly be described as an interlanguage.

Phonological feature

One of the HA consonants examined in my study is /f/ (Il-Hazmy 1975; Sieny 1972). I chose this variable because it has primarily been examined qualitatively in previous studies on migrants in the Gulf, for instance in the works of Hobrom (1996) and Naess (2008) (see Section 7.1). However,

in my study, I analyse it quantitatively. Similarly, in these qualitative studies, as /f/ reveals variation, I will discover if my findings confirm their results and if the production of /f/ is conditioned by the contextual variables.

3.10 Transcribing the Interviews

I transcribed the interviews myself orthographically in the Arabic language using Elan version 5.9 (Wittenburg *et al.* 2006), an annotation software for audio and video material. When used to transcribe linguistic data, it has many useful characteristics (Brugman *et al.* 2004): the audio file can be segmented into utterances that are manageable for transcription with the option of playback; individual varialbes can be searched for playback and coding.

In my study, auditory analysis was applied with the phonological feature. Up until recently, the most popular method for analysing phonological variation relied heavily on investigators' auditory judgments (Milroy & Gordon 2003: 144). Kerswill and Wright (1990) challenge the validity and reliability of this approach. In their experiment, thirteen phoneticians were tasked with transcribing utterances that included word-final /d/ assimilation to [g] in order to draw attention to the problems of the validity and reliability of auditory analysis. *Electropalatography* (EPG) technology, which captures the tongue's location in the mouth, was also employed to analyse the recorded utterances. They discovered that in several instances, the transcriptions of the phoneticians did not correspond to the acoustic data. They provide an explanation for this in terms of the phoneticians' prior knowledge of the phonological context of assimilation as well as their familiarity with articulatory phonetics, both of which had a negative impact on their judgment. Johnson (1993) compares the auditory and acoustic analyses of clicks and pulmonic sounds in Xhosa. To demonstrate the similarities between the two techniques when he investigated clicks, stops and fricatives, Johnson provides spectra for both of these analyses. He discovered that both the auditory and acoustic analyses reveal that the sounds' frequencies are identical, although the auditory spectrum offers a richer perspective than the acoustic version. According to Kerswill and Wright (1990: 273), the auditory and acoustic strategies used for transcription are generally plausible and interpretable and rarely conflict.

In connection with auditory analysis, one of the key issues is related to reliability (Milroy & Gordon 2003: 151). This concern pertains to consistency (Kerswill & Wright 1990: 258). Auditory analysis is more susceptible to subjectivity than acoustic analysis, which is one of the principal advantages (Milroy & Gordon 2003: 151).

My decision between acoustic and auditory analyses was determined by the aims of my study and the type of variable. /f/ displays discrete variants specifically [f, v, p, b] which are relatively easy to recognise auditorily compared to the continuous variables that exhibit many variants which range along a continuum, for instance vowels (Milroy & Gordon 2003: 144). I made several passes to improve the precision of my auditory coding. I also followed several of the suggestions made by Milroy and Gordon (2003: 151) to reduce possible risks and guarantee the analysis was strong. For example, I analysed a greater number of tokens and made sure there was inter-coder reliability by having a second coder painstakingly analyse even a small part of the data. To enhance the reliability of the coding process, I asked a Ph.D. colleague with training in auditory analysis to code some of the data independently. Subsequently, we discussed the similarities and differences extensively. After we investigated the differences, we reached an agreement that allowed us to make the required adjustment.

3.11 Coding of the Dependent Variables

There are two separate files for each interviewee. One of them is only for the interviewee's talk and the other one includes the interviewee's talk with annotation. In the annotation file, each variant of the linguistic variable was labelled with a unique code as shown in Table 2. This makes retrieving and accessing the tokens easy (Almoaily 2012: 128). In HA, the targeted morphosyntactic variables are categorical and the targeted phonological variable near-categorical (see Il-Hazmy 1975 & Sieny 1972). However, these vairables are treated as binary variables in my study, as supported by my data and observations from prior research including Albaqawi (2020), Almoaily (2012) and Naess (2008). Table 2 shows that, for the morphosyntactic features, any variant that has (+) indicates the typical feature of HA. In contrast, any variant that has (-) signifies different variants from that used in HA. To illustrate this, +DEF suggests that a definite article is used while -DEF denotes the variable is missing; +CONJ represents the presence of the conjunction marker, which is the standard form in HA, and -CONJ denotes the absence of this

variable. Regarding the phonological feature, I initially coded /f/ as having four variants: [f], [v], [b] and [p]. However, the frequency table in section (6.3) confirmed that I have many more [f] than other variants. Additionally, the preferred non-[f] variant for all groups is [p], followed by [v], with [b] negligible. Hence, it makes more sense to combine [v], [b] and [p] together and treat this variable as it has binary variants which are [f], the standard variant in HA, vs. non-[f]. Likewise, as HA does not have [p] and [v] in its phonemic inventory (see Tables 23 & 24) and to label these two variants, I used (*) as in [*b] and [*f] respectively (see Table 2). In the analysis, I will focus only on the frequency of +DEF, +CONJ and [f] as they are aligned with the research questions. By examining whether the independent variables have either a significant or insignificant impact on the production of HA characteristics employed by migrant workers in Madinah, it is possible to ascertain whether the varieties of Arabic they employ are indicative of an interlanguage or pidgin system.

Table 2: Arabic codes for the variants with their meanings

Arabic Code	Meaning
(ال التعريف +)	The definite article is used (+DEF)
(ال التعريف -)	The definite article is missing (-DEF)
(حرف العطف +)	The coordinating conjunction marker is used (+CONJ)
(حرف العطف -)	The coordinating conjunction marker is missing(-CONJ)
(ف)	/f/ is present [f]
(*ف)	/f/ is dropped and replaced by [v]
(ب)	/f/ is dropped and replaced by [b]
(*ب)	/f/ is dropped and replaced by [p]

Any word that was produced by the migrant workers and included any of the variants for the particular linguistic variable of this study was identified as a token. I ran the annotation file for each participant in AntConc software (Anthony 2011) to extract the tokens. Froehlich (2015) asserted that identifying patterns in a language by just reading the text would be difficult. However, AntConc is a useful toolkit with respect to finding eligible tokens.

To search for a specific variant in Antconc, I wrote for example, (+ال التعريف). The concordance view revealed whenever (+ال التعريف) appeared in the file and it illustrates the context around this variant which will help in establishing the effect of linguistic predictors. Additionally, it showed how many times the token occurred. Table 3 shows that across the whole speaker sample, there are 9941 tokens for the definite article, 2604 tokens for the coordinating conjunction marker, and 14206 tokens for /f/. Split by L1 group, we have these number of definite article tokens: Bengali (N=2945); Hindi-Urdu (N=4084); Tagalog (N=2912) (see Table 10). The tokens of the coordinating conjunction marker are distributed for each language group as follows: Bengali (N=1110); Hindi-Urdu (N=1006); Tagalog (N=488) (see Table 17). The tokens of the /f/ are distributed for each language group as follows: Bengali (N=5357); Hindi-Urdu (N=5324); Tagalog (N=3525) (see Table 30).

Table 3: The number of tokens for each linguistic variable

Feature	Tokens
The definite article	9941
The conjunction marker	2604
/f/	14206

3.12 Coding of the Independent Variables

Table 4 below lists each predictor with its predictor levels that are examined in my study. While most of these variables are categorical, six - specifically motivation, attitudes toward HA, Madinah and its population, LoR, age, proficiency and language anxiety - are treated as continuous variables given that they lack discrete predictor levels.

Building upon the rationale discussed in the introduction regarding how my study addresses gaps in previous research (see 1.2) and the significant influence of the external variables on the use of L2 features (see 2.5.1 - 2.5.12), I coded the data pertaining to these variables. Regarding the internal variables, I assigned codes to them using the justifications outlined in Sections 4.4 and 5.4.

Table 4: The entire list of predictors together with their coding for regression analysis.

	Predictors	Predictor levels and measurement		Type
1	**First language background**	a. Bengali b. Hindi-Urdu c. Tagalog		Categorical
2	**Sex**	Male/ female		Categorical
3	**Amount of Arabic and L1 used**	at work	a. 0%	Categorical
		at home	b. 25%	Categorical
		socially	c. 50% d. 75% e. 100%	Categorical
4	**Motivation**	1-7		Continuous
5	**Attitudes toward HA, Madinah and its people**	1-7		Continuous
6	**Length of residence**	3-25		Continuous
7	**Informal exposure by**	Listening to the Quran		Categorical
		Watching Arabic TV		
		Listening to Arabic radio	Yes/no	
8	**Formal Arabic instruction in the migrants' country at school-age.**	7-10		Categorical
		11-14		
		15-18	Yes/no	
	Formal Arabic instruction in the migrants' country outside school-age.	7-10		
		11-14		
		15-18	Yes/no	
	Formal Arabic instruction in the migrants' country after leaving school	Yes/no		
9	**Formal Arabic tuition in KSA**	Yes/no		Categorical

10	**Migration identity & Future plans**	a. Return to their country b. Remain in KSA c. No plans	Categorical
11	**Age**	26-54	Continuous
12	**Educational level in L1**	a. Primary school b. High school c. University and above	Categorical
13	**Proficiency (Level of Arabic speaking)**	1-10	Continuous
14	**Language anxiety**	1-7	Continuous
15	**Lexical semantics**	a. Day/part of the day b. Formulaic utterances c. Miscellaneous d. Place/city/country e. Clothes	Categorical
16	**Linguistic contexts**	a. Numbers b. Two phrases c. Two clauses d. More than two phrases	Categorical

3.13 Data Analysis

I analysed my data in accordance with Labov's (1972: 94) Principle of Accountability which states:

> any variable form (a member of a set of alternative ways of "saying the same thing") should be reported with the proportion of cases in which the form did occur in the relevant environment [emphasis mine], compared to the total number of cases in which it might have occurred.

I identified all the possible variants of the variable that I am interested in. The principle of accountability would be violated if not all the possible variants were identified.

Relative frequency illustrates how dependent variables are impacted by a variety of independent variables; however, it frequently reflects these impacts in isolation (Guy 1993: 23). Therefore, instead of analysing the influence of each independent variable in isolation, the mainstream variationist practice is to use multivariate analysis which can illustrate the simultaneous effect of various environmental factors on a dependent variable (Guy 1993; Tagliamonte 2006). This kind of effect is known as the 'principle of multiple causes' and can be examined by using regression (Young & Bayley 1996: 253).

Mixed methods are used in my study. Qualitative analysis is employed to examine verbal and nominal agreement in the speech of migrant workers in Madinah. Quantitative methods are employed to discover whether the internal and external factors impact migrant workers' use of the definite article, the conjunction marker and /f/ in comparison to HA native speakers who use them categorically. This type of analysis is essential to variationist sociolinguistics (Tagliamonte 2006: 12). Distributional analysis is a basic descriptive statistic used in this thesis to establish trends in the data. Regression analysis is used to establish whether constraints are significant. The p-value was set at 0.05.

The statistical language R was used in my research to analyse the data statistically (R Core Team 2012). R has become the de facto standard tool particularly in corpus-based studies of language variation and change. The 'lme4' package was employed for mixed-effect logistic regression modelling (Bates *et al*. 2015). It enables the examination of linguistic patterns that are present in the entire group, as well as the simultaneous investigation of individual variations (Drager & Hay 2012: 4). In addition, mixed-effects models have the capability to concurrently account for multiple sources of variation, including cross-items and repeated measures obtained from the same informants (Baayen, Davidson & Bates, 2008; Barr *et al.* 2013). Mixed-effect models can also include individual speaker as a random effect which makes the estimations and p-values more accurate (Gorman & Johnson 2013). Moreover, the risk of outliers which may skew the results, is reduced when using mixed effect models because only statistically significant factors are selected 'when they are strong enough to rise above the inter-speaker variation' (Johnson 2009: 365). Individual speaker was included as random effect in the models of all three linguistic variables. The word in which /f/ occurs was also considered a random effect in the models of /f/.

For each morphosyntactic and phonologica variable, there are four models. The first model includes all the speakers in the corpus. The second model includes L1 Bengali speakers, the third model consists of L1 Hindi-Urdu speakers, whereas the last model represents L1 Tagalog speakers. I used stepwise regression to deal with the multicollinearity problem. The last version of the findings from the 'stepwise regression' test was used as the basis for selecting the independent variables for each model. An essential component of this process is the use of the Akaike Information Criterion (AIC) as a heuristic for guidance. The AIC, which weighs the complexity of a model against its goodness of fit, is a measure of statistics employed to facilitate selecting the model that is the most suitable for the data (Burnham & Anderson, 2002). It is the model with the lowest AIC that is considered to be the best.

I have included only orthogonal variables and could not add more independent variables due to multicollinearity and non-orthogonal variables. Increasing the number of independent variables included in any regression model can cause a singularity problem which means that the random effect will be zero. A further issue as regards including too many independent variables is that all the factors become insignificant. I have also ascertained that occasionally, a random effect omits a participant because there is missing data in a certain predictor. As this independent variable does not significantly influence the use of the dependent variable, I have chosen the best model by excluding the variable and including the participant. Furthermore, a number of references were incorrectly chosen (or unordered) by default and were corrected using the relevel() function in R. For instance, I set Tagalog as the reference in the /f/ model that includes all the speakers in the corpus, while Bengali was the reference in the definite article models and the coordinating conjunction marker which include all sampled migrant workers.

With certain predictors, some predictor levels are conflated for quantitative analysis. As my study includes only 30 subjects, breaking down the data into too many predictor levels or cells in the tables meant that some data subsets were based on only a handful of speakers and results would therefore not be representative of social groups but reflect the preferences of an individual. Put differently, the findings might not provide a meaningful or robust insight into the impact of the internal and external factors unless some predictor levels are combined.

Chapter 4: Verbal and Nominal Agreement

4.1 Introduction

The aim of this chapter is to determine whether the Arabic varieties spoken by migrant workers in Madinah are a pidgin language or an interlanguage. The chapter concentrates on the use of three HA morphological features: subject-verb agreement in gender, number and person; noun-adjective agreement in gender, number and definiteness; and number marking on nouns. These features are highly indicative of pidgin status and I have examined them qualitatively.

Within the wider framework of this research, an examination of morphological features will generate vital knowledge regarding the categorisation of the migrant workers' Arabic varieties. Although the present chapter is dedicated to morphological features, subsequent chapters will cover syntactic and phonological variables. Diverse classifications may be proposed based on the analyses of these variables. For instance, the analysis of the morphological features may indicate that the Arabic varieties of the migrant workers are a pidgin, while the analysis of the syntactic and/or phonological features may suggest that they are an interlanguage. This possible difference may indicate the existence of a continuum, with the higher end potentially approaching an interlanguage. This particular finding could be considered the contribution made by my study.

In the following paragraphs, I will present the previous GPA studies. If the descriptive analysis in my study exhibits similar results, we can predict that the Arabic varieties of the migrant workers would be better described as a pidgin. These results specifically include simplification in the use of HA morphological features, for example a lack of verbal and nominal agreement. However, if agreement marking follows the usage patterns of L1 HA speakers, their Arabic varieties might be an interlanguage.

Arabic is a morphologically rich language with a complex inflection system, which makes learning the language challenging for migrant workers (Hobrom 1996: 77). Simplifying the subject-verb agreement criteria is one strategy exploited by the migrants. This technique displays conventional pidgin structures. Concerning person, number or gender, the verb in GPA does not agree with the noun (Albaqawi 2020; Almoaily 2012; Alsusut 2022; Bakir 2010; Hobrom 1996). The third-person

singular masculine form of the verb, on the other hand, is the one that is most often employed with all subjects, probably because it is the least marked form (Hobrom 1996). In example (9a) below, a migrant worker in the study conducted by Almoaily (2012) employed the verb *yi-dris* 'study', which is the third person singular masculine present form in preference to the first person singular past form in GA *daras-t* 'studied', in example (9b). According to Bakir (2010: 205), '[no] changes are applied to this form to indicate a difference in tense, mood, aspect, voice or agreement in gender, number or person with the subject'.

(9a) ana yi-dris kulliah
I PRS.3SG.M-study college
'I studied in college'
(Almoaily 2012: 85)

(9b) ana daras-t fi el-kulliah
I study-1SG.PST in DEF-college

Moreover, in GAP morphological inflection is infrequently employed to differentiate grammatical numbers and genders with nouns (Alghamdi 2014: 125). In example (10), after mentioning *arbaah* 'four', a migrant worker used *lughah* 'language' in its singular form rather than in its plural form *luɣat* 'languages' in GA (Almoaily 2012). Alghamdi (2014: 120) indicated that affixation is typically inserted arbitrarily. For example, a speaker may include the correct affix in one instance but omit it in another. However, this could also be considered a form of variation. Additionally, migrant workers frequently used adjectives that did not agree with the gender and number of their corresponding nouns (Albaqawi 2020: 252; Almoaily 2012: 148; Alsusut 2022: 335). According to Almoaily (2012) and Hobrom (1996), all nouns are utilised in the singular masculine form. Example (11) in the study undertaken by Albaqawi (2020: 144), illustrates a migrant worker using the adjective *kabeer* 'big' in its masculine form as opposed to *kabeerah*, the feminine form in GA.

(10) Ana mawjood arbaah lughah
I exist four.F language.**SG**.F
'I speak four languages'
(Almoaily 2012: 88).

(11) yabɣi ʕa'lah kabeer
want family.SG.F big. SG.**M**
'I want to make a big family'
(Albaqawi 2020: 144).

In the current chapter, the HA morphological features are examined in the Arabic speech of three older migrant speakers: BM3, a 44-year-old male Bengali speaker; HM3, a 54-year-old male Hindi-Urdu speaker; and FF4, a 51-year-old female Tagalog speaker. Each speaker was selected because they are the oldest and probably the most experienced person in their language group, having lived in Hijaz for more than 15 years. This may assist us to discover a clear linguistic pattern in their L2. Owing to gender differences in language usage, a mix of male and female speakers was included. This may provide us with a comprehensive analysis and mitigate gender bias.

Section 4.2 reviews how verbal and nominal agreement are employed in HA, as well as in migrant workers' L1s: Bengali, Hindi-Urdu and Tagalog. This section confirms how these Indo-Aryan, Austronesian and Semitic contact languages differ from each other in their use of the morphological features. In addition, it will demonstrate how L1 might affect the participants' acquisition of HA features if their Arabic varieties are an interlanguage. The descriptive analysis of the migrant workers' use of HA verbal and nominal agreement patterns is presented in Section 4.3. It reveals whether each language speaker has their own Arabic variety or whether all three migrant workers share a simplified variety of the Arabic language. I will conclude that, in general, the three migrant workers apply a simplified variety of Arabic characterised by a lack of morphological agreement.

4.2 Verbal and Nominal Agreement across HA and L1s

4.2.1 Agreement in Hijazi Arabic

Verbal agreement

UHA and BHA have two types of verbs: sound or strong verbs and weak verbs (Alqahtani & Sanderson 2019; Il-Hazmy 1975). Strong verbs are separated into two distinctive sub-types: regular and irregular. The stem or root of a regular verb does not include /w/ and /y/, distinguishing it from weak verbs (Alqahtani & Sanderson 2019: 7). Additionally, it excludes double sounds or

hamza /ʕ/, which are characteristic of irregular verbs. This indicates that the root of a regular verb consists only of consonant sounds, for example (k-t-b) for 'to write,' without geminates or hamza.

In UHA and BHA, the verb agrees with the subject in gender, number and person (Il-Hazmy 1975: 172; Omar 1975: 14 &18). Agreement patterns vary across tenses. Tables 5 and 6 illustrate the various suffixes used to indicate inflection for gender, number and person in the perfect (past) tense in UHA and BHA. Similarly, different affixes are employed to indicate inflection for gender, number and person across various tenses: the imperfect (present) tense, future tense and imperative. There are specific differences in verbal inflectional morphemes between the two varieties (for further details, see Il-Hazmy 1975; Omar 1975). Despite these differences, the key point is that verbs in both varieties inflect for gender, number and person, which is the focus of my analysis. It does not matter whether the participant uses the UHA affixes or BHA affixes. To address my research questions, the most important concern is whether verbal inflectional morphemes are used correctly.

Table 5: Subject-verb agreement in the past tense in UHA (adapted from Omar 1975: 13).

Subject	**Verb form** */daras/* 'studied'	**Suffix**
1SG	*/darast/*	-t
1PL	*/darasna:/*	-na:
2SG.M	*/darast/*	-t
2SG.F	*/darasti:/*	-ti:
2PL	*/darastu:*	-tu:
3SG.M	*/daras/*	∅
3SG.F	*/darasat/*	-at
3PL	*/darasu:/*	-u:

Table 6: Subject-verb agreement in the past tense in BHA (adapted from Il-Hazmy 1975: 82).

Subject	**Verb form** /rikib/ 'to ride'	**Suffix**
1SG	*/rikibt/*	-t
1PL	*/rikibna:/*	-na:
2SG.M	*/rikibt/*	-t
2SG.F	*/rikibti:/*	-ti:
2PL.M	*/rikibtu:/* */rikibtum/*	-tu:/-tum
2PL.F	*/rikibtin/*	-tin
3SG.M	*/rikib/*	∅
3SG.F	*/rikbat/*	-at
3PL.M	*/rikbaw/-* */rikbam/*	-aw/-am
3PL.F	*/rikban/*	-an

Nominal Agreement

Adjectives agree with the noun in gender, number and definiteness in UHA and BHA (Il-Hazmy 1975: 176; Omar 1975: 46). See the following three examples[10].

(12) ʃuft ri**dʒdʒ**a:l kab**i:**r
See.1SG.PST man.SG.M old.ADJ.SG.M
'I saw an old man'

[10] As a native speaker of HA, I have written these examples.

(13) ʃuft ri**dʒ**a:l kub**a**:r
See.1SG.PST men.PL.M old.ADJ.PL.M
'I saw old men'

(14) ʃuft ħorma kab**i:**r-**a**
See.1SG.PST woman.SG.F old.ADJ.SG-F
'I saw an old woman'

HA nouns inflect for gender, number, definiteness and possessive markers (Il-Hazmy 1975; Omar 1975: Sieny 1972). HA nouns consist of three 'number' forms: singular, dual and plural. The singular form of the noun is typically unmarked, the stem form. Dual nouns are indicated by the suffix /e:n/, for example, *madrasa* 'a school' becomes *madrasate:n* 'two schools'. In HA, there are three types of plurals: the regular masculine plural takes /i:n/ at the end of most male singular nouns, for instance *muhandis 'an engineer' becomes muhandis**i:n*** 'engineers'. In the regular feminine plural, most female singular nouns or inanimate singular nouns are marked with the suffix /a:t/, for example *bint* 'a girl' becomes *ban**a:t*** 'girls' and *imtiħa:n 'an exam'* becomes *imtiħa:n**a:t*** 'exams'. Likewise, a broken plural is formed by altering the vowel patterns among the root consonants, for instance, *ba:b* 'door' and the plural form is *ʔabwa:b* 'doors'.

With cardinal numbers, a plural inflectional morpheme is only used when the quantified noun falls within the range of three to ten, as illustrated in (15) and (16) (Il-Hazmy 1975; Omar 1975). When the number exceeds ten, the singular form of a noun is used, as in (17) and (18).

(15) Lijja hina: talaata ajja:m
To me here three day.**PL**
'I have been here for three days'.
(Omar 1975: 70).

(16) θala:θ sni:n
Three year.**PL**
'Three years'
(Il-Hazmy 1975: 138).

(17) Talatmijjat dula:r
Three hundred dollar.**S**
'Three hundred dollars'
(Omar 1975: 69)

(18) Mite:n xuru:f
Two hundred lamp.**S**
'Two hundred lamps'.
(Il-Hazmy 1975: 134)

4.2.2 Agreement in Bengali

Verbal agreement

The Bengali morphological system is rich in affixes. It includes over fifty different affixes to mark person, aspect, mood and tense (see Ray *et al.* 1966). It differs from Arabic in that verbs do not inflect for gender and number. Therefore, suffixes are similar for singular and plural subjects (David 2015: 182). In contrast, verbs in Bengali agree with the subjects in person and formality level (David 2015: 181; Klaiman 2009: 426).

Nominal Agreement

Klaiman (2009: 425), suggests that whereas morphology in Bengali is productive for verbs and minimal for nouns, it is non-existent for adjectives. There is an absence of adjectival morphology in this language in gender, case and number, and the loss of grammatical gender in Bengali makes it different from other south Asian languages (Chandra 2016; David 2015). Adjectives precede the nouns that they modify, but do not inflectionally agree with them (Klaiman 2009).

Bengali nouns do not exhibit gender differences (David 2015). Markers in Bengali signify case, number and humanness. Classifiers are bound morphemes. The plural marker for humans is /-(e)ra/. For instance, the singular form of the word 'girl' is *meye* and its plural form is *meyera* 'girls'. The plural marker for non-humans is /-gulo/ or /-guli/. For example, the singular form of the word 'river' is *nodi* and its plural form is *nodigulo* 'rivers'. Nouns do not require plural markers when they are preceded by numerals, as example (19) demonstrates.

(19) Tebile carṭe potrika royeche
Table four magazine.S be
'There are four magazines on the table'
(David 2015:77)

4.2.3 Agreement in Hindi-Urdu

Verbal Agreement

Generally speaking, the verb in Hindi must agree with a noun phrase that does not have a postposition immediately after it (Agnihorti 2013: 11). If a postposition does not follow the subject, the verb will agree with it in gender, number and person (Agnihorti 2013: 11), see example (20a). The subject is Ram which is third person singular masculine, so the verb is marked with -taa. In example (20b), Sita is the subject and it is third person singular feminine, so the verb is marked with -tii. Moreover, verbs inflect for tense, aspect and mood. Alternatively, when the subject is followed by a postposition, the verb does not agree with it but with the object. This is because Hindi is a verb-final language. In example (21), the verb does not agree with the subject but with the feminine object which is *rotii* 'bread'.

(20a) Raam seb khaa**taa** hai
Ram.3SG.M apple eat.3SG.M.
'Ram eats an apple'
(Agnihorti 2013: 10).

(20b) Siitaa roz nahaa**tii** hai
Sita.3SG.F everyday bath.3SG.F
'Sita bathes everyday'
(Agnihorti 2013: 11).

(21) Raam ne rotii khaaii
Ram.3SG.M bread.OBJ.F ate.F
'Ram ate bread'
(Agnihorti 2013: 12).

To form Urdu finite verbs, suffixes are added to the root of the verb to mark gender, number, tense, aspect and mood (Schmidt 1999). While Urdu intransitive verbs agree with the subject in number and gender, they do not agree with it in person. On the other hand, transitive verbs only form agreement with the subject in the present or imperfective tense when the subject is in the nominative case, regarding gender, number and person. In the perfective tenses - past tenses, the case of the subject is ergative. In this situation, there might be an agreement between the verb and the object when its case is absolutive.

Nominal Agreement

In Hindi, there are two categories of adjectives: variable and invariant (Agnihorti 2013: 98). Variable adjectives end with the long vowel /a:/ and agree with the noun in gender, number and case. In examples (22a) and (22b), the variation of *lamba:* 'tall' depends on the following noun if it is *larka:* or *larki:* 'boy' or 'girl' respectively. Invariant adjectives end with other vowels or consonants and do not change whatever the following noun's gender, number and case. Examples (23a) and (23b) show that the adjective *la:l* 'red' does not change either with *kapra:* which is masculine and *saari:* which is feminine. Moreover, the adjective precedes the noun.

(22a) lamb**a:** lark**a:**
tall.ADJ.M boy
'tall boy'
(Agnihorti 2013: 98).

(22b) lamb**i:** lark**i:**
tall.ADJ.F girl
'tall girl'
(Agnihorti 2013: 98).

(23a) la:l kapra:
red.ADJ cloth.M
'red cloth'
(Agnihorti 2013: 99).

(23b) la:l sa:ri:
red.ADJ sari.F
'red sari'
(Agnihorti 2013: 99).

Schmidt (1999: 34) reports that in Urdu, an adjective agrees with a noun in number, gender and case. The suffix -a which is used with the male singular is changed to the suffix -i with the female singular, see examples (24a) and (24b) below. Moreover, the adjective precedes the noun.

(24a) gahr**a** pani
deep.ADJ.**M** water
'deep water'
(Schmidt 1999: 34)

(24b) gahr**i** nimd
deep.ADJ.**F** sleep
'deep sleep'
(Schmidt 1999: 34)

In Hindi-Urdu, nouns have gender, number and case inflections (Koul 2008: 33; Schmidt 1999: 5). With a masculine singular noun: if it ends with /-a:/ such as in /larka:/ 'boy', the end of the

form will change to /-e:/ like in /larke:/ ‘boys’. Similarly, if it ends with a consonant or any other vowel then the form does not change at all, for example /kot/ ‘coat(s)’ and /ghar/ ‘house(s)’. Concerning a feminine singular noun: if it ends with /-i:/ such as in /larki:/ ‘girl’, /-i:/ will be omitted and /-iya:n / will be added to the stem of the word like in /larkiya:n / ‘girls’. In addition, if it ends with a consonant, the suffix /-e:n/ will be added like /kita:b/ ‘book’, which subsequently becomes /kita:be:n/ ‘books’. In Hindi, when a cardinal number is used before a collective noun, the noun frequently remains singular, as in example (25) (McGregor 1976: 63). In Urdu, cardinal numbers precede the noun that they describe. They are regarded as masculine, but are not inflected (Schmidt 1999: 228).

(25) Do payata cay
Two cup.**S** tea
‘Two **cups** of tea’
(McGregor 1976: 63)

4.2.4 Agreement in Tagalog

Verbal Agreement

In Tagalog, verbs do not inflect for subject, person and number (Serrat *et al.* 2007: 204). There is no obligatory distinction between verb forms that occur with singular and plural, as demonstrated in examples (26a) and (26b) (Schachter & Otanes 1983: 335). Moreover, they do not change due to formality or gender (Barrios 2011: 2). It is also worth mentioning that although verbs do not mark tense, they mark aspect (Serrat *et al.* 2007: 205).

(26a) Kumakanta siya
Sing he.3SG
‘He sings’
(Schachter & Otanes 1983: 335).

(26b) Kumakanta sila
Sing they.3PL
'They sing'
(Schachter & Otanes 1983: 335).

Nominal Agreement

In Tagalog, adjectives are inflected for number (Schachter & Otanes 1983: 229). To create the plural form, the particle *mga* is ordinarily inserted before the word it pluralises. For example, *tamad* is the singular adjective form which means 'lazy' and its plural is *mga tamad* (Serrat *et al.* 2007: 203). Duplicating the root word's first syllable occurs for most of the adjectives that start with *ma-* (Schachter & Otanes 1983: 230). For instance, *mabagal* is the singular adjective form which means 'slow', while the plural form of this adjective is *mababagal*. However, in general, the pluralisation of explicit adjectives is optional. To clarify, a pluralised adjective, as seen in example (27a), or a non-pluralised adjective, as in example (27b), is acceptable (Schachter & Otanes 1983: 229).

(27a) **Mga** tamad sila
PL particle lazy.ADJ.PL they
'They are lazy'
(Schachter & Otanes 1983: 229).

(27b) Tamad sila
Lazy.ADJ.SG they
'They are lazy'
(Schachter & Otanes 1983: 229).

There is a small group of adjectives that are borrowed from Spanish which inflect for gender (Schachter & Otanes 1983: 197). Example (28a) below demonstrates that *-a* is used at the end of an adjective to indicate female human beings or animals. Alternatively, *-o* is used at the end of the adjective to indicate male human beings or animals, abstractions, inanimate objects and groups that include people of both sexes, as shown in example (28b) (Schachter & Otanes 1983: 197).

(28a) Komik-**a** si Linda
Funny-**F** particle Linda
'Linda is funny'
(Schachter & Otanes 1983: 197).

(28b) Komik-**o** si Fred
Funny-**M** particle Fred
'Fred is funny'
(Schachter & Otanes 1983: 197).

Tagalog nouns are not inflected for case and number (Schachter & Reid 2009: 852). Nevertheless, specific particles or affixes are used to denote plurality, possession, or other grammatical relationships. There are gender markers for specific words that are derived from Spanish, such as *amiga* 'a female friend' or *amigo* 'a male friend'. The plural form of a noun is achieved by adding *mga* to the beginning of a noun such as in example (29a). However, using *mga* is optional, as in example (29b) (Schachter & Reid 2009: 852).

(29a) **Mga** **librong** para sa bata
PL.particle book for particle child
'books for the child/children'
(Schachter & Otanes 1983: 111).

(29b) **librong** para sa mga bata
a book for particle PL.particle child
'a book/ books for the children'
(Schachter & Otanes 1983: 111).

In the following example (30), *Bumbero* means 'a fireman' in the literal sense (Barrios 2011: 8), whereas the plural should be *mga bumbero*. Filipinos, on the other hand, only use the single and not the plural in everyday speech.

(30) Sunog! Sunog! Tawagin niyo ang bumbero!
Fire! Fire! Call you the fireman!
'Fire! Fire! Call the firemen!
(Barrios 2011: 8).

Mga which means 'approximately', 'about' or 'around,' is not 'plural' when it is employed with a time expression number or a cardinal (Schachter & Otanes 1983: 111). See example (31).

(31) Mga sampung anak
about ten child
'about ten children'
(Schachter & Otanes 1983: 111)

4.2.5 Summary

To summarise, Table 7 shows that the subject-verb agreement is marked in the superstrate and three substrate languages in different ways. In HA, the subject agrees with the noun in gender, number and person. This is also the case with Hindi, though only when the subject is not followed by a postposition. The subject also agrees with the noun in gender, number and person in Urdu when the verb is transitive. However, with intransitive verbs, the agreement pertains to number and gender. Whilst the verbal agreement in Bengali relates to person and formality level, it refers to aspect in Tagalog. If we ascertain that the L1 Hindi-Urdu speaker uses the subject-verb agreement in Arabic more than the L1 Bengali speaker and the L1 Tagalog speaker, we can predict that their Arabic varieties are an interlanguage.

With regard to the nominal agreement, the adjective agrees with the noun in gender, number and definiteness in HA. Bengali does not have noun-adjective agreement; the other two substrate languages have it. The adjective agrees with the noun in gender, number and case, in Hindi-Urdu, although in Hindi, it only agrees with variable adjectives. In Tagalog, the nominal agreement refers to number only. Adjectives borrowed from Spanish agree with the noun in gender. If we determine that the L1 Bengal speaker employs the noun-adjective agreement in Arabic less than the L1 Hindi-

Urdu speaker and the L1 Tagalog speaker, we can expect that their Arabic varieties are an interlanguage.

As the plural form of nouns occurs in all migrant workers' L1, it is expected that they can be employed in HA. Nonetheless, the dual form might be less used by them owing to its absence in their L1s. Regarding cardinal numbers 3-10, the plural form of nouns may well be less used by the L1 Bengali and Hindi[11] speakers because they typically use the singular form in their L1. With respect to the L1 Tagalog speaker, their use of inflections might be difficult to predict, because, as far as I am aware, there is no literature available as regards this particular subject.

Table 7: A cross-linguistic comparison of verbal and nominal agreement in the superstrate and three substrate languages.

L1	**Types of agreement**		
	Verbal agreement	**Nominal agreement**	
		N-ADJ agreement	**Number marking on N**
HA	Yes, for gender, number & person	Yes, for gender, number & definiteness	• SG, DU &PL • With number 3-10, the PL form is used. • With number 11 and above, the SG form is used
Bengali	Not for gender and number but for person and formality level.	No	• SG &PL • Often singular form after numbers

[11] Although in my study I have generally examined Hindi-Urdu speakers, HH3 speaks Hindi as his L1, in this qualitative analysis, I have not included Urdu speakers.

Hindi-Urdu	**In Hindi:** the verb agrees with the subject in gender, number and person if it is not followed by a postposition. **In Urdu:** 1. intransitive verbs agree with the subject in number and gender. 2. transitive verbs agree with the subject in number, gender and person.	**In Hindi:** 1. variable adjective agrees with the noun in gender, number and case. 2. invariant adjective does not change. **In Urdu:** an adjective agrees with a noun in gender, number and case	• SG &PL • **In Hindi**, the singular form is frequently used after numbers. • No data is available regarding the form of the nouns preceded by numbers **in Urdu.**
Tagalog	No, only for aspect	1. Yes, for number. However, the pluralisation of explicit adjectives' is optional. 2. Borrowing an adjective from Spanish inflects for gender	• No inflection for number, although *mga* can be applied optionally. • No data is available regarding the form of the nouns preceded by numbers.

4.3 Descriptive Analysis of Agreement

In this section, I will provide a descriptive analysis of how each of the three migrant workers apply the verbal and nominal agreement. The first sub-section focuses on the L1 Bengali speaker; sub-

sections 4.3.2 and 4.3.3 present results for the L1 Hindi-Urdu and Tagalog speakers, respectively. Each sub-section includes an excerpt of roughly 200 words taken from the participant's speech, how this speech would sound in HA and the translation into English. This excerpt is a section in which a migrant worker spoke at length without questions and was randomly chosen to mitigate research bias. The excerpts from the three speakers were chosen in such a way that they are comparable, specifically regarding conversational topics such as their life and family, in particular. Short data extracts from three speakers, rather than the entire interviews, were analysed due to time limitations. In each sub-section, selected instances exemplify a more general pattern for the speaker. I will begin each sub-section by analysing the verbal agreement which appears to be very uncommon in the speech employed by the three speakers. I will also describe the strategies that the migrant workers use as opposed to using the HA-inflected verb. I will then examine the nominal agreement, beginning with the noun-adjective agreement followed by number marking on nouns

4.3.1 BM3, a L1 Bengali Speaker

I. Excerpt from the Bengali participant's speech.

هدا أول أنا بابا في موت. أخُ تلاته كله مدرسه درس. هدا هكومه شغل بابا. هو في دكا. دكا مدينه كبير. إند المكتب في مسزد سلِّي إشا . ممكن اتنين ركات بادين هو تاه مسزد. بادين في جما قريب هقي. هو بابا سوا سوا شغل. هو يزي هو زيب خبر كدا. أنا أخو أختي روه دكا مستسفى. واهد أسبو كدا موت. أنا بنقلاديش. ممكن ستاشر سنه كدا زياده بابا موت.

في هقي أخويا. أخويا ياني هقي أختي رجال، هو ساودي أول. هو تالتو تاشيره. هو في رياد هو تلّا تاشيره. هو شغل بيزنيس. إن شالله أنا مايبقَ روه، مايبقَ روه مكان تاني. يزي مدينه ازليس مدينه. أبقَ روه، بنقلاديش بس. اهمدولله ميه ميه مدينه مكه هدا أحسن من كله دنيا. في كعبه في رسول سلى الله عليه وسلم هدا أحسن من كله دنيا مكه مدينه. والله أنا في بنقلاديش أنا سلي الله. روه ساوديا، انا أبقَ روه مدينه ولا مكه. اهمدولله انا يزي مدينه. أنا كل اسبوع سلي كلم الله. روه بره ساوديه مدينه ولا مكه. اهمدولله إزي مدينه. سوي هز. الهمدولله هز خلاس قبل سنه.

II. A Transcription of How a Native Speaker of HA would Speak.

لما أبويا مات، أخواني التلاته كانوا يدرسوا في المدرسة. أبويا كان يتشغل في وظيفة حكومية. هو كان في دكا. دكا مدينة كبيرة. جنب المكتب حقه في مسجد، صلى فيه العشاء. يمكن صلى ركعتين وطاح في المسجد. في جماعتنا في دكا، كانوا قريبين منه. هما يشتغلوا مع أبويا. هما جو وقالوا لنا. أخويا وأختي راحوا المستشفى في دكا. كان أبويا ما يقدر يتكلم. تقريبا أسبوع كدا ومات. أنا كنت في بنقلاديش. يمكن صار سطعشر سنه أو أكتر على موته.

في أخويا، أخويا، أقصد زوج اختي. أشتغَل في السعودية أول. هو طلّع لي التأشيرة. لما كان في الرياض، طلّع لي التأشيرة. هو يشتغل في التجارة. إن شاء الله أنا ما أبغ أنقل. ما أبغ أروح مكان تاني. جيت المدينة، أجلس في المدينة. إذا أبغ أغير، أروح بنقلاديش بس. الحمد لله المدينة ومكة ميه الميه أحسن من أي مكان في الدنيا. فيهم الكعبة وفيهم الرسول صلى الله عليه وسلم مدفون. مكة والمدينة أحسن من أي مكان في الدنيا. والله أنا لما كنت في بنقلاديش دعيت ربنا إني أروح السعودية. أنا أبغ أروح المدينة أو مكة. الحمد لله جيت المدينة. أنا كل أسبوع كنت أصلي وأدعي ربنا إني أسافر السعودية أروح المدينة أو مكة. الحمد لله جيت المدينة. حجيت. الحمد لله حجيت قبل سنه.

III. English Translation

When my father died, my three brothers were studying in school. My father was working in a government position. He lived in Dakka, which is a big city. Next to his office, there is a mosque, where he frequently prayed. He was praying Isha in the mosque during Ramadan. He had prayed two rak'ahs possibly when he suddenly collapsed. Our relatives who live in Dhaka, were very close to him and worked with my father. They came and told us that he was ill. My brother and sister went to the hospital in Dhaka but found that my father couldn't speak. About a week later when I was in Bangladesh, he died. It must be about sixteen years or more since his death.

My brother, I mean my brother-in-law. He worked in Saudi Arabia in the past and applied for my visa when he lived in Riyadh. He works in business. I don't want to move, God willing. I don't want to go somewhere else. I came to Madinah, I want to stay in Madinah. If I want a change, I will just go back to Bangladesh. Praise be to God, Al-Madinah and Makkah are better than any place in the world. They have the Kaaba and the Messenger, peace be upon him, is buried here. I swear, when I was in Bangladesh, I prayed to Allah that I would like to go to Saudi Arabia. I want to go to Madinah or Makkah. Thank God, I came to Madinah. Every week I prayed to Allah that I would go to Saudi Arabia, to Madinah or Makkah. Thank God, I came to Madinah. I performed Hajj about a year ago, praise be to God.

(A) Verbal Agreement

I- Using the HA Third Person Singular Masculine Present Form Instead of Other Subjects

In example (32a), BM3 makes use of the third-person singular masculine present form as opposed to the first singular past tense form. Example (32b) demonstrates how the sentence would appear in HA.

(32a) ʔana: **ji-zi:** Madinah
I **PRS.3SG**.M-come to Madinah
'I came to Madinah'

(32b) ʔana: **ʒi:-t** ʔal-Madinah
I come to-**1SG.PST** DEF-Madinah

In example (33a), BM3 was discussing his own plans for the future, but he used 3SG marking. In HA, the 1SG form of the verb is employed. The HA version of this example would be as shown in (33b).

(33a) ʔana: ma: **ji-bga** ru:h
I don't **PRS.3SG**.M-want to go.2SG.M.IMP
'I don't want to move'

(33b) ʔana: ma: **ʔa-bʏa** ʔa-ngul
I don't **PRS.1SG**.M-want to 1SG.M.PRS-move

BM3 utilised the third-person singular masculine present form of the verb, indicating wrong agreement with the subject in (33a). However, this particular migrant worker used the verb *ʔa-bʏa* 'want' correctly with the HA-inflected form in other examples, for instance the following in (34a):

(34a) ʔana: **ʔa-bga** ru:h Madinah walla Makkah
I **PRS.1SG**.M-want to go.2SG.M.IMP Madinah or Makkah
'I want to go to Madinah or Makkah'.

The HA version would be that in (34b):

(34b) ʔana: **ʔa-bʏa** ʔa-ru:ħ ʔal-Madinah aw Makkah
I **PRS.1SG**.M-want to 1SG.M.PRS-go DEF-Madinah or Makkah

The choice of inflectional morphology in non-HA-like example (33a) and in HA-like example (34a), could be attributed to the type of sentence. Regarding the negative sentence, BM3 did not use the verb like native speakers of HA would. However, with the affirmative sentence, he applied it in HA-like fashion.

II- Using the HA Third Person Singular Masculine Past Form more than other Subjects

In example (35a), BM3 mentions his three brothers who were all at school when their father passed away. He utilised the verb *daras* 'study', which is inflected for 3SG in the past tense. This appears to be in agreement with the subject *ʔaxu* 'brother' in some way[12] because the subject is also inflected for the singular form. Conversely, in HA, the verb *daras* 'study' would be inflected for 3PL, as explained in example (35b). Additionally, the subject *ʔaxu* 'brother' should also be inflected for the plural form *ʔaxwa:n* 'brothers' because BM3 mentions that he has three brothers. Likewise, in HA, the noun is used in the plural form when denoting the number 'three'.

(35a) ʔaxu talata kullu madrasa **daras**
Brother.SG three all school study.**3SG.M.PST**
'All my three brothers were studying at school'

(35b) ʔax-wa:n-i: ʔat-talata kullahum kanu: **ji-**drus-**u**:
Brother-PL-POSS DEF- three all COP **M.PRS**-study-**3PL**

fi: ʔa-madrasa
at DEF-school

There are some instances in the chosen excerpt where the subject agrees with the past verb. However, there have been some comments regarding this.

[12] In HA, we cannot say *ʔaxu daras.* The subject *ʔaxu* 'brother' should be combined with a possessive pronoun, such as *ʔaxu:ja:* 'my brother' *or ʔaxu:ha:* 'her brother' to agree with the verb *daras* 'study'.

In example (36a), BM3 used the overt pronoun *hwwa* 'he'. This is grammatically correct in HA, However, based on the context of the previous sentences in the excerpt above and to avoid repetition, the sentence in HA would comprise a covert subject, as observed in the (36b):

(36a) Mumkin	ʔitne:n	rakaat	bade:n	**hwwa**	**ta:h**		maszi:d
Possibly	two	rak'ahs	after that. ADV	he	fall.**3SG.M.PST**		mosque

'He had prayed two rak'ahs possibly when he suddenly collapsed'.

(36b) Jimkin	sˤalla:	rakʕa-te:n	baʕde:n	**tˤa:ħ**
Maybe	pray.3SG.M.PST	rak'ah-DU	after that. ADV	**fall.3SG.M.PST**

fi: ʔal-masʒid
in DEF-mosque

In example (37a), the subject agrees with the verb like in HA. However, *li:* 'for me' after the verb is missing. In contrast, in another example (38a), BM3 used the same verb with an incorrect suffix, which could be a performance error. Concerning these two previous examples, HA versions are demonstrated in (37b) and (38b).

(37a) Hwwa	fi:	Riyad,	hwwa	**talla**		ta:ʃi:ra
He	COP	Riyadh,	he	apply-**3SG.M.PST**		visa

'When he was in Riyadh, he applied for my visa'.

(37b) Lamma	ka:n	fi:	ʔr-riyadh	**tˤallaʕ**	li:	ʔat-ta:ʃi:ra
When	was	in	Riyadh	apply-**3SG.M.PST**	for me	DEF-visa

(38a) Hwwa	**talla-tu:**	ta:ʃi:ra
He	apply-**2PL.PST**	visa

'He applied for my visa'.

(38b) Hwwa **tˁallaʕ** li: ʔat-ta:ʃi:ra
He apply-**3SG.M.PST** for me DEF-visa

III- Using the HA Singular Masculine Imperative Form of the Verb as Opposed to the Inflected Verb

In example (39a), BM3 talks about his father and uses the 2SG imperative form of the verb instead of 3SG.M in the past tense. Moreover, in example (40a), he talks about himself and makes use of the 2SG imperative form of the verb too in preference to 1SG in the present tense. HA versions are demonstrated in examples (39b) and (40b).

(39a) **Salli:** ʔisha Ramadan
Pray.**2SG.M.IMP** Isha Ramadan
'He prayed Isha in Ramadan'

(39b) Huwwa **salla:** ʔa-ʔisha in Ramadan
He Pray.**3SG.M.PST** DEF- Isha in Ramadan

(40a) Ma: ji-bga: **ru:h** maka:n tani:
Don't PRS.3SG.M-want to go.**2SG.M.IMP** somewhere else
'I don't want to go somewhere else'

(40b) Ma: ʔa-bʁa: **ʔa-ru:ħ** maka:n tani:
Don't PRS.1SG.M-want to **1SG.M.PRS**-go. somewhere else

IV- Replacing the Verb with a Noun

In example (41a), BM3 has replaced the HA inflected verb *ji-ʃtaʁil* 'PRS. 3SG.M-work' with the noun *ʃuʁul* 'work'. Example (41b) demonstrates how the sentence would appear in HA.

(41a) Huwwa **ʃuʁul** business
He work trade
'He works in business'

(41b) Huwwa **ji-ʃtaɤil** fi: ʔa-tiʒa:ra
He **PRS. 3SG.M**-work in DEF. business

V- Verb Omission

Verb deletion is detected in BM3's speech. According to Almoaily (2012), in certain cases, when the meaning can be deduced from the context, the verb is omitted in GPA. On account of the absence of the verbal form, it is impossible to explain the subject-verb agreement in situations such as example (42a). In this example, the omitted verb is *ʔaʃtaɤal* 'work-3SG.M.PST' in HA. Likewise, in example (43a), the omitted verb is *sˤalla:* 'pray.3SG.M.PST'. These two examples would be in HA as observed in examples (42b) and (43b).

(42a) Huwwa Ø Saudi awwal
He Ø Saudi ADV
'He **worked** in Saudi Arabia in the past'.

(42b) Huwwa **ʔa-ʃtaɤal** fi: Saudi awwal
He work-3SG.M.PST in Saudi in the past.ADV

(43a) Mumkin Ø ʔitni:n rakaat bade:n hwwa ta:h maszi:d
Possibly Ø two rak'ahs after that. ADV he fall.3SG.M.PST mosque
'He **had prayed** two rak'ahs possibly when he suddenly collapsed'.

(43b) Jimkin **sˤalla:** rakʕa-te:n baʕde:n tˤa:ħ
Possibly pray.3SG.M.PST raka'ah-DU after that. ADV fall.3SG.M.PST

fi: ʔal-masʒid
in DEF-mosque

VI- Compound Verb

In their research, Naess (2008: 94) and Bakir (2010: 220) inferred that migrant workers used the verb sawwi 'to do/make' to create a compound verb from a noun, an adjective or another verb. In

example (44a), BM3 utilised a compound verb *sawwi* 'to do' + a noun as opposed to using the HA inflected verb, as illustrated in example (44b).

(44a) **Sawwi:** haz
Do Hajj.**N**
'I performed Hajj'

(44b) ʔana: **ħaʒʒi:-t**
I perform Hajj-1SG.PST

(B) Nominal Agreement

I- Noun-Adjective Agreement

Rather than using the feminine adjective *kabi:r-a* 'big' to agree with the feminine noun *madina* 'city', BM3 applied a masculine adjective in example (45a). In HA, it would be as shown in example (45b).

(45a) Dakka madina kabi:r
Dakka city.SG.F big.SG.**M**
'Dakka, which is a big city

(45b) Dakka madina kabi:r-a
Dakka.F city.SG.F big.SG-**F**

II- Number marking on nouns

In example (46a) and (47a), BM3 employed the singular form of the nouns instead of using the dual and plural forms, respectively. Examples (46b) and (47b) explain how the nouns would appear in HA.

(46a) Mumkin ʔitne:n rakaa
Possibly two rak'ah.**SG**
'Possibly two raka'as'

(46b) Jimkin rakʕa-te:n
Possibly rak'ah-**DU**

(47a) ʔaxu talata
Brother.**SG** three
'My three brothers'

(47b) ʔax-wa:n-i: ʔat-talata
Brother-**PL**-POSS DEF-three

4.3.2 HM3, a L1 Hindi-Urdu Speaker

I. Excerpt from the Hindi-Urdu participant's speech.

أنا يروه إجازه إن شالله سهر أسره انقليزي. سهر أسره إن شالله أنا روه اجازه. بادين هدا تياره ما أفته. تياره أفته ان شالله أنا سهر أسره روه. أنا في. أنا هرمه. هو في أكت. أكتك في ولد بنت سوي جواز سهر أسره. أنا روهْ إن شالله. كورونا كله كلاس، إن شالله تيب الهمدلله. ممكن روه أربا سهر. أربا شهر، كمسه. أنا كلَّم سته سهر ركِّب إجازه. أنا إجي أربا شهر قبل، مافي مسكله. ممكن مافي مالوم متى هدا قفل هدا تياره مشكله هدا. بادين هدا روه دبي، روه كتر، إجي فلوس جياده قالي. بادين كفيل مافي جيب فلوس كدا.

جيفور قريب. ممكن تلاته سائه أنا ألى تول بيت. ديلي، ديلي أسره سائه، أسره سائه بادين روه بيت. هدا جيفور هدا كريب. هدا تلاته سائه، الى تول روه بيت. هنا روه مدينه دمام ولا دبي ألى تول بيت. روه دبي هند واهد سائه نس كدا. زي كدا. أهسن هدا أنا جيفور هدا قريب هدا كويس. إيوا ديلي بييد. بادين أنا كلَّم ولد إنت تال. هو تالْ في مسكله بادين. هدا جيفور هدا. هدا كريب. أنا كلام إنت سقير سياره سيلْ ركِّب أقراد روه بيت ألى تول.

II. A Transcription of How a Native Speaker of HA would Speak.

أنا أبغ أروح إجازة إن شاء الله في شهر عشره بالميلادي. شهر عشره إن شاء الله أبغ أروح إجازة. الطيران دحين مقفل. لما الطيران يفتح، إن شاء الله انا شهر عشره أسافر. زوجتي عندها أخت. أختها عندها ولد وبنت حيتزوجوا في شهر عشره. أنا ابغ أروح إن شاء الله. كورونا كمان تختفي. إن شاء الله طيب الحمد لله. يمكن أروح أربعه شهور أو خمسه. أبغ أقول له سته شهور أبغ أخد إجازة. أنا رحت قبل كدا أربعه شهور. ما كان في مشكله. ما أعرف متى. تقفيل الطيران هذا مشكله. نروح على دبي أو قطر. ويكون أغلى. الكفيل ما راح يعطيني فلوس قد كدا.

جيفور قريبه. يمكن تلاته ساعات وأوصل البيت. ديلي ديلي عشره ساعات ألين أوصل البيت. هذه جيفور قريبة. تلاته ساعات وعلى طول أوصل البيت. من المدينة عالدمام أو دبي وعلى طول عالبيت. من دبي للهند ساعه ونص تقريبا. شي زي كدا. أحسن جيفور. قريبه وكويسه. إيوا هذه ديلي بعيده. أنا أقول لولدي يجيني. إذا يجيني على ديللي، حتكون مشكله. هذه جيفور قريبه. انا أقول له يجيب سيارته الصغيرة. يأخذ العفش ونروح البيت على طول.

III. English Translation

I would like to go on vacation. God willing, in the tenth month. In the tenth month, God willing, I will go on vacation. There are currently no flights. When the flights begin, I will travel in the tenth month, God willing. My wife has a sister whose son and daughter will both get married in the tenth month. I want to go, God willing. I hope that the virus will also disappear. God willing, it will be fine, praise be to God. I might go for four or five months. I want to tell the employer that I would like to go on vacation for six months. I went for four months previously and it wasn't a problem, but, I don't know when. This flight stopping is a problem, as we have to go to Dubai or Qatar, which is more expensive. The employer will not give me money like this.

Jaipur is close and I'll possibly arrive home in three hours. From Delhi, it's around ten hours to get home. Jaipur is close. From Madinah to Dammam or Dubai and then direct to my home. From Dubai to India it is about an hour and a half. Something like that. Jaipur is better as it is close. Yes, Delhi is far away. I ask my son to pick me up from the airport. If he comes to Delhi, it will be a problem. However, Jaipur is close, so I tell him to come in his little car. We can put our luggage in the car and go home immediately.

(A) Verbal Agreement

I- Using the HA Third Person Singular Masculine Present Form Instead of Other Subjects

Regarding the 1SG pronoun *ʔana:*, HM3 used the 3SG form of the present verb in place of the 1SG form in example (48a). The HA form of this specific sentence would be observed in example (48b).

(48a) ʔana: **ji-ru:h** ʔiʒaza
I **PRS.3SG.M**-go vacation
'I would like to go on vacation'

(48b) ʔana: ʔa-bʁa **ʔa-ru:ħ** ʔiʒaza
I PRS.1SG.M-want to **1SG.M.PRS**-go vacation

II- Using the HA Third Person Singular Masculine Past Form Rather than Other Subjects

In HA, the verb in example (49a) would be the 1SG present form of the verb. Example (49b) demonstrates how the sentence would appear in HA.

(49a) ʔana: **kallam** walad
I speak.**3SG.PST** son
'I ask my son' [13]

(49b) **ʔa-gu:l** li: waladi:
PRS.1SG.M-say to my son

III- Using the HA Singular Masculine Imperative Form of the Verb in Place of the Inflected Verb

In examples (50a), (51a) and (52a), HM3 employed the singular masculine imperative form of the verbs. The form of the HA verbs applied in examples (50b) and (51b) would be the 1SG.M present form and 3SG.M present form in example (52b), as shown below:

(50a) ʔana: **ru:h** ʔiʒaza
I go.**2SG.M.IMP** vacation
'I will go on vacation'

(50b) ʔana: **ʔa-ru:ħ** ʔiʒaza
I **1SG.M.PRS**-go vacation

[13] The literal translation is 'I say to my son'.

(51a) **Rakkib** ʔiʒaza

Build[14].**2SG.M.IMP** vacation

'I would like to go on vacation'

(51b) ʔa-bʁa **ʔa-xud** ʔiʒaza

PRS.1SG.M-want to **PRS.1SG.M**-take vacation

(52a) Huwa **ta:l** fi: muskila baadi:n

He come.**2SG.M.IMP** exist problem later

'If he comes to Delhi, it will be a problem'

(52b) ʔiza**:** **ji-ʒi:-ni:** ʕala Dili, ħatku:n muʃkila

If **PRS.3SG.M**-come to Delhi, it will be problem

IV- Replacing the Verb with a Noun

In (53a), HM3 has used the noun *kala:m* 'speech' where a native speaker of HA would use inflected verb *ʔa-gu:l* 'PRS.1SG.M-say'. In addition, in example (54a), the individual has used the noun *ma:lu:m* 'known' where a native speaker of HA would use inflected verb *ʔa-ʕrif* 'PRS.1SG.M-know'. Examples (53b) and(54b) demonstrate how the sentences would appear in HA.

(53a) ʔana: **kala:m** ʔinta sajjara sagi:r si:l

I speech you car little carry

'I tell him to come in his little car'.

(53b) ʔana: **ʔa-gu:l** lu: ji-ʒi:b sajjara-tu: ʔa-sˁaʁi:ra

I PRS.1SG.M-tell to him PRS.3SG.M-bring car-POSS DEF-little

[14] The literal translation is 'I want to take vocation'.

(54a) Ma-fi: **ma:lu:m**
No- exist known
'I don't know'

(54b) Ma: **ʔa-ʕrif**
No PRS.1SG.M-know

V- Verb Omission

In example (55a), HM3 omitted the verb where the meaning was apparent without it. In HA, the verb *ʔa-wsˁal* 'PRS.1SG.M-arrive'. The HA version would be in example (55b)

(55a) Mumkin tala:ta sa:ʔa ʔana: ʔalatu:l ∅ be:t
Possibly three hour.S I immediately ∅ home
'I'll possibly **arrive** home in three hours'.

(55b) Jimkin tala:ta sa:ʕ-a:t wa **ʔa-wsˁal** ʔa-lbji:t
Maybe three hour-PL CONJ PRS.1SG.M-arrive DEF-home

VI- Compound Verb

In (56a), HM3 produced the compound verb as opposed to using the third plural present form of the verb in HA. The HA version would be as shown in example (56b).

(56a) ʔukti-k fi: walad bint **sawwi** **ʒawaz**
Sister-your has son daughter make wedding.**V**[15]
'Her sister has a son and daughter who will both get married'

(56b) ʔuxta-ha: ʕindaha: walad wu bint **ħa-jitzwwaʒ-u:**
Sister-her has son and daughter will-**3PL.PRS.** get married-S

[15] *Wedding* is a noun which is used for a verbal function here. (The son and daughter of HM3's sister-in-law intend to marry, although they will not have a wedding celebration).

(B) Nominal Agreement

I- Noun-Adjective Agreement

In example (57a), HM3 employed the masculine form of the adjective with the feminine noun. This adjective agrees with the noun in HA and would be feminine and definite, as shown in (57b).

(57a) Sajjar-a sagi:r
Car.SG.F little.SG.**M**
'His little car'.

(57b) Sajjarat-u ʔa-sˤaɤi:r-**a**
Car.SG.F-POSS[16] DEF-little.SG-**F**

II- Number marking on nouns

In examples (58a) and (59a), HM3 used the singular form of the nouns as an alternative to using the plural form in HA, as they would be employed in (58b) and (59b).

(58a) ʔarba sahar
Four month.**SG**
'Four months'

(58b) ʔarbaʕa ʃuh**u:r**
Four month.**PL**

(59a) ʔasara sa:ʔa
Ten hour. **SG**
'Ten hours'

(59b) ʕaʃara sa:ʕ-**a:t**
Ten hour-**PL**

[16] The possessive makes the noun definite in HA.

In example (60a), HM3 used a singular form of the noun *tajja:ra* ‘airplane’ in preference to using the plural noun for HA, as revealed in example (60b).

(60a) Tajja:ra ʔa-ftah
Airplane.S IMP.2SG.M-open
‘When the flights begin’

(60b) Lamma: ʔa-tˁtˁajara:n ji-ftaħ
When DEF-flight.**PL** PRS.3SG.M.open

4.3.3 FF4, a L1 Tagalog Speaker

I. Excerpt from the Tagalog Participant’s Speech.

إيوا، اتنين ولد. هدا قبير تلاتين واهد تلاتين. بادين سقير تمنيه إسرين. هدا قبير شقل أبوظبي. كمان اتنين بنت. بنت سقير شقل *سنقابور.

أنا شقل سائه تلاته بادين روه سائه إهداسر. إجي سائه سائه ساباه قوم. سوي قسيل ترويس بادين إجي شقل. بادين شقل هنا سوي *ندابه هق همام. بادين إقلس. باد سوي كله *نديب سوي اقلس. بادين واهد سائه كمان سوي شوف همام *نديب. بس واهد سائه سوي *نديب. بادين إقلس سويه. سائه اهداسر روه بيت. إجاسه يوم هدا يوم جماه. بس واهد يوم جماه. سوي بيت. سوي *نديب بيت.

أشان في كورونا *مابي شقل. إقلس بيت. هدا واهد سهر بس إقلس بيت. *بلوس واهد شهر. هدا *بي شقل بس مامريدا. بادين *بي سويه سويه مامريدا شقل إجي هدا قورونا. *بي هدا تاني سديق هنا *مستسبى إجي . بادين *بي كورونا. إقلِس أربأتسر يوم *بي هدا بيت *قربه. هدا سيستر انت سديق إجي هنا بادين جيب أكل. أشان *مابي أكل. *بي هدا سديق واهد موت. ما أريف أنا. بس قلام *بي كورونا. *ما أريب أنا. أنا بس *أريب *بي هدا *مستسبى. هيا كلام *بي كورونا موت *مستسبى.

II. A Transcription of How a Native Speaker of HA would Speak.

إيوا، ولدين. الكبير عمره تلاتين، واحد وتلاتين. الصغير تمنيه وعشرين. الكبير يشتغل في أبو ظبي. كمان عندي بنتين. البنت الصغيرة تشتغل في سنغافورة.

دوامي يبدأ الساعة تلاته وأخلص الساعة إحد عش. لما تصير الساعة سبعه أقوم، أغسل وجهي و أتروش بعدين أروح الشغل. في الشغل هنا أنظف الحمامات. لما أخلص أجلس. لما أنظف كل الحمامات، أجلس. بعد ساعه كمان أشوف إذا الحمامات نظيفه. كل ساعه أنظف. بعدين أجلس شويه. الساعة إحد عشر أروح البيت. الإجازة يوم واحد بس، الجمعة. أنظف البيت. أنظف البيت.

عشان كان وقت كورونا ما كان في شغل. جلسنا في البيت. شهر واحد بس جلسنا في البيت. أعطونا فلوس شهر واحد. الممرضات بس كان عندهم شغل. الممرضات اللي اشتغلوا بعضهم جاتهم كورونا. في وحدة من صحابتي هنا في المستشفى جاتها كورونا. جلست أربع طعشر يوم في الغرفة في البيت. كانوا صحابتها يروحوا لها و يودوا لها أكل. عشان ما عندها أكل. وحدة من البنات ماتت. ما أعرف أنا. بس قالوا جاتها كورونا. ما أعرف أنا. أنا بس أعرف إنها تشتغل في المستشفى. قالوا جاتها كورونا وماتت في المستشفى.

III. English Translation

Yes, two sons. The eldest is thirty years old, I mean thirty-one years old. The youngest is twenty-eight years old. The eldest works in Abu Dhabi. I have also two daughters. The youngest daughter is currently working in Singapore.

My shift starts at three o'clock and I finish at eleven o'clock. I get up at seven o'clock, wash my face, take a shower and then go to work. At work I clean the toilets. When I finish cleaning the toilets I sit down. After an hour, I check if the toilets are dirty and clean them. I then sit down for a while. At eleven o'clock, I go home. I only have one day off, a Friday. On my day off, I clean the house.

Due to Corona, there was no work. We stayed at home for a month, although they paid us one month's salary. Nurses who worked, some of them were infected by the coronavirus. One of my friends, here in the hospital was infected by the virus. She had to remain in her room at home for fourteen days. At that time, her friends were going to her and taking her some food because she didn't have any. One of the girls who worked in the hospital died of the virus apparently. But I didn’t know her. They only said she had the virus and worked in the hospital.

(A) Verbal Agreement

I- Using the HA Third Person Singular Masculine Present Form as opposed to other Subjects

In example (61a), FF4 used *ʔi-ʒi:* 'PRS.3SG.M-come' instead of *ʒa:-ta-hum* '1SG.PST.come-F-PL.S'. The sentence in HA would be as in example (61b).

(61a) **ʔi-ʒi:** hada korona:
PRS.3SG.M-come DEM Corona
'infected by the coronavirus.

(61b) **ʒa:-ta-hum** korona:
1SG.PST.come-**F-PL.S** Corona

II- Using the HA Third Person Singular Masculine Past Form in Preference to Other Subjects

In HA, the verb in example (62a) would be the 3PL past form of the verb. Example (62b) demonstrates how the sentence would appear in HA.

(62a) bass **kallam** pi: korona
only speak.3SG.PST exist Corona
'They only said she had Corona'

(62b) bass **ga:l-u:** ʒa:taha: korona
only say-3PL.PST.S come.3SG.F.PST Corona

III- Using the HA Singular Masculine Imperative Form of the Verb in Place of the Inflected Verb

In examples (63a), (64a) and (65a), FF4 employed the verbs in their singular masculine imperative form. The verbs in HA for the sentences (63a) and (64a) would be in singular masculine present form, as shown in (63b) and (64b), respectively. In contrast, for the sentence (65a), the HA verb would be in the plural present form, as shown in example (65b).

(63a) Sa:ʔa ʔihdasar **ru:h** bij:t
O'clock eleven go.**SG.M.IMP** home
'At eleven o'clock, I go home'.

(63b) ʔa-ssa:ʕa ʔihdasar **ʔa-ru:ħ** ʔa-lbij:t
O'clock eleven **1SG.M.PRS**-go DEF-home

(64a) Sa:ʔa saba **gu:m**
O'clock seven get up.**SG.M.IMP**
'I get up at seven o'clock'.

(64b) ʔa-ssa:ʕa saba **ʔa-gu:m**
O'clock seven **1SG.M.PRS**-get up

(65a) Sadig ʔi-ʒi: hina: baadi:n **ʒi:b** ʔakil
Friend PRS.3SG.M-come here later. ADV bring.**SG.M.IMP** food
'Her friends were going to her and taking her some food'.

(65b) Saħbat.aha: ka:nu: ji-ru:ħu: laha: wi **ji-waddu:** laha:
Friends.her were PRS.3PL-go to her CONJ **PRS.3PL-**take to her

ʔakil
food

IV- Replacing the Verb with a Noun

In example (66a), FF4 used the noun *ʃugul* 'work' instead of the verb *ji-ʃtaʁil* 'PRS.3SG.M-work'. Likewise, in examples (67a) and (68a), the migrant worker used the nouns *mout* 'die' and *kala:m* 'speech' in place of the verbs *ma:t-at* 'die-3SG.F.PST' and *ga:l-u:* 'say-3PL.PST'. These three examples would be in HA as observed in the following examples (66b), (67b) and (68b).

(66a) Hada kabi:r **ʃugul** Abu Dabi
DEM old work Abu Dhabi
'The eldest works in Abu Dhabi'

(66b) ʔal-kabi:r. **ji-ʃtaʁil** fi: Abu Dabi
DEF-eldest PRS.3SG.M-work in Abu Dhabi

(67a) Hada sadi:g wahid **mout**
DEM friend one die
'One of the girls died'

(67b) Waħda min ʔal-bana:t **ma:t-at**
One of DEF-girls die-3SG.F.PST

(68a) Hijja: **kala:m** pi: korona
She speech exist Corona
'They said she had Corona'

(68b) **Ga:l-u:** ʒa:taha: korona
Say-3PL.PST.S come.3SG.F.PST Corona

V- Verb Omission

FF4 described the period during quarantine as a result of COVID-19. They stayed at home and received one month's salary. In example (69a), the meaning can be understood from the context where the migrant worker omitted the verb *ʔaʕtˤu-na:* 'give.3PL.PST-us'. The sentence in HA would be as shown in example (69b).

(69a) Ø pulu:s wahid ʃahar
Ø money one month
'They **paid** us one month's salary'.

(69b) ʔaʕtˁu-na: fulu:s ʃahar waħid
give.3PL.PST-us money month one

VI- Compound Verb

In examples (70a), (71a) and (72a), FF4 produced the compound verb *sawwi* 'make/do.IMP' + verb, noun and adjective instead of using the HA-inflected verb. In HA, the sentences would be as demonstrated in (70b), (71b) and (72b).

(70a) **Sawwi** gasi:l tarwi:s
Do wash.**N** shower.**N**
'I wash my face and take a shower'.

(70b) ʔa-ɤassil waʒ-hi: wa ʔa-trawwaʃ
PRS.1SG-wash face-POSS CONJ **PRS.1SG**-take a shower

(71a) **Sawwi** nadi:p biji:t
Make clean.**ADJ** house
'I clean the house'.

(71b) ʔa-nadˁdˁif ʔal-biji:t
PRS.1SG-clean DEF-house

(72a) **Sawwi** ʃu:f hammam nadip
Do see.**V** toilet clean
'I check if the toilets are clean'.

(72b) ʔa-ʃu:f ʔiza ʔal-ħammam-a:t nadˁi:f-a
PRS.1SG-see if DEF-toilet-F clean.F

After mentioning how FF4 applied the verbs differently to HA, there is one verb *ʔaʕrif* 'know' which explains the subject-verb agreement. See example (73).

(73) ʔana: bass ʔa-rip

I only **PRS.1SG**-Know

'I only know'.

(B) Nominal agreement

I- Noun-Adjective Agreement

In example (74a), the adjective does not agree with the noun in gender. While *bint* 'daughter' is feminine in HA, FF4 used the adjective in its masculine form. Example (74b) shows how the adjective would appear in HA.

(74a) Bint sagi:r

Daughter.SG.**F** little.**SG.M**

'The youngest daughter'

(74b) ʔal-bint ʔal-sagi:r-**a**

DEF-daughter.SG.**F** DEF-little.SG.**F**

II- Number marking on Nouns

In instances (75a) and (76a), FF4 used singular nouns instead of dual nouns. In HA, these nouns would appear as shown in (75b) and (76b).

(75a) ʔitne:n walad

Two son.**SG**

'Two sons'

(75b) Walad-**e:n**

Son-**DU**

(76a) ʔitne:n bint

Two daughter.**SG**

'Two daughters'

(76b) Bint-**e:n**
Daughter-**DU**

FF4 works as a cleaner of the toilets in a hospital that has many nurses. However, she used a singular form of the noun *mamarida* 'nurse' in example (77a) and a singular form of the noun *hamma:m* 'toilet' in example (78a). It can also be assumed from the context that these nouns would be plural in HA, as observed in example (77b) and (78b).

(77a) Pi: suwajja suwajja mamarida ʃugul ʔi-ʒi:
COP some some nurse.**S** work PRS.3SG.M-come

hada korona:
DEM Corona
'Nurses who worked, some of them were infected by Corona'.

(77b) ʔal-mumaridˤ-**a:t** ʔilli: ʔaʃtaʁal-u: baʕdˤahum ʒat-ahum
DEF-nurse-**PL** who work.3PL.PST some of them come.3SG.PST-O

korona:
Corona

(78a) Sawwi: nada:pa hag hamma:m
Do cleaning of toilet.**S**
'I clean the toilets'

(78b) ʔa-nadˤdˤif ʔal-ħamma:m-a:t
PRS.1SG-clean DEF-toilet-**PL**

4.4 Summary

This chapter has focused on how the oldest speaker in each language group - Bengali, Hindi-Urdu and Tagalog – mark agreement in their non-native varieties of HA. The descriptive analysis set out to discover if there is simplification in the HA morphological features used by these three

migrant workers – manifested as a lack of agreement between subject and verb as well as between adjective and noun, and as number marking on nouns. I hypothesized that if agreement marking was rare, the Arabic varieties of these three speakers might be best described as a pidgin varieties. However, if agreement marking is persuasive and correct, their Arabic varieties might be an interlanguage. The analysis shows that all the three speakers use a reduced verbal agreement system and that their Arabic speech exhibits a lack of the noun-adjective agreement. These findings support the pidgin hypothesis. However, the lack of data for investigating number marking on nouns makes it difficult to support either hypothesis.

Verbal Agreement

Compared to HA expectations, there is less subject-verb agreement in the speech employed by the three migrant workers. This finding is in line with Arabic pidgin studies in Saudi Arabia in general and GPA more specifically (Albaqawi 2020; Almoaily 2012; Alsusut 2022; Bakir 2010; Hobrom 1996). My selected fieldwork data indicates that verbal forms are represented in various ways using a range of strategies: informants consistently employ either the present or past form of the HA verb in the third person singular masculine, irrespective of the subject's gender, number or person. The three migrant workers also replace the HA present or past verb with the imperative form of the verb. In some cases, they use a related noun-form instead of the HA verb-form or omit the verb when its meaning is clear from the context. In using these strategies, BM3, HM3 and FF4 are similar to the GPA speakers in the studies completed by Albaqawi (2020) and Almoaily (2012). The three migrant workers also formed the compound verb by using sawwi ‘to do/make’ + a noun, an adjective or a verb instead of using an inflected-HA verb. This finding agrees with the results in the studies conducted by Almoaily (2012), Bakir (2010) and Naess (2008). Based on Table (7), I predicted that the Hindi-Urdu speaker would use the subject-verb agreement more than the Bengali and Tagalog speakers. However, the three migrant workers share the simplification of the verbal agreement, which suggests that their Arabic varieties might be better described as a pidgin.

Nominal Agreement

The three migrant workers predominantly used adjectives that did not agree with the nouns in gender and number. This finding mirrors previous findings reported by Albaqawi (2020: 252), Almoaily (2012: 148) and Alsusut (2022: 335). As an alternative strategy, the migrant workers in

my study utilised the singular masculine form of the adjective with every noun. Based on the literature relating to migrant workers' L1, I hypothesized that the Hindi-Urdu and Tagalog speakers would use the noun-adjective agreement more than the Bengali speaker. However, the results demonstrate that all three speakers share the same type of simplification.

Regarding the morphological inflection on nouns, the three speakers primarily employed nouns without any affixes added to the base-form. This result is in agreement with the GPA study conducted by Alghamdi (2014: 120). Nevertheless, because of limited data, it is difficult to decide – based on only this aspect of morphological marking - whether the Arabic varieties of the migrant workers in my study are a pidgin language or an interlanguage. The following are the details. First, I only have data relating to the plural marking on nouns in the speech of the L1 Hindi-Urdu speaker and the L1 Tagalog speaker. Essentially, they used the singular form of the nouns instead of using the plural form. Although nouns in Hindi-Urdu inflect for number, the plural marking is absent in the L2 speech of the L1 Hindi-Urdu speaker. With regard to the L1 Tagalog speaker, the absence of plural marking could be attributed to Tagalog nouns not inflecting for number. Moreover, although Tagalog includes the particle *mga* which can be applied to indicate plurality, its use is optional. Therefore, the optional use of this particle in the L1 of the Tagalog speaker may well explain the absence of the plural form in their L2 Arabic. Second, there are only three examples in my data that require dual marking on nouns. Nonetheless, the L1 Bengali speaker and the L1 Tagalog speakers did not use this marker in their L2 speech. This failure could be attributable to the influence of their L1s, as neither Bengali nor Tagalog inflects for dual marking. Third, I only have data relating to nouns with cardinal numbers from three to ten in the speech of the L1 Bengali speaker and the L1 Hindi-Urdu speaker. Essentially, they employed the singular form of the noun as opposed to using the plural form, probably because of the influence of their L1.

The following two chapters will explore the morphosyntactic features, specifically the definite article and the coordinating conjunction marker. This is followed by a chapter about the phonological feature, /f/. While the morphological features are examined using a qualitative method in the current chapter, the morphosyntactic and phonological features are analysed by means of a quantitative method. In the final chapter, the results from each of the different chapters

will be brought together to assess the combined evidence regarding the pidgin or interlanguage hypotheses.

Chapter 5: The Definite Article

5.1 Introduction

The focus in this chapter is on a common morpho-syntactic features in HA: an inflectional morpheme that characterise noun phrases, i.e., definiteness. To be specific, the focus is on the definite article */ʔal-/* 'the' in the Arabic speech of migrants who speak Bengali, Hindi-Urdu and Tagalog as their L1 and work in Madinah. As mentioned in Section 3.9, the definite article is a categorical variable in HA (see Il-Hazmy 1975 & Omar 1975). Nonetheless, my analysis treats this factor as a binary variable, as evidenced by my data and findings from earlier studies by Albaqawi (2020) and Almoaily (2012). The variable has two variants: +DEF indicates the use of a definite article, which is the standard form in HA, whereas -DEF indicates the absence of the variable.

My choice of the definite article is influenced by several quantitative and qualitative studies of GPA that have investigated this variable use in migrant workers' speech. If my literature review indicates that the independent variables such as L1, LoR and sex typically do not have a significant impact on the use of the definite article and the migrants mostly delete the variable, for example in the results of these studies, I can predict that their Arabic varieties are pidgin.

Almoaily (2012: 113) found that in GPA, the definiteness marker */ʔal-/* is typically absent, although certain participants infrequently produce it. He established that there is intra-group variation both among those speaking Punjabi and those speaking Malayalam as their L1. Almoaily argued that this intra-group variability might be due to influence from other independent variables which were not tested in his particular study, for instance the participants' attitude towards learning Gulf Arabic (GA), the various language learning abilities of the informants and their amount of exposure to GA. Almoaily also asserted that Bengali was the only language in his study that comprised a definite article, and that L1 Malayalam speakers produced the definite article more than the speakers of other languages. Based on these findings, he concluded that there was no significant correlation between participants' use of definite articles in their substrate languages and their production of the definite article in GPA. Regarding the effect of LoR, his results revealed

that long-term residents had non-significantly higher rates of definite article use than short-term residents.

In her study of female migrant workers, Albaqawi (2020) found that although definiteness can be found in some of the substrate languages such as Sundanese, Bengali and Tagalog and not others such as Sinhala, Malayalam and Punjabi, the predictor L1 does not play a significant role in their participants' use of the definite article in GPA. However, LoR has a significant effect on the use of the definite article. The long-term residents tend to shift towards using the definite article in GPA more than the newly arrived migrants. To ascertain gender variation, Albaqawi examined the effect of LoR and compared the use of the definite article by female migrant workers in her study with male migrant workers described in the study carried out by Almoaily (2012). In her research, she determined no statistically significant difference in the use of the definite article between men and women.

Qualitative studies also suggest that migrant workers regularly omit the definite article. This deletion, which suggests simplification in comparison to the lexifier language, signifies that the variety could be a pidgin, not an interlanguage. Alzubeiry (2015: 52) observed that the definite article is omitted in the speech of migrant workers with different L1s, including Malayalam, Tagalog, Indonesian, Bengali and Hindi-Urdu. Al-Azraqi (2011: 165) discovered that none of her participants, who are Asian-origin workers from Indonesia, the Philippines, Bangladesh and India, produced the definite article except with the lexical item *ilyum* 'today'. It is possible that the participants acquired this word as a single lexical item. In 2020, Al-Azraqi did not consider the definite article as a productive marker since migrant workers in her sample produced it only in three constructions: one token for *ilwalad* 'the boy' and two tokens for *alyum* 'today'. Moreover, Albakrawi (2012: 129) and Hobrom (1996: 78) found that migrant workers generally omitted the definite article except when using various formulaic utterances like *alhamdu lillah* 'thank God'. Instead of using the definite article, the participants in Albakrawi's (2012) study used pronouns to demonstrate definiteness, most typically the third person masculine pronoun *hu:*. Hobrom (1996: 82) found that the participants in his study who speak Kannada, Malayalam and Tamil used demonstratives rather than the definite article; he attributes this behaviour to the impact of their L1.

The use of the definite article in some of the L1s of the migrants in Madinah is controversial, whereas other migrants' L1s do not have overt articles. Furthermore, there is disagreement among academics as to whether or not languages that lack definite articles have a DP layer. Abney (1987) proposed the *DP Hypothesis* in syntax in which he considered a noun phrase to only be considered a DP when it has a functional determiner that the NP can serve as a compliment to. Since then, there has been no agreement among researchers in the literature regarding whether or not article-less languages project a DP. Zlatic (1997), Bošković (2008; 2009) and Bošković and Gajewski (2011) claimed that when a language does not have a definite article, it is considered a NP language and has no DP layer. Specifically, they believe that there is a difference between languages with articles and languages without articles in the nominal structure. Whereas nominals in languages that have articles are DP, they are NP in languages that do not have articles. However, assuming that these languages project a DP is not impossible. Due to Universal-DP Hypothesis, whether the article is present or absent in the noun phrase, the syntactic structure is universal (Pereltsvaig 2007). Hence, for the universalist, the DP is the default category. Given that the L1s of the migrant workers in my study may be classified as DP languages, it would not be surprising if L1 background influenced the migrant workers' variable use of the HA definite article. If the participants' L1 and the other contextual variables influence their use of the definite article, then their Arabic varieties may best be described as an interlanguage.

How definiteness is used in the superstrate language, Hijazi Arabic, and in the substrate languages (Bengali, Hindi-Urdu and Tagalog), is reviewed in Section 5.2. Section 5.3 presents the results of the distributional analysis. Section 4.4 presents the results of the regression analyses; analyses that distinguish speakers by L1 will uncover whether each language group has its own Arabic variety or if all three groups share a simplified variety of Arabic language. The results reveal that the three different L1 groups share the same Arabic variety, which is simplified because speakers generally delete the definite article. However, there is inter-group variation in terms of social constraints on variable use. I will discuss in Section 5.5 what these inter-group differences tell us about supporting the interlanguage hypothesis.

5.2 The Definite Article across HA and L1s

5.2.1 The Definite Article in Hijazi Arabic

While Arabic does not have an indefinite article that is equivalent to 'a' and 'an', the prefix */ʔal-/* 'the' is used to mark definiteness (Il-Hazmy 1975: 175; Omar 1975: 14), as example (79) shows[17].

(79) Ahmad fataħ **ʔal**-ba:b
Ahmad open.1SG.M.PST **DEF**-door
'Ahmad opened the door'

Based on the description provided by Arab grammarians in relation to the pronunciation of */ʔal-/* , */ʔal-/* is assimilated into the following sound when it precedes coronal sounds (Kenstowicz 1994; Abdelhady 2019: 9; Omar 1975: 4). To illustrate this, when */ʔal-/* is combined with the word */tuffaħ/* 'apples', it becomes */ʔat-tuffaħ*/. In contrast, when */ʔal-/* precedes non-coronal sounds, there is no assimilation (Kenstowicz 1994; Abdelhady 2019: 9). For instance, */ħali:b/* 'milk' is pronounced as /*ʔalħali:b*/.

Al Barrag and Alzahrani (2019: 269) argued that in HA, determiner (D) - which is the locus of the feature of definiteness - is the head of the noun phrase (NP). The determiner phrase (DP) includes a root which is c-commanded by the nominaliser and by gender, number and determiner. Since Arabic has an overt definite article, it is noticeable that it has a DP (Fehri 1999).

Nouns in HA are divided into definite and indefinite (Al Barrag & Alzahrani 2019). There are six types of definite nouns: personal pronouns, demonstrative pronouns, relative pronouns, construct-state construction, proper nouns and bare nouns that become definite when attached to the definite article (Al-Ansari 2007: 20). Other categories of nouns are considered indefinite nouns. My study focuses on the last three types of definite nouns for the reason that only bares, the construct state and a few of proper nouns can be attached to the definite article. This is illustrated in the following.

[17] The instances in the HA sub-sections of this chapter and the ones that follow were primarily provided by me, a native speaker of the language.

Bares are indefinite. However, when they are preceded with */ʔal-/,* they become definite (Al-Ansari 2007).

(80a) Mufta:ħ
Key.INDEF.SG
'a key'

(80b) **ʔal**-mufta:ħ
DEF- key.SG
'the key'

When indefinite nouns are specified by definite nouns, this is termed construct-state construction (Almalky 2020: 18).

(81) Mufta:ħ **ʔal**-ɣurfa
Key **DEF**-room.SG.F
'The room's key'

Proper nouns include names of people (Al-Ansari 2007), countries and cities.

(82) **ʔat-**Taif
DEF-Taif
'Taif'

The definite article is also attached to nouns as well as numerals and adjectives (Almalky 2020: 16).

(83) **ʔal**-bint **ʔal**-gami:l-a
DEF-girl.SG **DEF**-beautiful.SG.F
'The beautiful girl'

The following three sub-sections review definite article use in the three substrate languages of the migrant workers in my study[18].

5.2.2 The Definite Article in Bengali

Bengali is one of the languages that does not have definite articles (Thompson 2012). To make nouns definite, Bengali speakers add a small number of classifiers to nouns (Thompson 2012: 50). The suffix *–ta,* which is demonstrative and the default classifier in Bengali with its different forms, is used to express definiteness indirectly (Ray, Abdul Hai & Ray 1966). The different suffix-types inflect for size such as *-ti* 'that small' and for number, such as *-to* 'those two' and *-tin* 'those four' but not for gender. In English translation, this demonstrative and its forms can mean 'that/those individual unit(s)'. *-a* and *-i* ' the unit fraction/piece' and *gulo/guli* 'these/those' are also Bengali demonstratives which can be used to express definiteness indirectly (Ray, Abdul Hai & Ray 1966).

Alok and Mishra (2013) have argued that Indo-Aryan languages, such as Bengali, Magahi, Gujrati and Hindi, do not have the DP because overt articles do not exist in these languages. Alternatively, Syed and Simpson (2017: 2) maintained that whereas Bengali is a language without an article, it has a DP structure level. The distinct word-order alternations in Bengali make its patterns resemble a language with an article. Numeral> classifier> adjective> noun is the canonical order of a nominal phrase in Bengali that produces an interpretation of indefiniteness, as example (84) shows.

(84) Du-to **lal** **boi**
Two-CLF **red** **book**
'Two red books'
(Syed & Simpson 2017: 2)

The inverted order, however, which occurs when any adjective phrases and a noun precede the numeral and classifier, produces an interpretation of definiteness, as example (85) illustrates.

[18] In this and the following chapter, due to the absence of available empirical studies, the information provided in these sub-sections is derived from descriptive reference grammars.

(85) **lal** **boi** du-to
red **book** two-CLF
'The two red books'
(Syed & Simpson 2017: 2).

In Bengali, moving the noun and any preceding adjectival phrases in the phrasal complement of the classifier to the left has the same effect on interpretation that adding a definite article does in languages like Italian, English and German (Bhattacharya 1999a). This raises a complex question: whether having a DP in a language necessarily requires the presence of an article or determiner, or if it is related more to the specific structural positions within nominals. L1 Bengali speakers lack overt articles, but there may be indications of abstract DP structure. Bhattacharya (1999b) claimed that DP in Bengali has three layers. He added that the quantifier phrase (QP) is in the middle between the higher layer which is the DP and the lower layer which is the NP. Similarly, the classifier occurs as the head of QP (Chacón 2012: 9). The movement of the NP to the QP specifier is related to the specificity (Bhattacharya 1999b), whilst in relation to the DP specifier, it is associated with definiteness (Chacón 2012).

In addition, Syed and Simpson (2017) discussed Bošković's (2008) three generalisations to provide evidence that Bengali has a DP layer and does not behave like NP languages. Bošković (2008) claimed that only languages that have articles allow negative raising, which 'refers to the phenomenon in which an instance of negation in a higher clause can be understood to negate the content of a lower clause' (Syed and Simpson 2017: 4), the majority reading of 'most' and two lexical genitives in transitive nominals. However, Syed and Simpson ascertained that although Bengali does not have a definite article, it allows these generalisations. Therefore, Syed (2017) divided the languages of the world into three categories: (a) language with a definite article and DP layer like English, (b) language without a definite article, which does not have a DP layer, such as Russian, and (c) language without a definite article but with distinct word order alternation and comprise a DP, for example Bengali. Moreover, Syed and Simpson (2017) indicated that the relevant patterns of a few of Bošković's other generalisations do not occur in Bengali; thus, they cannot be scrutinised in this language. Likewise, seeking to examine other patterns which can be

tested in Bengali, for instance left branch extraction, is inconclusive because they do not provide a strong and reliable diagnostic to differentiate between DP and NP languages.

Regarding my research, if Bengali patterns like a DP language and the results reveal that L1 Bengali speakers regularly use the definite article when they speak Arabic, their variety can potentially be described as an interlanguage.

5.2.3 The Definite Article in Hindi-Urdu

As mentioned in the previous sub-section 5.2.2, that there is no article in Indo-Aryan languages comparable to the definite article 'the' in English (Junghare 1983: 116), with Hindi-Urdu being among these languages. Junghare (1983: 116) posits that conveying definiteness, a semantic notion, does not rely solely on syntactical and morphological markings; it can also be achieved through discourse mechanisms.

Syed (2017: 9) claimed that whereas Bengali behaves like a language that has articles, Hindi acts like an article-less language similar to Serbo-Croatian and Russian because it does not employ word order alternations to obtain the interpretation of definiteness.

On the other hand, Junghare (1983) and Agarwal (2019) argued that one of the ways to indicate definiteness in Hindi-Urdu is using the suffix -ko to mark the indirect object and the direct object when it is animate or human nouns. The following example (86) illustrates this.

(86)	maine	aj	kutte-**ko**	xarida
	I	today	dog-**DEF**	buy.PST

'I bought the dog today'

(Junghare 1983: 119)

Moreover, Junghare (1983) and Agarwal (2019) propose that topicalization is another device to refer to definiteness in Hindi-Urdu. Example (87a) shows that the sentence has indefinite meaning; however, when the NP is topicalized, the sentence can have a definite meaning in example (87b) (Junghare 1983:124).

(87a) bagice me **phul** the
garden in flowers were
'There were flowers in the garden'
(Junghare 1983: 124)

(87b) **phul** the bagice me
flowers were garden in
'There were flowers in the garden'
(Junghare 1983: 124)

However, it is not always that initial NP is a topic; as example (88) shows, the NP can be a subject and a topic (Junghare 1983: 124).

(88) paski ura
bird fly.PST
'A/The bird flew'
(Junghare 1983: 124)

Additionally, Agarwal (2019: 3) maintained that Hindi-Urdu nominals behave like languages that have overt determiners regarding syntactic properties, suggesting that determiners must be present. Agarwal established that Hindi-Urdu performs like DP languages in six of Bošković's (2008) generalisations, specifically negative raising, the majority reading of 'most', two lexical genitives in transitive nominals, inverse scope, sequence of tenses and negative constituents are focused. Only two generalisations can be made regarding Hindi-Urdu patterns with NP languages: its capacity for radical pro-drop and the absence of exhaustivity presuppositions in possessors. In addition, Agarwal showed evidence that Hindi-Urdu patterns similar to languages with articles. For instance, it is possible to have multiple lexical genitives within the same nominal phrase and the genitive marking on each of these genitives agrees with the head noun, which is similar to how determiners agree with the nouns they precede in NP languages. While the Polish language lacks a DP layer because it allows only one argument, the Arabic language and German allow two

arguments (Willim 2000). Following this, Agarwal suggests that Hindi-Urdu is perform similar to these languages and contains a DP layer.

With regard to the present study, even though Hindi-Urdu is a language without a definite article, it can be considered an article language. Therefore, we might expect Hindi-Urdu speakers to use the definite article in their Arabic variety.

5.2.4 The Definite Article in Tagalog

The status of Tagalog - whether it is a language with article or without article - is up for debate (Crosthwaite & Choy 2016: 1). According to Collins (2019), the Tagalog language does not have a definite article. In Tagalog, common noun phrases are preceded by specific grammatical function markers, such as *ang* and *ng* (Paul, Cortes& Milambiling 2015: 363). These markers are variably termed determiners, construction markers, proclitic or case markers (Reid 2002: 295). To establish whether *ang* and *ng* pattern with determiners, Cortes, Milambiling and Paul (2012: 2) applied the NP/DP generalisations of Bošković (2008) on Tagalog. If *ang* and *ng* are in D, Tagalog can be classified as a DP language. However, if they are not in D, but in other heads in the NP, Tagalog can be classified as NP language. The findings are inconsistent. They showed that Tagalog resembles both a DP language and an NP language, as shown in Table 8. While out of 13 tests, six indicate Tagalog patterns as a DP language, seven tests demonstrate that Tagalog is similar to a NP language. Therefore, Bošković's tests are unhelpful as regards deciding Tagalog NP/DP status. Moreover, to determine the syntactic category of *ang* and *ng*, Paul, Cortes and Milambiling (2015: 366) focused on the syntactic position D. The two properties of D are true of *ang* and *ng.* In regard to the noun phrase, they are functional heads. Moreover, they turn a predicate into an argument. Therefore, the syntactic test is compatible with the analysis of *ang* and *ng* as D. Schachter (1976: 495) mentioned that the *ang* phrase is a topic, not a subject and in the equivalent of English, a topic must always be marked as definite. Moreover, Schachter and Otanes (1983: 529) indicated that when an unmarked noun is preceded by the *ang* particle and it occurs in the predicate position, it is given a definite meaning. Without this, it typically expresses a generic or indefinite meaning. The answers to questions that start with *sino, ang ano* and *(ang) din* 'who, what and which' respectively are frequently definitive predicates such as:

(89) sino **ang** titser?
who **DEF** teacher?
'Who is the teacher?'
(Schachter & Otanes 1983: 529)

(90) **ang** American ang titser.
DEF American FOC.SG teacher
'The American is the teacher'
(Schachter & Otanes 1983: 529).

Table 8: Tagalog patterns as NP and DP language (Adapted from Cortes, Milambiling & Paul 2012: 2).

Tagalog patterns as a NP language	**Tagalog patterns as a DP language**
Inverse scope	Negative raising
Exhaustivity of possessives	Adjunct raising
Clitic doubling	Two lexical genitives
Radical pro-drop	Cross-clausal scrambling
Majority reading of 'most'	Obligatory# morphology
Possessors and adjective variable order	Left branch extraction
Second-position clitics	

Notwithstanding that the status of Tagalog as a DP or a NP language is controversial, my study predicts that if Tagalog speakers apply the definite article when they speak Arabic, this indicates that their variety might well be considered an interlanguage.

5.2.5 Summary

Table 9 below summarises the use of the definite article in the three substrate languages and in the superstrate language; it also indicates whether these languages can be considered a DP language. The substrate languages relevant to my study can convey definiteness even in an indirect way and may pattern as DP languages. If migrant workers use the HA definite article, their Arabic varieties

might be better described as interlanguage. However, the lack of differences between the migrant groups' use of the HA definite article could suggest that their varieties are pidgin.

Table 9: A cross-linguistic comparison of definite article and DP of the substrate languages and superstrate language

Feature	Language			
	HA	**Bengali**	**Hind-Urdu**	**Tagalg**
Definiteness	/ʔal-/ (prefix)	ø Markers (Expressed indirectly -ta)	ø Markers (the suffix -ko & topicalization)	ø Markers (ang)
DP	Yes	It can be	It can be	It can be

5.3 Overall Distribution of the Definite Article

This section examines Hijazi migrant workers' use of definite articles to determine the Arabic status of their speech, specifically whether it represents a pidgin or an interlanguage. Table 10 shows that the definite article is employed by migrant workers variably. The rates for using the definite article +DEF are consistently far lower than those for using the -DEF variant. Migrant workers as a whole use +DEF with a rate of (6%) and employ -DEF with a rate of (94%). There is a some rather negligible variation across the three different L1 speaker groups.

Table 10: Variant frequencies of the definite article across L1 speaker groups.

L1	The definite article variants				Total
	-DEF		+DEF		
	%	(N)	%	(N)	
Bengali	95%	(2794)	5%	(151)	2945
Hindi-Urdu	94%	(3827)	6%	(257)	4084
Tagalog	93%	(2715)	7%	(197)	2912
Total	**94%**	**(9336)**	**6%**	**(605)**	**9941**

5.4 Results of the Definite Article Regression Analyses

Based on the results shown in Table 10, migrant workers mostly delete +DEF. This indicates that their Arabic varieties may well be described as a pidgin. In this section, I present the results of regression analysis concerning the use of +DEF in four sub-sections. The first sub-section includes the model for each of the sampled migrant workers. This is followed by three additional sub-sections, each dedicated to a specific language group, demonstrating how each language group uses +DEF and whether or not their use of the Arabic definite article is conditioned by the contextual variables.

Each model in every sub-section has specific predictors that may not be tested in the other models in different sub-sections. As mentioned earlier in Section 3.11, the predictors for each regression model were selected based on the last version of the 'stepwise regression' findings, with the aim of addressing the issue of multicollinearity. I used the AIC as a guiding heuristic. The best-fit model is the one with the lowest AIC.

Moreover, the number of predictors in the lowest AIC models are greater than that presented in the following regression models. However, increasing the number of predictors in any of the subsequent regression models can lead to multicollinearity, non-orthogonal variables, singularity problems or render all the factors insignificant.

Because the definite article is used categorically in HA and, as far as I am aware, there is no study that examines the impact of an internal variable on the application of this feature, I examine the impact of lexical semantics on the use of the definite article. This variable and the L1 of the migrants were tested only in the first model that consists of all the sampled migrant workers for the reasons mentioned earlier. Lexical semantics was coded as ‘day/part of the day’, ‘formulaic utterances’, ‘miscellaneous’, ‘clothes’ and ‘place/city/country’. Albakrawi (2012: 129) and Hobrom (1996: 78) established that in their qualitative studies, the migrant workers applied the definite article in various formulaic utterances. Therefore, I included ‘formulaic utterances’ as a predictor level. I also included the 'grammatical category' as an internal variable. Bare nouns, the construct state and proper nouns serve as predictor levels concerning this variable. However, this variable has not been selected for testing in any model, for the reasons mentioned previously. In

addition, none of the models in sub-sections 5.4.2- 5.4.4 include any internal predictors for the same reasons.

As mentioned in Section 3.10, the majority of independent variables examined in my study are categorical. However, six are treated as continuous because they lack distinct predictor levels. These variables are motivation, attitudes toward HA, Madinah and its inhabitants, LoR, age, proficiency, as well as language anxiety.

5.4.1 Across all the Speakers.

Here I present the large general model, which includes all sampled migrant speakers from various L1 backgrounds, and treated them as a single social group. In this model, the impact of five predictors - lexical semantics, L1, the identity of the migrants, LoR and educational level in L1 - on the use of the definite article by all sampled migrant workers is examined. None of these predictors significantly affect variant choice, as Table 11 reveals. The predictor level 'formulaic utterances' demonstrates a trend towards favouring the use of the definite article compared to the reference 'clothes'.

Table 11: Mixed-effects results showing the contribution of independent variables on the use of definite article across all speakers in the corpus.

Predictor	Estimates	Std. Error	Z value	Pr (>|z|)
Intercept	-16.9785	9.5112	-1.785	0.0742 ·
Lexical semantics Clothes (reference)				
Day/part of the day	14.8044	9.4891	1.560	0.1187
Formulaic utterances	16.5994	9.4884	1.749	0.0802 ·
Miscellaneous	11.9084	9.4887	1.255	0.2095

Place/city/country	12.2247	9.4892	1.288	0.1977
L1 Bengali (reference)				
Hindi-Urdu	0.5889	0.5134	1.147	0.2513
Tagalog	0.0224	0.5860	0.038	0.9695
The identity of the migrant Permanent migration (reference)				
Temporary migration	0.1408	0.4881	0.289	0.7729
LoR	0.03895	0.0349	1.115	0.2649
Educational level in L1 University and above (reference)				
High school	-0.3047	0.5442	-0.560	0.5756
Primary school	-0.2768	0.6112	-0.453	0.6507
Random effect: speaker Variance: 1.089 Std. Dev: 1.043				

5.4.2 L1 Bengali Speakers

The model in Table 12 tests how three independent variables, LoR, language anxiety and motivation, influence the use of the definite article by L1 Bengali speakers. Whereas language anxiety and motivation do not significantly affect the use of the definite article in this model, LoR has a significant impact on its usage. The correlation between LoR and the use of the definite article is positive; as LoR in Hijaz of the L1 Bengali speakers increases, their use of the definite article increases.

Table 12: Mixed-effects logistic regression testing the effect of the independent variables on the use of the definite article by L1 Bengali speakers.

Predictor	Estimate	Std. Error	Z value	Pr (>|z|)
Intercept	-3.5382	1.4907	-2.374	**0.0176***
LoR	0.0584	0.0244	2.391	**0.0168***
Language anxiety	-0.1315	0.1613	-0.816	0.4147
Motivation	0.0979	0.2579	0.379	0.7044
Random effect: speaker Variance: 0.1623 Std. Dev: 0.4029				

5.4.3 L1 Hindi-Urdu Speakers

Table 13 shows the results of the mixed-effects logistic regression model for the use of the definite article by Hindi-Urdu migrants; it tests the effect of: LoR, language anxiety, sex and proficiency: Arabic level of the migrants at the time of the interview. It is evident that proficiency is the only predictor which significantly impacts the use of the definite article. The significance of this factor indicates that as the self-declared proficiency of L1 Hindi-Urdu speakers improves, the likelihood of using the local variant +DEF increases.

Table 13: Mixed-effects logistic regression testing the effect of the independent variables on the use of the definite article by L1 Hindi-Urdu speakers.

Predictor	Estimate	Std. Error	Z value	Pr (>|z|)
Intercept	-5.6258	0.9055	-6.213	**5.19e-10 *****
LoR	-0.0323	0.0228	-1.416	0.15682
Language anxiety	0.1227	0.0844	1.455	0.14571
Sex female (reference)				
Male	0.6508	0.3644	1.786	0.07413
Proficiency: Arabic level of the migrants at the time of the interview	0.6012	0.2223	2.704	**0.00685****
Random effect: speaker Variance: 0.1023 Std. Dev: 0.3198				

5.4.4 L1 Tagalog Speakers

Table 14 provides the results of the mixed-effect logistic regression model that examined the role of the social predictors of LoR, language anxiety and motivation on the use of the definite article by Tagalog speakers. None of these variables has a significant effect on the use of the definite article.

Table 14: Mixed-effects logistic regression testing the effect of the independent variables on the use of the definite article by L1 Tagalog speakers.

Predictor	Estimate	Std. Error	Z value	Pr (>|z|)
Intercept	-4.7135	2.5408	-1.855	0.0636 ·
LoR	-0.0308	0.0851	-0.362	0.7174
Language anxiety	-0.5664	0.6374	-0.889	0.3743
Motivation	0.8599	0.5539	1.552	0.1206
Random effect: speaker Variance: 0.8077 Std. Dev: 0.8987				

5.5 Summary

This chapter has focused on how L1 Bengali, Hindi-Urdu and Tagalog migrant workers in Madinah variably use the Arabic definite article. The analysis set out to establish the effect of carefully selected predictors on varaible use in order to determine if the migrants' Arabic varieties are best viewed as interlanguage or pidgin varieties. If the contextual factors significantly impact the migrants use of the definite article, their Arabic varieties might be described as interlanguage. However, if the predictors do not have any influence or have a marginal effect on the use of the definite article and the migrants mostly delete the variable, we may hypothesise that their Arabic varieties are pidgin.

The results in this chapter reveal that migrant workers in Madinah mostly delete the definite article. Table 15 below shows that there is limited evidence, specifically that the predictors have a significant effect on the use of the definite article supporting the interlanguage hypothesis. There are also a considerable number of null results. Null outcomes should be interpreted with caution as evidence of a pidgin, as they may not provide a solid foundation for such claims. The non-

significant effect of particular predictors; the identity of the migrant and educational level in L1 on the use of the definite article may be attributed to the smaller sample size (see Tables 43 & 44 in Appendix A). Based on my research, the evidence of an interlanguage hypothesis appears to be greater, as illustrated in the following paragraph.

Table 15: Summary of the predictors that reveal a significant effect on the use of the definite article supporting the interlanguage hypothesis.

Model	Predictor	Significant effect
All the speakers in the corpus	Lexical semantics	
	First language	
	The identity of the migrant	
	LoR	
	Educational level in L1	
L1 Bengali speakers	LoR	√
	Language anxiety	
	Motivation	
L1 Hindi-Urdu speakers	LoR	
	Language anxiety	
	Sex	
	Proficiency: Arabic level of the migrants throughout the interview	√
L1 Tagalog speakers	LoR	
	Language anxiety	
	Motivation	

The result obtained by L1 Bengali speakers regarding the significant impact of LoR is in line with the L2 studies undertaken by Drummond (2011), Flege, Bohn and Jang (1997) and Regan (2013b). It is also consistent with the study by Albaqawi (2020) which is related to GPA. Moreover, as proficiency in L2 is demonstrated to have a significant effect in the L2 studies completed by

Drummond (2010), Ulbrich (2021) and Young (1991), the self-reported level of Arabic at the time of the interview plays a significant role in the use of the definite article by L1 Hindi-Urdu speakers. These results support the interlanguage hypothesis. Additionally, the inter-group variation, which manifests as the significant impact of the length of residence on the use of the definite article by L1 Bengali speakers and a non-significant effect in the other three models, signifies that the Arabic varieties of the migrants are an interlanguage.

The migrants' L1 did not have a significant effect on the use of the definite article. Omitting the variable is a shared feature with GPA studies, for example Albakrawi (2012), Hobrom (1996), Albaqawi (2020), Almoaily (2012) and Alzubeiry (2015). Hence, this gives the impression that inflectional features, for instance the definite article, may be difficult for migrant workers to acquire. Moreover, the overt definite article in Tagalog language is debatable and Bengali and Hindi-Urdu do not have it but express it in indirect way. Likewise, the migrants L1s can be considered DP languages. Hence, we may expect that migrants can use the Arabic definite article in their speech. Seeing as the results in my study reveal that migrants predominantly delete the variable, this might be attributed to the transfer from their L1 languages, with their languages possibly being classified as NP languages. More specifically, this transfer supports the interlanguage hypothesis.

Exploring additional factors, such as the coordinating conjunction and /f/, may help in the discovery status of the Arabic varieties of the migrant workers in Madinah because concentrating solely on the analysis of one linguistic variable, +DEF, is unsatisfactory. The analysis of the second linguistic variable, the coordinating conjunction marker, in the subsequent chapter, will add more support for either the interlanguage or pidgin hypotheses.

Chapter 6: The Coordinating Conjunction Marker

6.1 Introduction

Coordinating conjunctions are generally understood to be simple to learn; this is due to simplistic ideas pertaining to parallelism. Therefore, they have not received much attention in the study of L2 acquisition (Chiu 2004: 2). There are many coordinating conjunction markers in HA, such as *wa/wu* 'and', *laakin* and *bass* 'but', and *walla*/*willa* and *aw* 'or' (Il-Hazmy 1975: 189; Omar 1975: 87-88; Sieny 1972: 41). This chapter concentrates on the linguistic variable *wa/wu* 'and' in the Arabic speech of migrant workers in Madinah who speak different L1s: Bengali, Hindi-Urdu and Tagalog. The use of the coordinating conjunction is not variable in HA – speakers of L1 HA use it categorically (see Il-Hazmy 1975 & Omar 1975). However, the coordinating conjunction marker is regarded as a binary variable in my analysis, as supported by my data and the findings obtained by Albaqawi (2020), Almoaily (2012) and Naess (2008). In the migrant workers' speech, this variable has two separate forms: +CONJ, which reveals the presence of a conjunction marker (typical of HA), whereas -CONJ represents the non-use of the variable.

The literature on interlanguage denotes that several independent variables, such as L1, LoR and sex, can significantly influence the acquisition of L2 features (see sections 2.5.1, 2.5.2 & 2.5.4). Previous studies on GPA, which I will discuss in the following paragraphs, illustrate how migrant workers use the conjunction marker variably and how the aforementioned variables notably affect its use. The significant impact of contextual variables on the use of conjunction markers by migrant workers in GPA studies encouraged me to consider the Arabic varieties of migrant workers in Madinah. If I ascertain that contextual variables significantly affect the migrants' use of the conjunction marker, should we subsequently consider their Arabic varieties to be an interlanguage or pidgin? My hypothesis is that if the migrant workers' use of the conjunction marker is significantly conditioned by contextual factors, then their Arabic varieties can be considered interlanguage. However, if migrants predominantly omit the conjunction marker and contextual predictors do not influence its use, their Arabic L2 might be characterised as a pidgin variety.

In her qualitative study, Naess (2008: 53) found that although the Arabic coordinating conjunction marker *wa* 'and' is used in GPA, migrants mostly omitted it, as illustrated in the following example taken from an L1 Sinhala migrant worker .

(91) Lazim ʃu:f ma:ma: Ø ba:ba: da:xil ʤidda
Must see.IMP mother Ø father in Jeddah
'You have to see your mother **and** father in Jeddah'
(Naess 2008: 53).

Almoaily (2012) hypothesised that L1 Bengali speakers may use the conjunction markers less than L1 Punjabi and L1 Malayalam speakers when they speak Arabic, for the reason that use of these markers is optional in Bengali while it is obligatory in the other two languages. Almoaily ascertained that whereas the general tendency for all of these male migrants is to drop the coordinating conjunction markers, as demonstrated in example (92), a number of them employed several of these markers. The most common conjunction markers are *wa* 'and' *willa* 'or' and *lakin* 'but', as in example (93). In addition, the substrate language of these migrant workers has a significant effect on the use of the conjunction markers. Punjabi speakers used the conjunction markers at a rate of (25.3%), while Bengali speakers and Malayalam speakers employed them at rates of (9.4%) and (11.7%), respectively. This finding still supports the claims about transfer.

(92) fi: axu: juni:virsi:ti: Ø sistar juni:virsi:ti:
COP brother university Ø sister university
'My brother **and** sister study at the university'
(Almoaily 2012: 92).

(93) Wahid **wu** nuss fi: il-li:l
One **and** half in DEF-night
'One **and** a half at night'
(Almoaily 2012: 92).

Regarding length of residence, it appears that this predictor also has a significant effect on the use of the conjunctions. Almoaily (2012) discovered that in all the three language groups, Bengali, Malayalam and Punjabi, the frequency of use of the conjunctions was much greater among long-term migrants, who have lived in the Gulf for ten years or more than among the short-term migrants who have resided in the Gulf for five years or less. In addition, as data reveals that there is variability in the use of conjunction markers between migrants who are in the same group.

Albaqawi (2020) examined the use of Arabic conjunction markers by female migrant workers who are from different heritage language backgrounds. Their substrate languages were Sinhala, Bengali, Tagalog, Malayalam, Punjabi and Sunda. Albaqawi noted that using the conjunction markers is optional in Sinhala, but obligatory in Malayalam, Tagalog, Punjabi and Sunda. Her results revealed conjunctions were significantly less frequent among Sinhala speakers (10.2%) and Bengali speakers (12.9%) compared to Tagalog (23.3%), Malayalam (19.3%), Punjabi (22.7%) and Sunda (24%) speakers. Regarding the influence of LoR, Albaqawi's results show that there is a positive correlation between LoR in the Gulf and the use of the conjunctions across all six language groups. In addition, female migrants who lived in the Gulf for more than ten years used the conjunction markers with a frequency of (42%), while those who stayed five years or less used them (13%). The difference between the two groups is significant. To determine gender variation, Albaqawi investigated the impact of LoR and compared the use of the conjunction markers by female migrants in her study with male migrants documented in Almoaily's (2012) study. Albaqawi established that females tend to use the conjunctions significantly more than males.

To my knowledge, none of the pidgin studies conducted on migrant workers in the Gulf countries in general or Hijaz in particular have examined the following independent variables that are examined in my study (Al-Azraqi 2020; Albaqawi 2020; Albaqawi & Oakes 2019 Alghamdi 2014; Aljutaily 2018; Al-Shurafa 2014; Almoaily 2012; Hobrom 1996; Naess 2008), namely: linguistic contexts, exposure to informal Arabic, motivation, formal instruction in their country of origin, sex and the migrants' identity. My research will fill this gap.

Section 6.2 demonstrates how the coordinating conjunction marker is used in the superstrate language, which is Hijazi Arabic, as well as the substrate languages, namely Bengali, Hindi-Urdu

and Tagalog. This section shows how these Indo-Aryan, Austronesian and Semitic contact languages are different from each other in the use of the coordinating conjunction. Furthermore, it will show how L1 might influence the participants' acquisition of HA features if their Arabic varieties are interlanguage. The findings of the distributional analysis are presented in Section 6.3. The findings of the regression analyses are presented in Section 6.4. These analyses, which differentiate speakers based on their L1, may help to determine whether each language group has its own Arabic variety or whether all three groups share a simplified variety of the Arabic language. I will argue that – broadly speaking - the results in the three models reveal that the three different L1 groups share the same simplified Arabic variety: speakers from all three groups generally delete the coordinating conjunction. However, there is still inter-group variation. Although there is limited evidence exhibiting the significant effect of the independent variables on the use of the conjunction marker by migrant workers in Madinah, the inter-group variations appear to be more effective in supporting the interlanguage hypothesis in comparison to the null results which are in favour of the pidgin hypothesis.

6.2 The Coordinating Conjunction Marker across HA and L1s

This section demonstrates how the coordinating conjunction marker is used across the three substrate languages, i.e., Bengali, Hindi-Urdu and Tagalog. The differences between them might impact the migrants' use of +CONJ if their varieties are interlanguage. This may possibly appear as a variation that indicates the effect of their L1, for instance. Alternatively, if the migrant groups do not use the conjunction frequently, external factors might have negligible influence. Thus, if the results reveal a pattern that appears unusual compared to what I predict in the interlanguage, then the Arabic varieties of the migrants might be better described as pidgin. In sub-section 6.2.1, I introduce how the conjunction marker is used in HA. Sub-sections 6.2.3 to 6.2.5 are for the coordinating conjunction across the three L1s.

6.2.1 The Coordinating Conjunction Marker in Hijazi Arabic

In HA, the coordinating conjunction marker is *wa/wu* 'and' (Il-Hazmy 1975: 189; Sieny 1972: 41). *Wu* and *wa* are used interchangeably. The following examples, (94a) and (94b), illustrate how the conjunction marker can combine two phrases and two clauses in HA.

(94a) Ahmad **wu** Kalid raaħ-u
Ahmad **and** Khalid go out.PST-2PL.M.
'Ahmad **and** Khalid went out'

(94b) Kul ʔal-xija:r **wa** ʃrab ʔal-ħali:b
Eat.IMP DEF- cucumber **and** drink.IMP DEF-milk
'Eat the cucumber **and** drink the milk'

Each connected component is preceded with *wa* in the list with more than two phrases (see Omar 1975: 174 & 188), as the following example demonstrates:

(95) Tˤajib **wu** ʒijb tˤaħina **wu** salatˤa **wu** bibsikula:
Ok **and** bring.IMP tahini **and** salad **and** pepsi
'Ok **And** bring tahini, **and** salad **and** a Pepsi'
(Omar 1975: 188)

In HA, single cardinal numbers are combined with whole tens using *wa* **(**Kheshaifaty 1997: 28; Omar 1975: 68) as shown in example (96). The coordinating conjunction can also be used to connect numbers that are over one hundred. Example (97) demonstrates this. Regarding fractions, the coordinating conjunction marker is also used as illustrated in example (98).

(96) Xamsa **wu** sitti:n
Five **and** sixty
'Sixty-five'

(97) Mijja **wu** sittah
One hundred **and** six
'One hundred **and** six'

(98) Waħid **wu** nusˁ
One **and** half
'One **and** a half'

6.2.2 The Coordinating Conjunction Marker in Bengali

Using conjunction markers is optional in Bengali (see Racova 1983; Ray, Abdul Hai & Ray 1966; Thompson 2012; Nasrin & van der Wurff 2009). One of the ways to combine words or sentences in Bengali is to apply asyndetic linkage where no conjunction is used. However, there is no variationist study of this phenomenon in Bengali. *Ar* 'and' is repeatedly left out when it combines two words that are synonyms, antonyms or have an associated meaning (see Nasrin & van der Wurff 2009: 105; Ray, Abdul Hai & Ray 1966: 58). Examples (99) and (100) illustrate how phrases can be joined in Bengali. Asyndetic is also applied when more than two words are combined, as shown in example (101) (Ray, Abdul Hai & Ray 1966: 59).

(99) Ma Ø bap
Mother Ø father
'Mother **and** father'
(Nasrin & van der Wurff 2009: 105; Ray, Abdul Hai & Ray 1966: 58)

(100) Cear Ø tebil
Chair Ø table
'Chair **and** table'
(Nasrin & van der Wurff 2009: 105)

(101) Lamba sikkhito Ø samponno
Tall educated Ø wealthy
'Tall, educated **and** wealthy'
(Ray, Abdul Hai & Ray 1966: 59).

Another technique is to make use of conjunction markers like *ar* 'and' when combining phrases and clauses (David 2015; Ray, Abdul Hai & Ray 1966). Example (102) illustrates how the phrases

are connected. In example (103), the use of the conjunction marker is similar to English, i.e., before the final item in a list.

(102) Kolkatta **ar** dhaka
Kolkatta **and** dhaka
'Kolkatta **and** Dhaka'.
(David 2015: 166).

(103) Ja-be thag-be **ar** kha-be
Go-FUT stay-FUT **and** eat-FUT
'Will go, stay **and** eat'
(Ray, Abdul Hai & Ray 1966: 59).

Regarding numbers*, ar* 'and' is not used between them in Bengali, such as *cowsotti* 'sixty-four' and *ekso tin* 'one hundred and three' (Ray, Abdul Hai & Ray 1966: 49-50). In addition, this linker is not used in *der* and *aray* 'one and a half' and 'two and a half'.

6.2.3 The Coordinating Conjunction Marker in Hindi-Urdu

Hindi-Urdu includes the conjunction 'aur,' which can connect two words as well as two phrases or clauses (Agnihorti 2006: 213; Koul 2008: 279&282; Schmidt 1999: 220). Examples (104) and (105) below illustrate how this connecter is applied in Hindi and Urdu.

(104) Mohan **aur** Sohan seb khaa rahe he
Mohan **and** Sohan apple eat AUX.PRG.M.PL AUX.PRS
'Mohan **and** Sohan are eating apples'
(Agnihorti 2013: 213).

(105) Ham ne cae pi **aur** pakora khae
We tea drink.PST **and** fritters eat.PST
'We drank tea **and** ate fritters'
(Schmidt 1999: 220).

When *aur* connects more than two phrases or clauses, it appears before the final conjunct (Koul 2008: 279; Schmidt 1999: 228). The following examples (106) and (107) are Hindi and Urdu respectively.

(106) Amar khel rahaa he Mohan gaane sun
Amar play AUX.PROG.M.SG AUX.PRS Mohan songs listen

rahaa he **aur** Sham tivi dekh rahaa
AUX.PROG.M.SG AUX.PRS **and** Sham TV see AUX.PROG.F.SG

he
AUX.PRS
'Amar is playing, Mohan is listening to songs, **and** Sham is watching television'
(Koul 2008: 279).

(107) Bara qamis, cauda rumal **aur** pandra jore jurrab
Twelve shirts, fourteen handkerchiefs **and** fifteen pair socks
'Twelve shirts, fourteen handkerchiefs **and** fifteen pairs of socks'
(Schmidt 1999: 228).

The coordinating conjunction is not used between numbers in Hindi-Urdu (Koul 2008: 99; Schmidt 1999: 228). See examples (108) and (109) below.

(108) Do sɔ Ø ek
Two hundred Ø one
'Two hundred **and** one'
(Koul 2008: 99).

(109) Do sau Ø pachattar
Two hundred Ø seventy-five
'Two hundred **and** seventy-five'
(Schmidt 1999: 228).

In addition, Hindi-Urdu speakers do not use *aur* with fractions. For instance, in Hindi *saarhe caar* 'four and a half' is utilised (Koul 2008: 92). Additionally, in Urdu, *derh* 'one and a half' is applied (Schmidt 1999: 233).

6.2.4 The Coordinating Conjunction Marker in Tagalog

At 'and' is the coordinating conjunction marker in Tagalog. In the following example (110), *at* is used to connect two words (Schachter & Otanes 1983: 540).

(110) Maganda **at** mayaman si Rosa
Beautiful **and** rich FOC.SG Rosa
'Rosa is beautiful **and** rich'
(Schachter & Otanes 1983: 540).

When *at* is used to connect more than two words, phrases or clauses, there are two specific methods (Schachter & Otanes 1983: 541). The first one uses the coordinating conjunctions, whilst the second is deleting the coordinating conjunctions except before the last word, which can be seen in English. Examples (111a) and (111b) illustrate this.

(111a) Kumanta**'t** sumayaw **at** tumawa**'t** naglaro ang
Sing.PST-**and** dance.PST **and** laugh.PST-**and** play.PST DEF

mga bata
child.PL

'The children sang **and** danced **and** laughed **and** played'
(Schachter & Otanes 1983: 541).

(111b) Kumanta sumayaw tumawa**'t** naglaro ang mga bata
Sing.PST dance.PST laugh.PST-**and** play.PST DEF child.PL
'The children sang, danced, laughed **and** played'
(Schachter & Otanes 1983: 541).

The linker *at* 'and' is employed in complex numbers in Tagalog (Schachter & Otanes 1983: 201). When any number that represents a multiple of ten is combined with another number, the connector occurs between them (see examples 112 & 113 below). Optionally, *at* is shortened to *'t,* for instance when it is preceded by [n] like in *daan at* 'hundred and' becomes *daa't* and [n] is dropped.

(112) Tatlumpu**'t** dalawa
Thirty-**and** two
'Thirty-two'
(Schachter & Otanes 1983: 201)

The coordinating conjunction *at* can also be used with fractions, for example:
(113) Isa**'t** kalahati
One-**and** half
'One **and** a half'
(Schachter & Otanes 1983: 201)

6.2.5 Summary

Table 16 is a summary of how the coordinating conjunction marker is used in the superstrate language and the three substrate languages. These differences may influence the migrants' use of +CONJ if their Arabic varieties are interlanguage. For example, the prediction is that Bengali speakers will use conjunctions the least, while Hindi-Urdu and Tagalog speakers would exhibit variations in specific linguistic contexts. However, no difference between the groups of the migrants and not making frequent use of +CONJ may suggest that their varieties are pidgin.

Table 16: The use of the coordinating conjunction marker in the substrate languages and superstrate language.

Language	The use of the coordinating conjunction marker		
	Between (two phrases or two clauses)	**Link more than two phrases.**	**Between numbers**
HA	Yes	Yes	Yes
Bengali	It is optional. 1. Ø 2. use it anywhere.	It is optional 1. Ø 2. only before the last word.	Ø
Hindi-Urdu	Yes	Only before the final conjunct	Ø
Tagalog	Yes	1. use it anywhere. 2. use it only before the last word.	Yes

6.3 Overall Distribution of the Coordinating Conjunction

This section examines the use of the coordinating conjunction marker by migrant workers in Hijaz, Madinah in order to discover their Arabic varieties; whether they are pidgin or interlanguage. The results in Table 17 reveal that all migrant speakers combined employed +CONJ with less frequency than -CONJ. The different L1 speaker groups mirror this pattern. There are distinct differences among these migrant groups, i.e., L1 Tagalog speakers used +CONJ more than the other two groups.

Table 17: Variant frequencies of the coordinating conjunction marker across L1 speaker groups.

L1	The coordinating conjunction variants				Total
	-CONJ		+CONJ		
	%	(N)	%	(N)	
Bengali	87%	(968)	13%	(142)	1110
Hindi-Urdu	89%	(892)	11%	(114)	1006
Tagalog	80%	(390)	20%	(98)	488
Total	**86%**	**(2250)**	**14%**	**(354)**	**2604**

6.4 Results of the Coordinating Conjunction Marker Regression Analyses

Based on the findings seen in Table 17, although migrant workers mostly delete the conjunction marker, the variation across the three language groups indicates their Arabic varieties might be described as an interlanguage. In this section, I present more comprehensive results concerning the use of +CONJ. The first sub-section includes all the sampled migrant workers. Sub-sections 6.4.2 – 6.4.4 investigate the variation of the conjunction marker throughout the three L1 groups.

For the reason that the coordinating conjunction marker is used categorically in HA and, as far as I am aware, there is no study that examines the effect of an internal variable on the use of this variable, I examine the influence of linguistic contexts on the use of the conjunction marker. I have created this predictor based on the literature of the coordinating conjunction marker use in the three L1s. The tokens for this independent variable were coded as 'numbers', 'two phrases', 'two clauses' and 'list of more than two phrases (non-final & final)'. If linguistic contexts and other contextual variables exhibit an influence on the use of +CONJ, the migrants' Arabic varieties could possibly be an interlanguage.

'Linguistic contexts', which is the internal variable, is examined only in the first model. The stepwise regression was employed to select predictors for each model used. This approach was adopted to minimise multicollinearity. The best model in each case was determined by the lowest

AIC value. These models differ across the three language groups as they include specific predictors in limited numbers. For example, the model for L1 Hindi-Urdu speakers includes different predictors compared to the models pertaining to L1 Bengali speakers and L1 Tagalog speakers. The addition of more variables in any model could introduce various issues (refer to Section 5.4 for more details).

Most of the factors examined in the following models are categorical variables, except for three variables, notably motivation, age and LoR, which are treated as continuous variables due to a lack of individual predictor levels.

6.4.1 Across all the Speakers

This big general model in Table 18 includes data from all migrant workers, treating them as a single social group with several differences in relation to their backgrounds. It includes seven predictors: linguistic contexts, first language, exposure to informal Arabic, amount of Arabic and L1 used, motivation, formal Arabic instruction in their country of origin and sex. Most predictors have a non-significant effect. Exceptions are: (1) 'linguistic contexts' , with, 'numbers' and 'two phrases' favouring and 'list of more than two phrases' disfavouring the use of +CONJ; (2) 'amount of Arabic and L1 used' , with, 'those who predominantly used their L1' disfavouring the use of +CONJ in contrast to the reference which is 'those who generally used Arabic'.

Table 18: Mixed-effects results showing the contribution of independent variables on the use of coordinating conjunction marker use across all speakers in the corpus.

Predictor	Estimate	Std. Error	Z value	Pr (>|z|)
Intercept	-3.3223	2.0706	-1.605	0.1086
CONJ in different contexts Two clauses (reference)				
Numbers	2.5261	0.2687	9.400	**<2e-16*** **
Two phrases	0.5967	0.2479	2.406	**0.0161***

List of more than two phrases (final& non-final)	-1.0579	0.3327	-3.180	**0.00147****
First language Bengali (reference)				
Hindi-Urdu	-0.3822	0.6556	-0.583	0.5599
Tagalog	0.0140	0.72611	0.019	0.9846
Exposure to informal Arabic No exposure (reference)				
Quran only	-0.6773	0.7278	-0.931	0.3521
Quran and more	0.2271	0.7728	0.294	0.7688
Amount of Arabic and L1 used Those who generally used Arabic (reference)				
Those who mostly used their L1	-1.2541	0.5734	-2.187	**0.0287***
Motivation	0.0597	0.3212	0.186	0.8526
Formal instruction in their country of origin Mostly no (reference)				
Mostly yes	0.3463	0.5429	0.638	0.5235
Sex Female (reference)				
Male	0.4381	0.5469	0.801	0.4231

Random effect: speaker Variance: 1.521 Std. Dev: 1.233				

6.4.2 L1 Bengali Speakers

This sub-section examines the impact of three predictors - age, sex and amount of Arabic and L1 used - on the use of the coordinating conjunction marker by L1 Bengali speakers. The mixed-effect logistic regression results in Table 19 reveals that only the age predictor has a significant effect on the use of +CONJ. The correlation between age and the use of +CONJ is negative, i.e., as the age of the L1 Bengali speakers increases, their use of the conjunction marker decreases. The amount of Arabic and L1 used, along with sex did not have a significant effect.

Table 19: Mixed-effects results for the effect of the independent variables on coordinating conjunction marker use by L1 Bengali speakers.

Predictor	Estimates	Std. Error	Z value	Pr (>|z|)
Intercept	0.50173	1.23016	0.408	0.6834
Amount of Arabic and L1 used Those who generally used Arabic (reference)				
Those who mostly used their L1	-0.92004	0.55086	-1.670	0.0949 .
Age	-0.08423	0.04138	-2.035	**0.0418 ***
Sex Female (reference)				
Male	0.75583	0.58927	1.283	0.1996
Random effect: speaker Variance: 0.3317 Std. Dev: 0.576				

6.4.3 L1 Hindi-Urdu Speakers

The findings as regards use of the coordinating conjunction marker by L1 Hindi-Urdu speakers are discussed in this sub-section. The model in Table 20 includes five fixed factors which are sex, formal Arabic instruction in Hijaz, formal Arabic instruction in the country of origin, the identity of the migrant and age. The mixed-effect logistic regression demonstrates that sex has a significant effect on variant choice: males favour +CONJ while females disfavour it. All the other predictors are statistically non-significant.

Table 20: Mixed-effects results for the effect of the independent variables on coordinating conjunction marker use by L1 Hindi-Urdu speakers.

Predictor	Estimate	Std. Error	Z value	Pr (>|z|)
Intercept	-1.34140	2.82809	-0.474	0.63528
Sex Female (reference)				
Male	2.75588	0.85376	3.228	**0.00125 ****
Formal Arabic instruction in Hijaz Those who did not receive formal instruction (reference)				
Those who received formal instruction	-1.78120	1.00435	-1.773	0.07615.
Formal Arabic instruction in the country of origin Mostly no (reference)				
Mostly yes	1.31405	1.24342	1.057	0.29060
The identity of the migrant Permanent migration (reference)				

Temporary migration	-1.74048	1.12449	-1.548	0.12167
Age	-0.03929	0.06644	-0.591	0.55424
Random effect: speaker Variance: 0.7476 Std. Dev: 0.8647				

6.4.4 L1 Tagalog Speakers

This sub-section shares the results for the use of the coordinating conjunction marker by L1 Tagalog speakers. This model in Table 21 tests four fixed factors which are LoR, age, amount of Arabic and L1 used and educational level in L1. None of these fixed factors played a significant role except the amount of Arabic and L1 used. L1 Tagalog speakers who reported using more Arabic than Tagalog favoured +CONJ, while those who reported using an equal amount of Tagalog or more Tagalog than Arabic disfavoured it.

Table 21: Mixed-effects results for the effect of the independent variables on coordinating conjunction marker use by L1 Tagalog speakers.

Predictor	Estimate	Std. Error	Z value	Pr (>|z|)
Intercept	1.20073	2.68072	0.448	0.6542
Length of residence	-0.10657	0.19737	-0.540	0.5892
Age	-0.04957	0.10924	-0.454	0.6500
Amount of Arabic and L1 used. Those who generally used Arabic (reference)				

Those who mostly used their L1	-1.94755	0.88195	-2.208	**0.0272 ***
Educational level in L1 High school (reference)				
University	0.66837	0.58328	1.146	0.2518
Random effect: speaker Variance: 0.3324 Std. Dev: 0.5766				

6.5 Summary

This chapter has focused on the use of the Arabic coordinating conjunction marker; +CONJ, by migrant workers in Madinah who speak Bengali, Hindi-Urdu and Tagalog as their L1. The aim was to establish whether their Arabic varieties are interlanguage or pidgin. If the contextual variables have an impact on migrants' use of +CONJ, it suggests their Arabic varieties can be interlanguage. On the other hand, if these predictors do not show any influence, and the migrants do not use +CONJ regularly, then their varieties might be better described as pidgin.

The findings in this chapter show that migrant workers in Madinah mostly delete the coordinating conjunction marker. They share this feature with migrants in the GPA studies, as observed in the studies undertaken by Albaqawi (2020: 137), Almoaily (2012: 92) and Naess (2008: 53). It is evident from Table 22 below that there is limited evidence, i.e., few predictors that display a significant effect on the use of the conjunction marker supporting the interlanguage hypothesis, and numerous null results. However, I have found it challenging to interpret null results as clearly favouring the pidgin hypothesis. The sample size could be the reason for the non-significant effect of some predictors; particularly formal instruction in their country of origin, as well as the identity of the migrant and educational level in L1 on the use of the conjunction marker (see Appendix A). According to my findings, the support for an interlanguage hypothesis appears to be stronger, as I will illustrate in the following section.

Table 22: Summary of the predictors that reveal a significant effect on the use of the coordinating conjunction marker supporting the interlanguage hypothesis.

Model	Predictor	Significant effect
All the speakers in the corpus	Linguistic contexts	√
	First language	
	Exposure to informal Arabic	
	Amount of Arabic and L1 used	√
	Motivation	
	Formal instruction in their country of origin	
	Sex	
L1 Bengali speakers	Amount of Arabic and L1 used	
	Age	√[19]
	Sex	
L1 Hindi-Urdu speakers	Sex	√
	Formal Arabic instruction in Hijaz	
	Formal Arabic instruction in the country of origin	
	The identity of the migrant	
	Age	
L1 Tagalog speakers	Length of residence	
	Age	
	Amount of Arabic and L1 used	√
	Educational level in L1	

The findings obtained by all the sampled migrant workers and L1 Tagalog speakers regarding the significant influence of the amount of Arabic and L1 used on the use of +CONJ is in line with previous L2 studies conducted by Flege, Frieda and Nozawa (1997), Flege, Yeni-Komshian and Liu (1999), Major (2014) and Piske, MacKay and Flege (2001). Similarly, sex has a significant

[19] Age has a significant negative effect in relation to the L1 Bengali speakers' model.

effect on the use of the conjunction marker by L1 Hindi-Urdu speakers. Males employ the conjunction marker more frequently than females (for further details, see 8.3.2). Additionally, linguistic contexts significantly affects the use of +CONJ among all sampled migrant workers. Age has a significant impact on the use of +CONJ by L1 Bengali speakers. Thus, as the age of these participants increases, their use of the L2 feature decreases. All of these significant effects of the independent variables on the use of the conjunction marker by migrant workers in Madinah indicate that their variety might be better described as an interlanguage.

In addition, the inter-group variations propose that the migrants' Arabic varieties are interlanguage. To illustrate that, the significant impact of sex, age and amount of Arabic and L1 used on the use of +CONJ in selected models and their non-significant effect in other models support the interlanguage classification.

The analysis of the third linguistic variable which is the phoneme /f/ in the following chapter will provide additional evidence to support either a pidgin hypothesis or an interlanguage hypothesis.

Chapter 7: /f/

7.1 Introduction

The focus of this chapter is on production of the HA fricative /f/. The production of [f], the standard form and the variant categorically used in HA (see Il-Hazmy 1975 & Sieny 1972), is examined in the Arabic speech of the three migrant worker groups in Madinah. Austronesian languages differ from Indo-Aryan and Semitic languages as /f/ is not part of the former's phonemic inventory. Given that the L1s differ in the phonemic status of /f/, then, it would not be surprising if L1 background influenced the mgirant workers' variable production of HA /f/. If the participants' L1 influences their production of /f/, then their Arabic varieties might possibly be described as an interlanguage.

Several qualitative GPA studies that described this linguistic variable confirm that there is variation in the production of /f/. Naess (2008: 32) determined that to some extent, migrant workers who speak Tagalog, Chavacano, Sinhala and Javanese produce [p] instead of /f/ because /f/ does not exist in the phonological inventories of these languages. However, it does not appear that any one speaker displays this behaviour consistently. For instance, a Javanese speaker produced [p] and [f] as two allophones for a single phoneme /f/ in the syllable-final position in two different words. Moreover, in his study of Indian workers who speak Tamil, Malayalam and Kannada as their first language, Hobrom (1996: 34-35) determined that they replaced many consonants in Saudi Arabic with other consonants. The consonants that they produced share a number of features with the consonants in Saudi Arabic. For instance, a migrant worker who speaks Tamil as their L1, alternated /f/ with [b] and [v] like in /fi:/ 'there is'. Hobrom clarified that Tamil does not have the voiceless labiodental fricative [f], but it does have its voiced equivalent [v]. Since /f/ does not exist in their L1, they use [v] as an alternative. In GPA studies, migrant workers produced [f, v, b, p] as allophones of /f/. In my study, /f/ was initially coded with four variants: [f], [v], [b] and [p]. The frequency Table in section 7.3 confirmed, however, that I have far more [f] than other variants. Furthermore, concerning every group, the preferable non-[f] variant is [p], followed by [v], with [b] being negligible. Consequently, it makes more sense to combine [v], [b] and [p] and regard this variable as comprising binary variants, [f] vs. non-[f].

Section 7.2 reviews the literature and presents the phonemic inventories of the superstrate language Hijazi Arabic, including both its Bedouin and Urban dialects, as well as the substrate languages Bengali, Hindi-Urdu and Tagalog. In Section 7.3, the results of the distributional analysis are reported. An analysis of migrant speakers' production of [f] is presented in Section 7.4 using four models: one that combines data from all speakers, and three further models – one for each of the L1 groups. These analyses will confirm whether each of these language groups has its own Arabic variable system for variable (f). Broadly speaking, the findings in the three models reveal that the three different L1 groups, each exhibit different patterns of use for (f) when speaking L2 Arabic, supporting an interlanguage hypothesis for migrants' Arabic varieties.

7.2 /f/ across HA and L1s

This section illustrates how /f/ is produced across the three substrate languages, i.e., Bengali, Hindi-Urdu and Tagalog. The differences between them in regard to L1 may influence their production of /f/ in L2. Hence, we have the following evidence: their Arabic variety can potentially be described as an interlanguage. This might manifest as inter-group variation that, for example, reveals the impact of their L1 on their production of /f/. Alternatively, if any of the external factors have no effect on the production of /f/ and if the results illustrate a trend that appears unusual in comparison to what I would expect in an interlanguage, which is inter-group variation, then migrant speakers' Arabic variety might be better described as a pidgin variety. In sub-section 7.2.1, I describe how /f/ is produced in HA. The production of /f/ across the three L1s is covered in subsections 7.2.2 to 7.2.4.

7.2.1 /f/ in Hijazi Arabic

While there are 26 consonant phonemes in UHA (Sieny 1972), there are 28 consonant phonemes in BHA (Il-Hazmy 1975) (see Tables 23 and 24 for both varieties). Only BHA has /ɵ/, /ð/ and /ðˤ/. Only UHA has /zˤ/. The consonant phoneme /f/, which is the focus of my study, exists in both varieties of Hijazi Arabic. As mentioned in Section 3.2, I have not been able to determine which BHA speakers or UHA speakers the migrant workers in my study engage with the most, or which of their two varieties they are most exposed to. Accordingly, my study's analysis focuses on /f/, a feature shared by both social groups and observed in both their varieties.

Table 23: UHA consonant phonemes (Adapted from Sieny 1972: 4).

Place of articulation → / Manner of articulation ↓	Labial	Alveolar	palatal	Velar	Pharyngeal	Glottal
Stop	b	t, tˤ d, dˤ		k, g		ʔ
Flap		r				
Affricative			ʒ			
Fricative	**f**	s, sˤ z, zˤ	ʃ	x, ɣ	ħ, ʕ	h
Nasals	m	n				
Lateral		l				
Semivowels	w		j			

Table 24: BHA consonant phonemes of (Adapted from Il-Hazmy 1975: 41).

Place of articulation → / Manner of articulation ↓	Labial	Labiodental	Dental	Interdental	Dento-alveolar	Emphatic	Plato-alveolar	Palatal	Velar	Pharyngeal	Glottal
Stop	b		t, d	dˤ		tˤ			k, g		ʔ
Affricative							dʒ				
Fricative	w	**f**		ɵ, ð	s, z	sˤ, ðˤ	ʃ	j	x, ɣ	ħ, ʕ	h
Nasals	m				n						
Liquid			l				r				

7.2.2 /f/ in Bengali

Bengali consists of 30 phonemic consonants including two semivowels which are /w/ and /j/ (Barman 2009; Kostić & Das 1972) (see Table 25). The bilabial plosive /pʰ/, is frequently pronounced as the labiodental fricative allophone [f] (Kostić & Das 1972).

Table 25: Bengali consonant phonemes (Adapted from Barman, 2009; Kostić & Das 1972).

Place of articulation → / **Manner of articulation ↓**	**Labial**	**Dental / alveolar**	**Retroflex**	**Palato-alveolar**	**Postalveolar**	**Velar**	**Glottal**
Stops	p, **pʰ** b, bʰ	t, tʰ d, dʰ	ʈ, ʈʰ ɖ, ɖʰ	ʃ		k, kʰ g, gʰ	
Affricative					ʧ, ʧʰ ʤ, ʤʰ		
Fricative		s					h
Nasals	m	n				ŋ	
Lateral		l					
Approximant			r		j	w	

7.2.3 /f/ in Hindi-Urdu

There are 40 consonants in Hindi (Agnihorti 2006: 244). They are shown in Table 26. Of these, 35 sounds are inherited from Sanskrit whereas the remaining five sounds, i.e., [q, x, ɤ, f, z], are borrowed from other languages. The labio-dential fricative /f/exists in Hindi (Koul 2008: 17).

Table 26: Hindi consonant phonemes. (Adapted from Agnihorti 2006: 246; Koul 2008: 12)

Place of articulation → / **Manner of articulation ↓**	**Labial**	**Labio-dental**	**Dental /alveolar**	**Retroflex**	**Alveopalatal**	**Velar**	**Glottal**
Stops	p, p^h b, b^h		t, t^h d, d^h	ʈ, $ʈ^h$ ɖ, $ɖ^h$		k, k^h g, g^h, q	
Trill			r				
Flap				ɽ, $ɽ^h$			
Affricative					tʃ ʤ, $ʤ^h$		
Fricative		**f**	s, z	š	ʃ, ʒ	x, ɤ	h
Nasals	m		n	ɳ	ɲ	ŋ	
Lateral			l				
Semi-vowels		v			j		

Urdu consists of 38 consonants, including the labiodental fricative /f/ (Mirdehghan 2010: 11; Raza *et al.* 2009: 43). See Table 27 below.

Table 27: Urdu consonant phonemes (Adapted from Raza et al. 2009: 43).

Place of articulation → / **Manner of articulation ↓**	**Labial**	**Labio-dental**	**Dentals/ alveolars**	**Retroflex**	**Alveolar**	**Alveopalatal**	**Velar**	**Uvular**	**Glottal**
Stops	p, p^h b, b^h		ʈ, $ʈ^h$ ɖ, $ɖ^h$		t, t^h d, d^h		k, k^h g, g^h	q	ʔ
Trill					r				

Flap				ɾ					
Affricative						tʃ, tʃʰ ʤ, ʤʰ			
Fricative		**f**, v			s, z	ʃ, ʒ	x, ɤ		h
Nasals	m				n		ŋ		
Lateral					l				
Approximant						j			

7.2.4 /f/ in Tagalog

Tagalog has sixteen phonemic consonants which occur in words of Tagalog origin (Schachter & Reid 2009: 834). They are shown in Table 28. There are other phonemes that occur only in loanwords which are regularly heard in the speech of many Tagalog speakers, especially those who have learnt English in higher education. For instance, the phoneme /f/ is used primarily in proper names with foreign origins, such as */flu:ra/* 'Flora' and in certain English and Spanish borrowings like */kafiti:rja/* 'cafeteria' (Schachter & Reid 2009: 836). Furthermore, some speakers of non-Manila dialects of Tagalog who are not educated in the capital Manila might produce /p/ in a relatively consistent way in cases where educated Tagalog speakers in Manila produce /f/ (Schachter & Otanes 1983: 22).

Table 28: Tagalog consonant phonemes (Adapted from Schachter & Reid 2009: 835).

Place of articulation → **Manner of articulation** ↓	**Labial**	**Dental**	**Alveolar**	**Palatal**	**Velar**	**Glottal**
Stop	p, b	t, d			k, g	ʔ
Trill			r			
Affricative						
Fricative			s			h
Nasal	m	n			ŋ	
Lateral			l			
Semivowels	w			j		

7.2.5 Summary

To summarise, Table 29 below reveals that the phoneme /f/ exists in HA and Hindi-Urdu, while in Bengali, /p^h/ is often represented as [f]. However, it is not found in the Tagalog phonemic inventory, although it appears in specific words that are borrowed from English and Spanish in the speech of those with higher education in English. Thus, these differences may demonstrate variation between the migrant workers' groups when they produce /f/ in Arabic, indicating that their Arabic varieties are interlanguage.

Table 29: A cross-linguistic comparison of /f/ in the superstrate and three substrate languages.

Languages	**/f/**
HA	/f/
Bengali	/p^h/ is often represented as [f]
Hindi-Urdu	/f/
Tagalog	Ø • Only in certain English and Spanish borrowings. • In the speech of those with higher education in English.

7.3 Overall distribution of /f/

The use of the variable /f/ by migrant workers in Hijaz; Madinah, is examined in this section. Table 30 reveals that the [f] is the preferred variant across all speaker groups, particularly those with Hindi-Urdu and Bengali as their L1. L1 Tagalog use [f] the least of all groups, but it is still the majority variant. Regarding the non-[f]-variants, for all three groups, [p] is the most frequent one, followed by [v] then [b] which are negligible across groups. Essentially, L1 Tagalog speakers have more than four times the rate of [p] compared to L1 Bengali and L1 Hindi-Urdu speakers. This high frequency of [p] by L1 Tagalog speakers is at the expense of [f] rather than other co-variants. Rates for [v] and [b] are quite similar across the three groups. Thus, it is basically with [f] and [p] that the L1 Tagalog speakers pattern differently in relation to the Bengali and Hindi-Urdu speakers.

Table 30: The frequency of four variants of /f/

L1	/f/ variants								
	[f]		[v]		[b]		[p]		Total
	%	(N)	%	(N)	%	(N)	%	(N)	
Bengali	88.1 %	(4718)	3%	(161)	0.1%	(4)	8.8%	(474)	5357
Hindi-Urdu	90%	(4793)	2.8%	(150)	0.1%	(6)	7.1%	(375)	5324
Tagalog	53.8%	(1898)	1.5%	(52)	0.1%	(2)	44.6%	(1573)	3525
Total	**80.3%**	**(11409)**	**2.6%**	**(363)**	**0.1%**	**(12)**	**17%**	**(2422)**	**14206**

7.4 Results of /f/ Regression Analysis

Based on the results in Table 30, the variation across the three language groups indicate their Arabic varieties may well be described as an interlanguage. More detailed findings related to the production of [f] are presented in this section. The first sub-section consists of all the sampled migrant workers. In this model, the L1 of the migrants is included as a predictor for [f]-realization in the model. If L1 and other external factors exhibit an effect on the production of [f], the migrants' Arabic variety could possibly be an interlanguage. Sub-sections 7.4.2 to 7.4.4 explore [f]-variaiton across the three L1 groups. In each model, only two independent variables are examined i.e., fewer than in models in the previous two chapters. The last versions of the stepwise regression models with the lowest AIC include numerous independent variables. However, the chosen variables in the regression models are the only orthogonal predictors, and attempting to include any more variables causes multicollinearity and non-orthogonal variables. A table that displays the mixed-effect logistic regression model is provided in each sub-section.

Given that /f/ is a categorical variable in HA and there are no variation studies as regards this variable, I have created a number of internal variables like 'location of [f]' and 'the following segment' to test if they can condition the production of /f/ by migrant workers in Madinah. The predictor levels with respect to the location of [f] are onset, nucleus and coda, whereas the predictor levels of the following segment are consonant, vowel and pause. These two variables have not been included in any of the following models for the reasons that are mentioned earlier.

The independent variables tested in the following models are mostly categorical. In contrast, five predictors, namely motivation, attitudes toward HA, Madinah and its inhabitants, language anxiety, age and LoR, are treated as continuous variables because they lack distinct predictor levels.

7.4.1 Across all the speakers

In this large general model, I combined all migrant speakers, who are from different L1 backgrounds, and treated them as a single social group. The model consists of eight factors: first language, exposure to informal Arabic, amount of Arabic and L1 used, motivation, formal Arabic instruction in the country of origin, sex, attitude, and language anxiety. Only five out of these variables significantly affect migrant speakers' use of /f/. Table 31 shows that in comparison to the reference group, set as Tagalog speakers, L1 Bengali speakers and L1 Hindi-Urdu speakers favour [f]. These results confirm the distributional results. These consistent results across both analyses indicate a clear preference for [f] in the Arabic varieties of the migrant workers, particularly among L1 Hindi-Urdu and L1 Bengali speakers, supporting the interlanguage hypothesis. Regarding 'exposure to informal Arabic' and comparing the reference which is the speakers who reported that they are not exposed to informal Arabic, there is no significant difference between these speakers and those who reported that they are only exposed to the Quran. However, those who reported that they listen to the Quran and are exposed to Arabic via TV and radio tend to favour [f] in comparison to the reference. Additionally, motivation plays a significant role in the production of [f]; migrant workers who are highly motivated to learn Arabic, use [f] more than those who are lacking motivation. Whereas male speakers favour [f], female speakers disfavour it. Attitude has a significant effect on the production of [f] too; those with more positive attitudes towards Madinah, living there, its inhabitants and their variety, have significantly higher rates of [f]. In contrast, the amount of Arabic and L1 used, formal Arabic instruction in the country of origin and language anxiety did not significantly affect [f] use.

Table 31: Mixed-effects results showing the contribution of independent variables on the production of [f] across all speakers in the corpus.

Predictor	Estimate	Std. Error	Z value	Pr (>|z|)
Intercept	-9.5826	2.6927	-3.559	0.000373***
First language Tagalog (reference)				
Bengali	2.1736	0.6426	3.377	0.000732***
Hindi-Urdu	2.7134	0.6979	3.888	0.000101***
Exposure to informal Arabic No exposure (reference)				
Quran only	0.3372	0.5868	0.575	0.565545
Quran and more	1.6263	0.6460	2.517	0.011821*
Amount of Arabic and L1 used Those who generally used Arabic (reference)				
Those who mostly used their L1	-0.2757	0.4626	-0.596	0.551111
Motivation	0.6659	0.2723	2.445	0.014470*
Formal instruction in their country of origin Mostly no (reference)				
Mostly yes	0.4039	0.5120	0.789	0.430143
Sex				

Female (reference)				
Male	2.5868	0.4558	5.675	1.39e-08***
Attitudes toward HA, Madinah and its inhabitants	0.6855	0.3373	2.033	0.042094*
Language anxiety	-0.0392	0.2181	-0.180	0.857329
Random effect: speaker Variance: 1.164 Std. Dev: 1.079 **Random effect: word** Variance: 5.979 Std. Dev: 2.445				

7.4.2 L1 Bengali speakers

The mixed-effect logistic regression model in Table 32 tests the influence of only two predictors, specifically motivation and gender, on the production of [f] by L1 Bengali speakers. Table 32 reveals that both of these fixed factors have a significant effect on the production of [f]. Motivation has a significantly positive effect on the production of /f/. L1 Bengali speakers who are highly motivated to learn Arabic, employ [f] more than those who are lacking motivation. In addition, while males tend to favour [f], females disfavour it.

Table 32: Mixed-effects results for the effect of selected independent variables on the production of [f] by L1 Bengali speakers.

Predictor	Estimates	Std. Error	Z value	Pr (>|z|)
Intercept	-5.2014	1.7302	-3.006	0.00264**
Motivation	1.1367	0.2608	4.359	1.31e-05***
Sex Female (reference)				
Male	1.7996	0.3461	5.199	2.00e-07***
Random effect: speaker Variance: 0.2418 Std. Dev: 0.4918 **Random effect: word** Variance: 5.8349 Std. Dev: 2.4156				

7.4.3 L1 Hindi-Urdu speakers

For L1 Hindi-Urdu speakers, I tested the impact of two predictors on their production of [f]: the amount of Arabic and L1 used, together with length of residence. The mixed-effect logistic regression results in Table 33 demonstrate that only the amount of Arabic and L1 used has a significant effect. L1 Hindi-Urdu speakers who reported that the amount of L1 that they use is generally equal to or more than the amount of Arabic, disfavour [f]. By contrast, those who reported that the amount of Arabic employed is more than the amount of L1 utilised, favour [v].

Table 33: Mixed-effects results for the effect of the independent variables on the production of [f] by L1 Hindi-Urdu speakers.

Predictor	Estimate	Std. Error	Z value	Pr (>|z|)
Intercept	4.12759	0.93137	4.432	9.35e-06***
Amount of Arabic and L1 used. Those who generally used Arabic (reference)				
Those who mostly used their L1	-1.95163	0.843	-2.315	0.0206*
Length of residence	0.07957	0.04570	1.741	0.0816
Random effect: speaker Variance: 1.094 Std. Dev: 1.046 **Random effect: word** Variance: 10.647 Std. Dev: 3.263				

7.4.4 L1 Tagalog speakers

The focus in this sub-section is on the findings of the production of [f] by L1 Tagalog speakers. Two of the predictors, namely amount of Arabic and L1 used and age, are examined in this model. Only age plays a significant role in the production of [f], as demonstrated in Table 34. The correlation between age and the production of [f] is negative; as the age of the L1 Tagalog speakers increases, their production of [f] decreases.

Table 34: Mixed-effects results for the effect of the independent variables on the production of [f] by L1 Tagalog speakers.

Predictor	Estimate	Std. Error	Z value	Pr (>|z|)
Intercept	8.96640	3.91756	2.289	0.0221*
Amount of Arabic and L1 used. Those who generally used Arabic (reference)				
Those who mostly used their L1	-1.22343	1.44187	-0.848	0.3962
age	-0.20146	0.09479	-2.125	0.0336*
Random effect: speaker Variance: 3.922 Std. Dev: 1.980 **Random effect: word** Variance: 4.658 Std. Dev: 2.158				

7.5 Summary

This chapter examined the production of /f/ in the L2 Arabic spoken by migrant workers in Madinah who speak Bengali, Hindi-Urdu and Tagalog as their L1. The aim was to add more evidence to establish – in conjunction with results presented in other chapters – whether their Arabic varieties are interlanguage or pidgin. If external factors influence migrant workers' production of /f/, their Arabic varieties might be better described as interlanguage. However, if they have no influence and the results show an unusual trend, deviating from what I would expect in an interlanguage, particularly inter-group variation, then their varieties may be described as pidgin.

Table 35 demonstrates that the findings in the large general model (see 7.4.1) and the L1 Bengali speakers' model (see 7.4.2) tend to indicate that the migrants' Arabic varieties are interlanguage. Most of the fixed factors in the former model and all the external factors in the latter have a significant effect on the production of [f]. These results are generally in line with L2 studies which show variation, such as the research conducted by Abu-Rabia and Kehat 2004; Adamson and Regan 1991; Drummond 2011; Flege, Frieda and Nozawa 1997; Major 2004; Mougeon and Rehner 2001; Mougeon, Rehner and Nadasdi 2004 and Thompson 1991. Moreover, the impact of the migrants' L1 on the production of [f] in the large general model supports previous L2 studies (e.g. Flege, Frieda & Nozawa 1997; Mougeon & Rehner 2001 and Mougeon, Rehner & Nadasdi 2004), and it also supports the results in the qualitative studies of GPA (Hobrom 1996; Naess 2008). Tagalog speakers may exhibit a different pattern in the rate of [f] compared to other L1 groups, influenced by the absence of /f/ in their L1 inventory. They produce [p] with high frequency, and this occurs at the expense of [f] rather than other co-variants.

The individual models I presented concerning L1 Hindi-Urdu speakers and L1 Tagalog speakers include only two independent variables each (for the reasons mentioned in Seciton 7.4). The results in these two models suggest that the Arabic varieties of the migrants are interlanguage too. While one of the two predictors in each model presents a significant effect on the production of /f/, the non-significant effect of the other variable does not necessarily indicate a pidgin classification. It is important to note that null results should not be interpreted as supporting a null hypothesis. Likewise, the predictor 'amount of Arabic and L1 used' has a significant effect on the use of [f] by L1 Hindi-Urdu speakers and a non-significant effect by L1 Tagalog speakers. This inter-group variation suggests that the Arabic varieties of the migrants are interlanguage.

Table 35: Summary of the predictors that reveal a significant effect on the production of [f] supporting the interlanguage hypothesis.

Model	Predictor	Significant effect
All the speakers in the corpus	First language	√
	Exposure to informal Arabic	√
	Amount of Arabic and L1 used	
	Motivation	√
	Formal instruction in their country of origin	
	Sex	√
	Attitudes toward HA, Madinah and its inhabitants	√
	Language anxiety	
L1 Bengali speakers	Motivation	√
	Sex	√
L1 Hindi-Urdu speakers	Amount of Arabic and L1 used	√
	Length of residence	
L1 Tagalog speakers	Amount of Arabic and L1 used	
	Age	√[20]

I will return to the results of the three dependent variables, i.e., the definite article, the conjunction marker and /f/ in more detail in the subsequent chapter, which comprises the discussion and conclusion.

[20] Age has a significant negative effect in relation to the L1 Tagalog speakers' model.

Chapter 8: Discussion and Conclusion

8.1 Introduction

My study set out to investigate whether the L2 Arabic varieties spoken by migrant workers in Madinah, Saudi Arabia, are pidgin or interlanguage varieties. I hypothesise that the migrant workers' Arabic L2 can be considered a pidgin variety if the following apply: they have reduced verbal and nominal systems; they typically delete the morphosyntactic features and substitute /f/ with other variants; contextual predictors negligibly impact their L2 production, and their Arabic L2 is simplified compared to Hijazi Arabic (HA), the local variety of Arabic. However, if their use of the morphological features follows the usage patterns of L1 HA speakers and their use of morphosyntactic and phological features is significantly conditioned by contextual factors, the migrant workers' Arabic L2 can be considered interlanguage varieties.

The qualitative analysis of the morphological features reveals that the Arabic speech of the three migrants, who speak Bengali, Hindi-Urdu and Tagalog as their L1, lacks both the subject-verb agreement and the noun-adjective-agreement. This result may possibly support the pidgin hypothesis. Conversely, my quantitative analyses including all the participants in my study have established that the definite article /ʔal-/ 'the' and the coordinating conjunction marker /wa/ 'and' are predominantly omitted by migrant workers in Madinah. This demonstrates simplicity compared to the HA variety. Not speaking at a level approaching L1 proficiency, does not necessarily imply speaking a pidgin. It is evident from Table 36 that although there are a considerable number of null results, a significant impact on the production of the definite article and conjunction marker is confirmed by a restricted set of independent variables. These significant factors can show variations across various language groups and linguistic variables. Concerning the production of /f/, migrants typically produce it when they speak Arabic. Table 36 also reveals that most contextual variables significantly affect the migrants' production of /f/. Thus, the significant impact of social predictors across various linguistic variables and speaker groups suggests that the Arabic varieties of migrant workers in Madinah can be better described as an interlanguage. Thus, the morphological analysis proposes that the Arabic varieties of the migrant workers might be a pidgin, while the morphosyntactic and phonological analyses indicate that they

could be an interlanguage. This may indicate the Arabic varieties of the migrant workers studied in this thesis are located on a continuum, where the higher end may be closer to an interlanguage.

Table 36: Summary of the predictors that reveal a significant effect on the use of HA features supporting the interlanguage hypothesis.

Feature	Model	Predictor	Significant effect
The definite article	**All the speakers in the corpus**	Lexical semantics	
		First language	
		The identity of the migrant	
		LoR	
		Educational level in L1	
	L1 Bengali speakers	LoR	√
		Language anxiety	
		Motivation	
	L1 Hindi-Urdu speakers	LoR	
		Language anxiety	
		Sex	
		Proficiency: Arabic level of the migrants throughout the interview	√
	L1 Tagalog speakers	LoR	
		Language anxiety	
		Motivation	
The coordinating conjunction marker	**All the speakers in the corpus**	CONJ in different context	√
		First language	
		Exposure to informal Arabic	
		Amount of Arabic and L1 used	√
		Motivation	

		Formal instruction in their country of origin	
		Sex	
	L1 Bengali speakers	Amount of Arabic and L1 used	
		Age	√[21]
		Sex	
	L1 Hindi-Urdu speakers	Sex	√
		Formal Arabic instruction in Hijaz	
		Formal Arabic instruction in the country of origin	
		The identity of the migrant	
		Age	
	L1 Tagalog speakers	Length of residence	
		Age	
		Amount of Arabic and L1 used	√
		Educational level in L1	
/f/	**All the speakers in the corpus**	First language	√
		Exposure to informal Arabic	√
		Amount of Arabic and L1 used	
		Motivation	√
		Formal instruction in their country of origin	
		Sex	√
		Attitude	√
		Language anxiety	
	L1 Bengali speakers	Motivation	√
		Sex	√

[21] Age has a significant negative effect in relation to the L1 Bengali speakers' model.

	L1 Hindi-Urdu speakers	Amount of Arabic and L1 used	√
		Length of residence	
	L1 Tagalog speakers	Amount of Arabic and L1 used	
		Age	√[22]

In the preceding paragraph, I provided a summary of the results of my study and its interpretation. The details will be further investigated in this particular chapter. I will explore the evidence for interlanguage status, whilst also considering the evidence supporting possible pidgin status. I will assess the conflicting evidence, examine the differential impact of social predictors across various linguistic variables and speaker groups, and discuss the implications of my research for GPA studies and contact languages more generally. I will also illustrate the opposite approach in the results of the morphological features on the one hand and the morphosyntactic and phonological features on the other. By addressing these points, I aim to make an original contribution to knowledge in the field.

First, I will discuss the morphological features. Then, I will address the first three research questions that I outlined in Section 1.3. These questions are related to the impact of the internal factors and the L1 of the migrant workers on their use of HA features. Section 8.3 answers the last research question which examines the influence of social factors on the migrants' use of HA features. The last section of this chapter is the conclusion of my study.

The Morphological Features

As Sebba (1997: 39) indicated, pidgins have a minimal or non-existent number of inflections, such as those denoting, number, gender, tense, case or grammatical agreement. The qualitative analysis of the three old speakers reveal that verbs have reduced inflections, similar to other GPA studies (Albaqawi 2020; Almoaily 2012; Alsusut 2022; Bakir 2010; Hobrom 1996). As an alternative, the migrant workers developed specific strategies to express the verbal form (for more details, see Chapter 4). Although there is variation in the L1 of the migrants in regard to using the verbal agreement, the verbs principally do not agree with the nouns in gender, number and person.

[22] Age has a significant negative effect in relation to the L1 Tagalog speakers' model.

Likewise, the qualitative analysis demonstrates a lack of inflections in the nominal agreement. Not unlike the GPA studies conducted by Albaqawi (2020: 252), Almoaily (2012: 148) and Alsusut (2022: 335), the adjectives did not agree with the noun. In actual fact, I expected to find the noun-adjective agreement in the Arabic speech of the L1 Hindi-Urdu speaker and the L1 Tagalog speaker, on account of the impact of their L1. In spite of this, the results illustrate that the simplification is not only in the speech of the L1 Bengali speaker whose L1 does not in include this sort of agreement, but also in the speech of the other two speakers. With regard to the number marking on nouns, there is a scarcity of data generally, which makes it difficult to decide whether the Arabic varieties of the migrants is a pidgin language or an interlanguage (for more details see 4.4).

Acquiring the HA subject-verb agreement and noun-adjective-agreement appears to be challenging for the migrant workers in my study. This type of simplification implies that their Arabic varieties could be described as a pidgin language. We will ascertain in the following if the answers to the research questions pertaining to the morphosyntactic and phonological features support the results in the qualitative analysis or demonstrate a different approach and support the interlanguage hypothesis.

8.2 The Influence of Internal Factors on the Migrants' Use of HA Features

My research questions were formulated to establish whether the Arabic varieties of the migrants are pidgin or interlanguage. Below, I will address each research question in turn to provide evidence in support of my conclusion about that status of the examined varieties.

1. Do migrant workers use features of HA categorically or variably? If variably, what are the linguistic constraints?

While the definite article, the coordinating conjunction marker and the phoneme /f/ are used categorically in HA (see Il-Hazmy 1975 & Omar 1975), migrant workers in Madinah use these features variably. They delete the definite article and the conjunction marker more than they use them (see Tables 10 & 17), but they produce more [f]- than non-[f]-variants (see Table 30). To establish whether these patterns are indicative of migrants' L2 Arabic varieties being pidgin or

interlanguage varieties, I explored the effect of multiple linguistic constraints on variant choice. Based on my reading of the literature, I hypothesised that if there is variation across various linguistic variables and across various language groups in the Arabic varieties of the migrants, their Arabic L2 varieties may be classed as interlanguage varieties. However, no variation does not necessarily prove the pidgin hypothesis, it is basically just one of the possible explanations for the results.

In the analysis of the definite article, 'lexical semantics' was included as a linguistic factor. This variable was coded as 'day/part of the day', 'formulaic utterances', 'miscellaneous', 'clothes' and 'place/city/country'. 'Lexical semantics' did not constrain migrant workers' use of the definite article in the model that included all sampled migrant workers. Although the use of the definite article was greater in formulaic utterances than non-formulaic constructions, the difference was not statistically significant (see Table 11). My quantitative findings do not therefore provide compelling evidence to confirm the qualitative findings reported by Albakrawi (2012) and Hobrom (1996), i.e., that discovered migrant workers typically did not use the definite article, with the exception of a few formulaic statements. The non-significant effect of this predictor on the use of the definite article by the migrants cannot be interpreted as clearly favoring the pidgin account.

I incorporated 'linguistic contexts' as an internal variable in the analysis of the coordinating conjunction marker. The predictor levels of this variable determine whether the conjunction marker is applied to join 'numbers', 'two phrases', 'two clauses' or 'a list containing more than two phrases'. 'Linguistic contexts' significantly impacted the production of the conjunction marker in the model that includes all the speakers in the corpus (see Table 18), with 'numbers' and 'two phrases' favouring and 'list of more than two phrases' disfavouring the use of +CONJ. This result suggests that the Arabic varieties of the migrants are interlanguage varieties. Close scrutiny of the results shows that the migrant workers predominantly use the conjunction marker in the context of 'numbers'. This could be a frequency effect[23]. Migrants use +CONJ between numbers because they may hear it frequently, seeing as it is related to discussing their salary and daily activities such

[23] My research has not examined the impact of HA on the speech of the migrants owing to time limitations and as it is not among the aims of the study. Research conducted in the future will have an opportunity to examine that particular aspect.

as buying things from supermarkets and shops. Likewise, numbers are also related to time. In HA, the conjunction marker ‘and,’ which means ‘past’ is used to report time. For example,

(114) assaʕa xamsa **wu** nuss
O’clock.SG.F five.SG.F **and** half
‘It is half past five’
(Sieny 1972: 118).

I speculate that migrants use +CONJ between numbers owing to its frequent association with the timing of their shiftwork, specifically when indicating their work starts and finishes. For this reason, it might be easy for them to memorise and learn. Therefore, in this sense, language-internal conditioning could be related to the nature in which the migrant workers acquire and use L2 Arabic. Almoaily (2012: 92) indicated that the general tendency for the migrant workers in his GPA study was to omit conjunctions; but he referenced several examples of particular conjunctions being used in his data. One of these examples is the use of +CONJ between numbers (see example 115 below). The finding in my study supports his example.

(115) Wahid **wu** nuss fi il-li:l
One **and** half in DEF-night
‘One and a half at night’
(Almoaily 2012: 92).

On the other hand, migrants use +CONJ in the context of a ‘list of more than two phrases’ the least. This pattern could be ascribed to the absence of sufficient exposure. Migrants may rarely encounter the conjunction marker in an environment comprising a ‘list of more than two phrases’. This is especially true if most of the language they hear at work is L2 Arabic spoken by co-workers. This assertion could be verified by more research that explores the language environments that migrants are likely to experience and how those contexts might affect the migrants' use of the language.

Regarding the impact of the two internal factors on the migrants' use of HA features, then, the results provide conflicting results for the definite article and the conjunction marker (see Table 37). With the former, I found a no significant effect. However, it is quite challenging to interpret this finding as an indication that Arabic varieties of migrants are pidgin, as a null result does not confirm a null hypothesis. With the latter, I found significant effect, suggesting that their varieties are interlanguage varieties. Table 37 does not account for the impact of any internal variable on the production of /f/ because only orthogonal predictors were tested (for more details, see Section 7.4)

Table 37: The impact of the internal variables on DEF and CONJ by all the speakers in the corpus

Internal variable	Significant effect	Non-significant effect
Lexical semantics		DEF
Linguistic contexts	CONJ	

Additional evidence, then, is needed to confirm the status of migrant workers' L2 Arabic. Below, I turn to the second research question to do so.

2. Do migrant workers with different heritage language backgrounds, i.e., Bengali, Hindi-Urdu, and Tagalog, share the same underlying grammar of variation regarding the selected linguistic features?

In the context of the L1 of migrants, Bengali, Hindi-Urdu and Tagalog, there is an argument concerning abstract structure on one hand and a question of overt marking on the other hand, particularly concerning the use of the definite article (see 5.2.2 - 5.2.4). While Bengali and Hindi-Urdu lack over definiteness marker, they can convey it indirectly. However, definiteness is controversial in Tagalog. All these languages exhibit characteristics that may make them suitable to be considered and function as DP languages.

My findings show that the migrant speakers of these languages in my sample mostly delete the Arabic definite article (see Table 10). This finding is consistent with findings reported in GPA

studies conducted by Al-Azraqi (2011), Albakrawi (2012), Albaqawi (2020), Almoaily (2012), Alzubeiry (2015) and Hobrom (1996). Furthermore, L1 does not significantly influence migrant workers' use of the definite article (see Table 11). This finding mirrors the results in the GPA studies completed by Albaqawi (2020) and Almoaily (2012). Conceivably, negative transfer is the reason why the migrant workers' use of the definite article is not conditioned by their L1. It is possible that NP languages act as the L1 for migrants. Likewise, the HA definite article is a bound morpheme that could present a challenge for migrant workers who are seeking to acquire it. This result appears to be consistent with other research that has determined that, regardless of their L1, L2 learners tend to produce free morphemes earlier and use them with greater precision than bound morphemes (Zobl & Liceras 1994). Thus, the combined group of migrant workers use a simplified version of HA characterized by at least one shared characteristic: a regular omission of the definite article. This characteristic is also by GPA speakers. However, these findings do not indicate that the migrants' Arabic language use can be classified as a pidgin.

While the use of the coordinating conjunction marker differs in migrants' L1s (see 6.2.2 - 6.2.4), my results indicate that all three individual L1 groups mostly delete the Arabic conjunction marker (see Table 17). This finding is in line with the findings of Naess (2008) who found that migrants in Buraimi, Oman and Al-Ein, UAE mostly omitted the conjunction marker. In addition, my migrants' L1s do not significantly influence their use of the conjunction marker (see Table 18). Although the conjunction marker is a free morpheme in HA, it remains challenging for the migrants in my study to acquire it. This might be due to the migrants' L1, seeing as using the coordinating conjunction is not obligatory in all the linguistic contexts (see Table 16). Thus, the non-significant effect of the migrants' L1 on their use of the conjunction marker may not indicate that their Arabic varieties are pidgin.

Several migrant workers in my study had acquired a degree of Arabic and Quran knowledge in their country of origin prior to arriving in Saudi Arabia. This might contradict the emergence of a pidgin to some extent given that pidgins generally develop among people who do not share a common language and need to communicate. So it is expected that a number of the participants in my study could understand and communicate in Arabic when they first arrived in Madinah and their Arabic varieties may be interlanguage not pidgin. However, the findings that reveal that they

primarily deleted the definite article and the coordinating conjunction marker imply that what they had learned has not led them to use these features categorically, which is how L1 HA speakers use them. Moreover, most of the participants reported that when they arrived in Saudi Arabia 'they could not speak Arabic at all'. The non-significant impact of migrants' L1 on the use of the definite article and the coordinating conjunction marker should be interpreted with caution, as null results do not confirm a pidgin classification.

Further proof is required to validate the status of migrant workers' Arabic L2. To achieve this, I will examine the third research question below.

3. Do migrant workers with different heritage language backgrounds use their own distinct system of Arabic L2 which might be affected by their L1? If yes, what are these distinguishing characteristics?

The three individual language groups have their own distinct systems of /f/-usage in Arabic; their respective L1s exert a significant effect on their production of /f/ (see Table 31). This finding is consistent with other L2 research which, for example, determined the significant effect of L1 on the L2 use of the negative particle in French (Mougeon, Rehner & Nadasdi 2004; Mougeon & Rehner 2001). In my study, L1 Bengali speakers and L1 Hindi-Urdu speakers produced /f/ more than the L1 Tagalog speakers. L1 Tagalog speakers do not have /f/ in their L1 phonemic inventory, although they have /p/ (see 7.2.4). This may explain why they produce /f/ the least and /p/ the most compared to the other two language groups (see Table 30). The variation and the influence of the migrants' L1 in the /f/ model indicate that the migrants' Arabic varieties are distinct interlanguage varieties.

My findings suggest that migrant workers may demonstrate a greater degree of success in acquiring the phonological HA feature compared to the morphosyntactic HA features I investigated. I should approach this conclusion with a degree of caution owing to the limited comparison of variables in this study. It is crucial to distinguish between learning to produce the sounds and learning phonological patterns. My study primarily focused on the former. Future research could explore the migrants' proficiency in learning and applying phonological patterns to

provide a more comprehensive understanding of their language acquisition skills. Table 38 reveals that whereas my research participants produce [f] 80.3% of the time, they use the conjunction marker 14% and the definite article 6% of the time. Specifically, given that /f/ is a non-marked sound, a common and shared phoneme across languages (Aljutaily 2018) , the effect of L1 phonemes on L2 phonemes is evident. Regarding the morphosyntactic features, the coordinating conjunction marker was used with higher frequencies than the definite article. This may be because in HA, the former is a free morpheme while the latter is a bound morpheme, with the former known to be acquired more easily than the latter. Notwithstanding that the migrant workers' occasional use morphosyntactic features, their variety is not identical to the L1 HA variety.

Table 38: The frequency of using HA features.

Feature	%
+DEF	6%
+CONJ	14%
[f]	80.3%

Thus, to determine the status of migrant workers' variety of Arabic L2, Table 39 illustrates the impact of migrant workers' L1 on HA features. The non-significant effect of this variable on migrants' use of the definite article and the coordinating conjunction marker does not necessarily indicate a pidgin classification. It is essential to note that null results should not be interpreted as supporting a null hypothesis. However, the variation across various linguistic variables and the impact of the migrants' L1 in the use of /f/ suggest an interlanguage classification. Below, I discuss evidence for the effect of social predictors on observed patterns of L2 Arabic variation in an attempt to determine the Arabic status of the migrants.

Table 39: The impact of the migrants' L1 on their use of HA features by all the speakers in the corpus

Predictor	Significant effect	Non-significant effect
L1		DEF
		CONJ
	/f/	

8.3 The Influence of Social Factors on the Migrants' Use of HA Features

This section focuses on the fourth research question.

4. Do social factors impact migrant workers' variable use of selected HA features?

The results of my study confirm that there are 1) four social variables that show significant variation across various language groups, 2) seven social variables that show significant variation across various linguistic variables, 3) seven social variables that neither show significant variation across various language groups nor various linguistic variables. These three categorisations have enabled me to discover the status of the Arabic varieties of migrant workers in Madinah, and whether they are a pidgin or an interlanguage. I will begin by exploring the consistent effect of the four predictors in the first category.

8.3.1 Social Variables Show Significant Variation Across Various Language Groups.

In my research, four social factors - namely length of residence, sex, age and amount of Arabic and L1 used, - illustrate significant inter-group variation (see Table 40). This is when a social variable has a significant effect on the use of a linguistic variable in one language group's model but not that of others. For example, Table 40 reveals that sex significantly affects the use of the coordinating conjunction in the model that includes L1 Hindi-Urdu speakers. However, it did not play a significant role in the other language groups' models, i.e., across all the speakers and L1 Bengali speakers for the same variable. The significant impact of social predictors across speaker groups suggests that the Arabic varieties of migrant workers in Madinah can be better described as an interlanguage.

Table 40: Social variables show significant variation across various language groups.

Social variables	Dependent variable	Significant effect	Non-significant effect
		Model	Model
LoR	DEF	L1 Bengali speakers	all the speakers in the corpus
			L1 Hindi-Urdu speakers
			L1 Tagalog speakers
Sex	CONJ	L1 Hindi -Urdu speakers	all the speakers in the corpus
			L1 Bengali speakers
Age	CONJ	L1 Bengali speakers[24]	L1 Hindi-Urdu speakers
			L1 Tagalog speakers
Amount of Arabic and L1 used	CONJ	all the speakers in the corpus	L1 Bengali speakers
		L1 Tagalog speakers	
	/f/	L1 Hindi-Urdu speakers	all the speakers in the corpus
			L1 Tagalog speakers

In my study, length of residence was found to have a significant effect on L1 Bengali speakers' use of the definite article. Those who resided in Hijaz for a long time used +DEF more than those who stayed for a short time. This finding is in line with the L2 studies conducted by Drummond (2011), Regan (2013b) and Flege, Bohn and Jang (1997) and the GPA study undertaken by Albaqawi (2020), which reveal the effect of this predictor on the L2 acquisition of the English glottal variation in /t/, French *ne* deletion, English vowels and Arabic definite article. As the LoR of the participants in these studies increases, their acquisition of the local variants increases too. However, no difference was discovered between those with short versus long periods of residence in Hijaz in the other three models pertaining to the definite article which include all the speakers in the corpus, L1 Hindi-Urdu speakers and L1 Tagalog speakers. The reason that LoR only has an effect on the production of the definite article by L1 Bengali speakers and not on the other three models might be due to the number of participants: the number of L1 Bengali speakers who have lived in Hijaz for over 10 years is more than the number of L1 Hindi-Urdu speakers and L1 Tagalog

[24] Age has a significant negative effect in relation to the L1 Bengali speakers' model.

speakers who have stayed for the same period (see Table 1). The significant effect of LoR on the use of DEF in the L1 Bengali speakers' model suggests that their Arabic variety is interlanguage.

An important finding was the significant effect of sex in the model that tested its effect on L1 Hindi-Urdu speakers' use of the coordinating conjunction marker. Male migrants used HA variants more than females[25]. A possible explanation for this might be the impact of length of residence, which is not examined in my model. The average length of residence for male L1 Hindi-Urdu speakers is longer than that for female L1 Hindi-Urdu speakers. While the average length of residence for males 15.8 years, the average length of residence for females 5.2 years. The effect of sex in my study differs from the L2 studies completed by researchers such as Adamson and Regan (1991), Major (2004) and Thompson (1991), who argued that female L2 English speakers were more accurate native-like in pronunciation, and also used [ŋ] and other prestigious forms more than males. I do not have information pertaining to LoR in these studies. Similarly, this independent variable has a different effect on the use of the conjunction marker in comparison to what Albaqawi (2020) discovered in her GPA research. The discrepancy in my findings is fairly unexpected, albeit the social environments are generally comparable to one another. Although Albaqawi investigated the impact of LoR and compared the use of the conjunction markers by female migrants in her study with male migrants documented in Almoaily's (2012) study, establishing that females tend to use the conjunctions significantly more than males. My investigation uncovered the opposite: male Hindi-Urdu speakers used the coordinating conjunction marker more than females. The differences in the impact of sex on the use of the conjunction marker in my study and Albaqawi's (2020) study may be attributable to sampling differences. In the study undertaken by Albaqawi, the female participants are domestic helpers who work as housemaids and nannies. These types of participants were excluded from my study because they work in the homes of HA native speakers. Therefore, they might be exposed to the HA dialect more than other females who work as messengers and cleaners and males who work as sellers. In fact, even with excluding housemaids, who were included in the study performed by Albaqawi but not in mine, we cannot be certain who the migrants communicate with precisely and what input

[25] I am unable to ascertain whether or not there is an interview impact until both the male and female interviewers interview each respondent.

they receive. Consequently, in general, work could be the reason why male speakers in my study used +CONJ more than females. With males working as sellers and drivers, they might receive more input in the local variety than females who work as cleaners and messengers. Future research on how different kinds of occupations may influence the conjunction marker' use will make a significant contribution to the growth of this field.

Sex impacts the production of +CONJ by L1 Hindi-Urdu speakers but it does not affect +CONJ when it is produced by all the speakers in the corpus and L1 Bengali speakers. This could be attributed to their L1. In various linguistic contexts, the coordinating conjunction marker is more optional in Bengali than in Hindi-Urdu (see Table 16).

Concerning the models that tested the effect of age on the use of the coordinating conjunction marker, the predictor has a significant negative effect on the L1 Bengali speakers only. A possible explanation for this negative effect might be that young L1 Bengali speakers have been exposed to informal Arabic more than old L1 Bengali speakers. While age significantly affects the use of CONJ in the model that includes L1 Bengali speakers, it did not exhibit any significant effect in the models that contain L1 Hindi-Urdu speakers and L1 Tagalog speakers. The reason for this could be that although all the participants in my study are adults and passed the critical period for language acquisition, most L1 Bengali speakers reported that they had learnt Arabic and the Quran since they were approximately seven years old. Conversely, less than half of the L1 Hindi-Urdu speakers and L1 Tagalog speakers reported that they had learnt Arabic and the Quran from an early age. In other words, most L1 Hindi-Urdu speakers and L1 Tagalog speakers learned Arabic whilst residing in Madinah. It is conceivable that their learning of the Arabic language is sufficient to facilitate reciting the Quran and helping them to communicate at work but not beyond.

My findings reveal that the predictor ‘amount of Arabic and L1 used’ significantly affects variant choice in the CONJ models that include all speakers and the models that include only L1 Tagalog speakers, together with the models for /f/ that include only L1 Hindi-Urdu speakers. These results are in line with those obtained by Flege, Frieda and Nozawa (1997), Flege, Yeni-Komshin and Liu (1999), Major (2014), together with Piske, MacKay and Flege (2001), which have shown that using L1 frequently has a negative impact on L2 pronunciation accuracy and proficiency while

regular L2 use has a positive effect on L2 proficiency. 'Amount of Arabic and L1 used' significantly affects the use of the coordinating conjunction marker in the model that includes L1 Tagalog speakers. However, it did not play a significant role in the L1 Bengali speakers' model for the same variable. This may possibly reflect the impact of the migrants' L1. While the coordinating conjunction is predominantly used in the Tagalog language, it is largely optional in Bengali (see Table 16). 'The amount of Arabic and L1 used' also affected the production of [f] by L1 Hindi-Urdu speakers, whereas it did not influence the production of [f] in the model that included L1 Tagalog speakers. This could also be the result of the impact of the migrants' L1. While /f/ exists in the Hindi-Urdu phonemic system, it is absent in Tagalog. These findings suggest that, when CONJ does not exist in the substrate language or if it is optional in the substrate language, the fact that migrants use Arabic more than their L1 does not affect their use of CONJ and [f]. The impact of the amount of Arabic and L1 employed on the use of HA features in some models implies that the Arabic varieties of the workers are interlanguage varieties.

Generally speaking, the four social variables mentioned above consistently have a significant impact on the production of HA features across speaker groups, implying that the Arabic varieties of migrant workers are interlanguage varieties.

8.3.2 Social Variables that Show Significant Variation Across Various Linguistic Variables.

In my research, seven social factors - specifically proficiency in Arabic, attitude towards Madinah, living there, etc, sex, amount of Arabic and L1 used, motivation, exposure to informal Arabic and age – illustrate significant variation across various linguistic variables (see Table 41). For instance, Table 41 reveals that whereas sex significantly affects the use of the conjunction marker in the model that includes only L1 Hindi-Urdu speakers, it did not exhibit any significant effect on these speakers' use of the definite article. The significant impact of social predictors across various linguistic variables implies that the Arabic varieties of migrant workers can be best described as an interlanguage.

Table 41: Social variables that show significant variation across various linguistic variables.

Social variables	Significant effect		Non-significant effect	
	Dependent variable	Model	Dependent variable	Model
Proficiency: migrants' level of Arabic speaking at the time of the interviews	DEF	L1 Hindi-Urdu speakers		
Attitude towards Madinah, living there, its inhabitants and their Arabic variety	/f/	All the speakers in the corpus		
Sex	CONJ	L1 Hindi -Urdu speakers	DEF	L1 Hindi-Urdu speakers
	/f/	all the speakers in the corpus	CONJ	all the speakers in the corpus
		L1 Bengali speakers		L1 Bengali speakers
Amount of Arabic and L1 used	CONJ	all the speakers in the corpus	/f/	all the speakers in the corpus
		L1 Tagalog speakers		L1 Tagalog speakers
Motivation	/f/	all the speakers in the corpus	CONJ	all the speakers in the corpus
		L1 Bengali speakers	DEF	L1 Bengali speakers
Exposure to informal Arabic	/f/	all the speakers in the corpus	CONJ	All the speakers in the corpus

Age	/f/	L1 Tagalog speakers[26]	CONJ	L1 Tagalog speakers

L1 Hindi-Urdu speakers' self-professed proficiency in L2 Arabic has a significant effect on their production of the definite article. Those who reported that they are proficient in Arabic used significantly more +DEF than those who reported their lack of proficiency. This finding is consistent with other research which discovered the impact of this predictor on the L2 acquisition of English h-dropping, (ing), glottal variation in /t/, the plural -s, as well as phonetic accommodation in German (Drummond 2010; Ulbrich 2021; Young 1991). The significance of this predictor indicates that – for L1 Hindi-Urdu speakers at least - the likelihood of using the local variant +DEF increases as the migrant workers' self-declared proficiency improves. This result supports the interlanguage hypothesis.

My results indicate that migrant workers' attitudes towards Madinah, living there, its inhabitants and their Arabic variety have a significant effect on the production of [f] across participants; those with more positive attitudes have significantly higher rates of [f] than those who have a negative attitude. This finding mirrors the results reported by Drummond (2011), who found that the shift towards the local variant of STRUT in Manchester, UK, was more likely among migrants who have a positive attitude towards Manchester, its residents, and their way of life. Drummond indicated that the significant effect of this variable on the acquisition of STRUT is an aspect of identity. It refers to the connection between language use and the expression of one's identity. Specifically, it may involve how participants associate with a linguistic variant as a way of aligning themselves with a particular group or community, influencing the way they speak as a means of expressing their identity. In contrast to a negative attitude, which demonstrates a lack of acquisition as regards the [f] variant, this implies a wish to remain distinct and a determination to maintain migrant worker's L1 identity, a positive attitude reveals movement towards the target variant [f] and suggests a readiness to integrate into the target culture, Hijaz.

Sex has a significant influence on the production of +CONJ by L1 Hindi-Urdu speakers, but a non-significant effect on their production of +DEF. This might be due to the type of morphosyntactic

[26] Age has a significant negative effect in relation to the L1 Tagalog speakers' model.

feature. The coordinating conjunction is a free morpheme while the definite article is a bound morpheme. For social reasons such as exposure, the former might be easier for male Hindi-Urdu speakers to acquire than the latter. This finding might be similar to the study completed by Zobl and Liceras (1994) who determined that, irrespective of of their L1, L2 learners tend to produce free morphemes earlier and use them with greater precision than bound morphemes. Sex is also significant in the models that include (i) all the speakers in the corpus, and (ii) L1 Bengali speakers when they produced /f/. However, this predictor did not influence (i) all the speakers in the corpus, and (ii) L1 Bengali speakers when they produced +CONJ. This could be the impact of their L1. In Bengali, /ph/ is frequently produced as /f/ (see Table 29). Conversely, the conjunction marker is not applied in all linguistic contexts (see Table 16).

While the 'amount of Arabic and L1 used' significantly affects the use of CONJ in the models that comprise all sampled migrant workers and L1 Tagalog speakers, it shows no significant effect in either of these models when participants produce /f/. A possible explanation for this might be the influence of their L1. Tagalog has a conjunction marker, but /f/ does not exist in the Tagalog phonemic inventory.

My study ascertained that motivation has a significant effect on the production of /f/ by all the speakers in the corpus and by L1 Bengali speakers. Those who are highly motivated to learn Arabic use [f] more than those who lack motivation. My quantitative findings confirm the qualitative finding reported by Abu-Rabia and Kehat (2004). Their study suggests that motivation might be the reason behind the success of late L2 starters acquiring the L2 accent. While motivation significantly affects the production of /f/ in the model that includes L1 Bengali speakers, it did not exhibit any significant effect on their adoption of the definite article. These results suggest that acquiring the phonological feature might be easier than the morphosyntactic feature for this language group, possibly due to their L1 influence. In addition, motivation significantly impacts the production of /f/ in the model that includes all the speakers in the corpus, although it did not exhibit any significant influence in their use of the conjunction marker. This finding suggests that migrant speakers with different L1s may find it easier to acquire the phonological L1 HA feature than the morphosyntactic feature. Future research could benefit from further exploration of this phenomenon. In general, the impact of this predictor suggests that acquiring the phonological HA

feature is more straightforward than acquiring the morphosyntactic HA features – irrespective of whether they are bound or free morphemes. The variation within the same language group model across various linguistic variables suggests that the Arabic varieties of these migrants are an interlanguage.

Exposure to informal Arabic was found to be statistically significant in the model for [f] which includes all the speakers in the corpus. Listening to the Qur'an alone does not seem enough input for migrant workers to produce high rates of [f] (see Table 31). However, when the migrant workers in my study report being exposed to the Arabic language through watching television and listening to the radio alongside listening to the Qur'an, they produce more target-like features. The significance of this predictor suggests the importance of exposure to the target language across different media because the likelihood of using the target variant [f] increases as the migrant workers' exposure to informal Arabic diversifies and increases. In contrast, exposure to informal Arabic did not significantly impact the use of CONJ in the model that includes all the speakers in the corpus. Again, the difference between the significant effect of this predictor on the production of /f/ and the non-significant effect on the use of CONJ may be due to the type of linguistic variable. Future studies will benefit from examining whether acquiring HA phonological features is more straightforward than acquiring morphosyntactic features. The significant impact of exposure to informal Arabic on the production of /f/ within the same language group model, as opposed to its impact on the use of CONJ, implies that the Arabic varieties of the migrants are an interlanguage.

Age has a significant negative effect in the model that tested its effect on L1 Tagalog speakers' production of /f/. This negative effect suggests that the non-HA variety being learned might be an interlanguage. A possible explanation for this could be similar to what I have previously discussed in the model that examines the impact of age on the use of the coordinating conjunction marker by L1 Bengali speakers in the earlier categorisation (7.3.2). Young L1 Tagalog speakers have been exposed to informal Arabic more than old L1 Tagalog speakers. Therefore, the production of /f/ by old L1 Tagalog speakers does not mirror the production of this feature by HA speakers. Age significantly affects the production of /f/ in the model that includes L1 Tagalog speaker, although

it did not influence their use of CONJ. This might again be related to the type of linguistic variable analysed.

In general, the seven social variables mentioned above exert a significant impact on the production of specific HA features within the same language group model. However, these independent variables may not necessarily influence other HA features within the same language group. This variation across various linguistic variables suggests that the Arabic varieties spoken by migrant workers can be regarded as interlanguage varieties.

8.3.3 Social Factors that Neither Show Significant Variation Across Various Language Groups nor Across Various Linguistic Variables.

Seven predictors, specifically the identity of the migrant, educational level in L1, language anxiety, formal instruction in their country of origin, formal Arabic instruction in Hijaz, motivation and length of residence, do not demonstrate any significant differences either between language groups or between different variables in this sub-section (see Table 42). These findings do not support the results of the L2 studies undertaken by Abu-Rabia and Kehat (2004), Abu-Ghararah (1999), Aida (1994) Alderson (2005), Al Kendi (2021), Bosmans and Hurda (2016), Drummond (2010), Elliott (2018), Flege, Bohn and Jang (1997), Krashen *et al.* (1978), Krashen and Seliger (1976), Krashen, Seliger and Hartnett (1974), MacIntyre and Gardner (1989), Phillips (1992) and Regan (2013b). These studies demonstrate that the above-mentioned predictors significantly affect English proficiency, English achievement, Arabic phonological sensitivity and the production of coda consonant clusters, performance in Japanese, English glottal variation in /t/, French *ne* deletion, FACE and GOAT vowels, /ing/ in English, French phonological attainment, learning and the production of French vocabulary, along with speaking French. Given that a number of the aforementioned seven predictors are correlated with other significant predictors which cause confounds in the analysis and others may be attributable to the limitation of the sample size, considering the non-significant impact of these variables on the migrants' use of HA features suggesting pidgin varieties is challenging (see Appendix A).

While Albaqawi (2020) and Almoaily (2012) established that LoR significantly influences the use of the conjunction markers, this predictor did not impact the use of the conjunction marker by L1

Tagalog speakers. This might reflect the quality of input that the migrants received from their local community, which includes more +CONJ. As Almoaily (2008) discovered, conjunctions are the only linguistic feature that the locals use with the standard form 72.2%, while the non-standard form was 27.8%. The conjunction marker is a categorical variable in Arabic; however, the variation in the speech of the locals might be referred to as 'foreigner talk'. According to Ferguson (1971), when native speakers simplify their own language when interacting with others who have scarcely any proficiency in that language, this is termed 'foreigner talk'. The locals HA speakers were not included as participants in my study. Future studies may fill this gap and explore whether the L1 HA speakers' use of this variable when they communicate with the migrants influences the migrants' speech. In relation to the non-significant effect of LoR on the production of /f/ in the model that includes L1 Hindi-Urdu speakers, this might be due to /f/ already existing in the phonemic inventory of their L1. Therefore, it may be relatively straightfoward for them to produce /f/ in their L2 Arabic as well.

Table 42: Social variables that neither show significant variation across various language groups nor across various linguistic variables.

Social variable	Dependent variable	Model
The identity of the migrant	DEF	All the speakers in the corpus
	CONJ	L1 Hindi-Urdu speakers
Educational level in L1	DEF	All the speakers in the corpus
	CONJ	L1 Tagalog speakers
Language anxiety	DEF	L1 Bengali speakers
		L1 Hindi-Urdu speakers
		L1 Tagalog speakers
	/f/	All the speakers in the corpus
Formal instruction in their country of origin	CONJ	All the speakers in the corpus
		L1 Hindi-Urdu speakers
	/f/	All the speakers in the corpus
Formal Arabic instruction in Hijaz	CONJ	L1 Hindi-Urdu speakers
Motivation	DEF	L1 Tagalog speakers

LoR	CONJ	L1 Tagalog speakers
	/f/	L1 Hindi-Urdu speakers

Ordinarily, the quantitative analysis of the morphosyntactic and phonological features denotes that the Arabic varieties of the migrant workers in my study are an interlanguage. Alternatively, the qualitative analysis undertaken on the morphological features suggest that their Arabic varieties are a pidgin. Thus, I can conclude that the Arabic varieties of the migrant workers in my study are located on a continuum, where the higher end may be closer to an interlanguage.

8.4 Conclusion

It can be concluded that the three individual language groups of the migrant workers in Madinah have reduced verbal and nominal systems. The lack of the subject-verb agreement and the noun-adjective-agreement signify that their Arabic varieties could be best described as a pidgin language. Migrants also share the characteristic of simplification concerning the use of morphosyntactic features. They omit the definite article and the coordinating conjunction marker more frequently than they use them. Granting that most of the contextual predictors included in the models that are associated with these two linguistic variables do not have a significant influence on the choice of the variant, the variations across various linguistic variables and throughout various language groups signify that the L2 Arabic spoken by the migrants represent an interlanguage. Conversely, migrants primarily produce [f] where their L1 influences are apparent. Independent variables in the models pertaining to the phonological feature typically have significant effects on the production of [f] and support the interlanguage hypothesis. The analysis of morphological features and the analysis of morphosyntactic and phonological features point towards opposite conclusions. I tentatively propose that the Arabic varieties of the migrant workers studied in this thesis are located on a continuum, where the higher end may be closer to an interlanguage.

Future research will benefit from a larger sample size. Although thirty participants are manageable for sociolinguistics research, increasing the the number of participants would enhance the precision and reliability of statistical results in identifying differences among the participants. Other ethnic groups can be found in Saudi Arabia. A future study could broaden the data set to include

additional ethnicities, e.g., Ethiopian, Nigerian, Indonesian, Panjabi and Sinhala to provide a more detailed representation of the migrants' diversity and to investigate the substratal variations among the participants. Likewise, it would be fascinating to compare the Arabic variety of migrant workers from Africa who speak, for example, Amharic and Luganda as their first language to the Arabic variety of workers from East Asia. This could demonstrate language variation and the impact of the L1. Analyses of a larger set and a wider range of variables will confirm whether my hypothesis about the Arabic varieties of the migrants in Madinah is correct.

References

Abaalkhail, Faisal. 1998. 'Syllabification and metrification in Urban Hijazi Arabic: between rules and constraints'. Unpublished PhD dissertation, University of Essex.

Abdelhady, S. 2019. 'A sociolinguistic study of the use of the definite article (ʔal) in the Jordan Valley'. *SKY Journal of Linguistics* 32: 7-41.

Abed, Lowai. 2018. 'Gulf Pidgin Arabic: An empirical investigation into its emergence and social attitude'. Unpublished PhD dissertation, Trinity College Dublin.

Abney, Steven. 1987. 'The English noun phrase in its sentential aspect'. Unpublished PhD dissertation, Massachusetts Institute of Technology.

Abu-Ghararah, A. H. 1999. 'Learning anxiety and English language achievement of male and female students of university and secondary stages in Al-Madinah Al-Munawwarah; A comparative research study'. *King Abdulaziz University Journal* 12(1): 3-29.

Abu-Rabia, S. and S. Kehat. 2004. 'The critical period for second language pronunciation: Is there such a thing?'. *Educational Psychology* 24(1): 77- 97.

Adamson, H. D. and V. Regan. 1991. 'The acquisition of community speech norms by Asian immigrants learning English as a second language'. *Studies in Second Language Acquisition* 13: 1–22.

Agarwal, H. 2019. 'To Dº or not to Dº: Nominal structure in Hindi-Urdu'. *Inquiry: A Journal of Undergraduate Research* 23: 36.

Agnihorti, R. K. 2006. *Hindi: An Essential Grammar.* London: Routledge.

Aida, Y. 1994. 'Examination of Horwitz, Horwitz and Cope's construct of foreign language anxiety: The case of students of Japanese'. *The Modern Language Journal* 78(2): 155-168.

Aikhenvald, A. 2002. 'Language contact in Amazonia'. In A. Aikhenvald (ed) *The Languages of the Amazon.* Oxford: Oxford University Press. 68-98.

Alansari, A. 1970. *A masterpiece of lovers and friends in knowing the genealogy of Madinah dwellers.* Tunis: Almaktaba Al-Atika.

Al-Ansari, G. 2007. *The Roots of Gold in Knowing the Words of the Arabs*. Egypt: Dar Assalaam.

Al-Azraqi, M. 2011. 'Pidginisation in the eastern region of Saudi Arabia: Media presentation'. In R. Bassiouney (ed) *Arabic and the Media.* Boston: Brill. 159-173.

Al-Azraqi, M. 2020. ‘Pidginisation in Abha, Saudi Arabia: Pathways to multifunctionality’. *Journal of Pidgin and Creole Languages* 35(1): 37-61.

Albakrawi, H. 2012. ‘The linguistic effect of foreign Asian workers on the Arabic Pidgin in Saudi Arabia’. *Language* 2(9): 127-133.

Albaqawi, N. S. 2016. ‘Unity and diversity within pidginised Arabic as produced by Asian migrant workers in the Arabian Gulf’. *Journal of Second Language Teaching and Research* 5: 253-290.

Albaqawi, Najah Salem. 2020. ‘Gender variation in Gulf Pidgin Arabic’. Unpublished PhD dissertation, University of Wolverhampton.

Albaqawi, N. and M. Oakes. 2019. 'Compiling and analysing a corpus of transcribed spoken Gulf Pidgin Arabic based on length of stay in the Gulf'. Paper presented at the 3rd Workshop on Arabic Corpus Linguistics. July, Cardiff.

Al Barrag, T. and S. Alzahrani. 2019. ‘The formation of simplex nouns in Urban Hijazi Arabic: A distributed morphology approach’. *Sino-US English Teaching 16*(7): 269-292.

Albirini, A. 2016. *Modern Arabic Sociolinguistics: Diglossia, Variation, Codeswitching, Attitudes and Identity*. London: Routledge.

Alderson, J. C. 2005. *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.

Alghamdi, E. 2014. ‘Gulf Pidgin Arabic: A descriptive and statistical analysis of stability’. *International Journal of Linguistics 6*: 110-126.

Alhamadi, Nesreen. 2015. ‘Language transfer errors in speaking among Saudi Arabia students; a comparative study between Students in Saudi Arabia and Students in the UK’. Unpublished PhD dissertation, University of Leeds.

Alhazmi, Laila. 2018. ‘Perceptions of Hijazi Arabic Dialects: An attitudinal approach’. Unpublished PhD dissertation, University of Sheffield.

Alhazmi, M. Z. 2020. ‘Urban Pidgin and Bedouin L2 in the Hijaz’. *Anthropological Linguistics* 62(4): 391-413.

Al-Jehani, Nasir. 1985. ‘Sociolinguistic stratification of Arabic in Makkah’. Unpublished PhD dissertation, The University of Michigan.

Al-Jehani, N. 1990. ‘Colour terms in Mecca: A sociolinguistic perspective’. *Anthropological Linguistics 32*(1–2): 163–74.

Aljutaily, Mohammad. 2018. 'The influence of linguistic and non-linguistic factors on the variation of Arabic marked consonants in the speech of Gulf Pidgin Arabic: acoustic analysis'. Unpublished PhD dissertation, University of Georgia.

Al Kendi, Azza. 2021. 'Foreigner-directed speech and L2 speech learning in an understudied interactional setting: The Case of foreign-domestic helpers in Oman'. Unpublished PhD dissertation, Newcastle University.

Almalky, Abdullah. 2020. 'The active particle in Hijazi Arabic: An LFG Perspective'. Unpublished PhD dissertation, University of Essex.

Almoaily, Mohammad. 2008. 'A data-based description of Urdu Pidgin Arabic'. MA dissertation, Newcastle University.

Almoaily, Mohammad. 2012. 'Language variation in Gulf Pidgin Arabic'. Unpublished PhD dissertation, Newcastle University.

Al-Mozainy, Hamza. 1981. 'Vowel alternations in a Bedouin Hijazi Arabic dialect'. Unpublished PhD dissertation, University of Texas.

Al-Musnad, B. I. 2018. 'The role of motivation and attitude in second language learning: A study of Arabic language learning among foreign female nurses in Riyadh, Saudi Arabia'. *Journal of Applied Linguistics and Language Research* 5(1): 157-183.

Alok, D and S. K. Mishra. 2013. 'No DP-analysis of Indo-Aryan Languages'. *Aligarh Journal of Linguistics* 3: 208-219.

Alqahtani, F. and M. Sanderson. 2019. 'Generating a lexicon for the Hijazi dialect in Arabic'. In *International Conference on Arabic Language Processing*. France. 3-17.

Alshammari, Wafi. 2010. 'An investigation into morpho-syntactic simplification in the structure of Arabic based pidgin in Saudi Arabia'. Unpublished MA dissertation, Mu'tah University.

Alshammari, Wafi. 2018. 'The development of and accommodation in Gulf Pidgin Arabic: verbal and pronominal form selection'. Unpublished PhD dissertation, Indiana University.

Alshraidah, A. 2011. Pidgin language.. the language of migrant workers between the harm and necessary. Alriyadh Newspaper. 14th Apr. Last accessed 01 November 2020, from: http://www.alriyadh.com/623569.

Al-Shurafa, N. 2014. 'Linguistic functions of fii/maafii in Hejaz Gulf Pidgin Arabic'. *Journal of King Khalid University* 21(1): 11-27.

Alsusut, A. 2020. ‘Pidgin language: sociolinguistic study (Southern Saudi Arabia is a model). In Gh. J (ed) *The written saying in the history of the South*. Riyadh: Alhumaidhi Press. 301-358.

Alzaidi, Muhammad. 2014. ‘Information structure and intonation in Hijazi Arabic’. Unpublished PhD dissertation, University of Essex.

Al-Zubeiry, H. Y. A. 2015. ‘Linguistic analysis of Saudi pidginised Arabic as produced by Asian foreign expatriates’. *International Journal of Applied Linguistics and English Literature* 4(2): 47-53.

Andersen, R. W. 1979. ‘The relationship between first-language transfer and second-language overgeneralization: Data from the English of Spanish speakers’. In R. W. Andersen (ed) *The Acquisition and Use of Spanish and English as First and Second Languages*. Washington, D.C.: TESOL. 43–58.

Ansaldo, U and S. Matthews. 2004. ‘The origins of Macanese reduplication’. In G. Escure and A. Schwegler (eds) *Creoles, Contact, and Language Change: Linguistic and Social Implications*. Amsterdam: John Benjamins Publishing. 1-20.

Anthony, L. 2022. *AntConc (Version 4.0.11)*. Tokyo: Waseda University.

Araujo, L. and P. Dinis da Costa 2013. *The European Survey on Language Competences: School Internal and External Factors in Language Learning*. Luxembourg: Publications Office of the European Union.

Arnold, J. and H. D. Brown. 1999. ‘A map of the terrain’. In J. Arnold (ed) *Affect in Language Learning.* Cambridge: Cambridge University Press. 1–24.

Avram, A. 2015. ‘On the developmental stage of Gulf Pidgin Arabic’. Paper presented in Arabic Varieties: Far and Wide. In *Proceedings of the 11th International Conference of AIDA*. Bucharest. 87-98.

Baayen, R. H., D. J. Davidson and D. M. Bates. 2008. ‘Mixed-effects modeling with crossed random effects for subjects and items’. *Journal of Memory and Language* 59(4): 390-412.

Barr, D. J., R. Levy, C. Scheepers and H. J. Tily. 2013. ‘Random effects structure for confirmatory hypothesis testing: Keep it maximal’. *Journal of Memory and Language* 68(3): 255-278.

Barrios, J. 2011. *Tagalog for Beginners: An Introduction to Filipino, the National Language of the Philippines*. US: Tuttle Publishing.

Berk, R. A. and J. M. Adams. 2001. 'Establishing rapport with deviant groups'. In J.M. Miller and R. Tewksbury (eds). *Extreme Methods: Innovative Approaches to Social Science Research*. Boston: Allyn and Bacon. 58–71.

Birdsong, D. 2007. 'Nativelike pronunciation among late learners of French as a second language'. In O. Bohn and M. J. Munro, (eds) *Language Experience in Second Language Speech Learning*. Amsterdam: John Benjamins. 99-116.

Baker, W. 2010. 'Effects of age and experience on the production of English word-final stops by Korean speakers'. *Bilingualism: Language and Cognition 13*(3): 263-278.

Bakir, M. J. 2010. 'Notes on the verbal system of Gulf Pidgin Arabic'. *Journal of Pidgin and Creole Languages* 25(2): 201-228.

Bakker, P. 1994. 'Pidgins'. In J. Arends, P. Muysken and N. Smith (eds) *Pidgins and Creoles an Introduction.* Amsterdam; Philadelphia: J. Benjamins. 25-38.

Bakker, P. 2008. 'Pidgins versus creoles and pidgincreoles'. In S. Kouwenberg and J. Singler (eds) *The Handbook of Pidgin and Creole Studies*. Oxford: Blackwell. 130-157.

Barman, B. 2009. A contrastive analysis of English and Bangla phonemics. *Dhaka University Journal of Linguistics* 2(4): 19-42.

Baranowski, M. and D. Turton. 2020. 'TD-deletion in British English: New evidence for the long-lost morphological effect'. *Language Variation and Change* 32(1): 1-23.

Barnard, M. 2005. 'Discomforting research: Colliding moralities and looking for 'truth' in a study of parental drug problems'. *Sociology of Health and Illness* 27(1): 1-19.

Bates, D., M. Mächler, B. Bolker and S. Walker. 2014. 'Fitting linear mixed-effects models using lme4'. *Journal of Statistical Software* 67(1): 1-48.

Bayley, R. 1996. 'Competing constraints on variation in the speech of adult Chinese learners of English'. In R. Bayley and D. R. Preston (eds) *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins. 97-120.

Bayley, R. 2007. 'Second language acquisition: A variationist perspective'. In Bayley, R and C. Lucas (eds) *Sociolinguistic Variation: Theories, Methods and Applications*. Cambridge: Cambridge University Press. 133-144.

Bhatia, T. K. 1987. *A History of the Hindi Grammatical Tradition: Hindi-Hindustani Grammar, Grammarians, History and Problems*. Leiden: Brill.

Bhattacharya, T. 1999a. 'The structure of the Bangla DP'. Unpublished PhD dissertation, University College London.

Bhattacharya, T. 1999b. 'Specificity in the Bangla DP'. *Yearbook of South Asian Languages and Linguistics* 2: 71-99.

Bialystok, E. 2002. 'Acquisition of literacy in bilingual children: A framework for research'. *Language Learning* 52(1): 159-199.

Bigelow, M. and E. Tarone. 2004. 'The role of literacy level in second language acquisition: Doesn't who we study determine what we know?'. *TESOL Quarterly* 38(4): 689-700.

Birdsong, D. 1992. 'Ultimate attainment in second language acquisition'. *Language* 68: 706-755.

Birman, D. 2005. 'Ethical issues in research with immigrants and refugees'. In J. E., Trimble and C. B Fisher (eds) *The Handbook of Ethical Research with Ethnocultural Populations and Communities*. London: Sage Publications. 155-178.

Bongaerts, T., B. Planken and E. Schils. 1995. 'Can late starters attain a native accent in a foreign language? A test of the critical period hypothesis'. In D. M. Singleton and Z. Lengyel (eds) *The Age Factor in Second Language Acquisition: A Critical Look at the Critical Period Hypothesis*. Clevedon: Multilingual Matters. 30- 50.

Bongaerts, T., C. van Summeren, B. Planken and E. Schils. 1997. 'Age and ultimate attainment in the pronunciation of a foreign language'. *Studies in Second Language Acquisition* 19: 447-465.

Booth, T. and W. Booth. 1994. 'The use of depth interviewing with vulnerable subjects: Lessons from a research study of parents with learning difficulties'. *Social Science and Medicine* 39(3): 415-424.

Bosmans, D. and S. Hurd. 2016. 'Phonological attainment and foreign language anxiety in distance language learning: a quantitative approach'. *Distance Education* 37(3): 287–301.

Bošković, Ž. 2008. 'What will you have, DP or NP?' In E. Elfner and M. Walkow (eds) *Proceedings of the North East Linguistic Society.* Amherst, MA: BookSurge Publishing. 101–114.

Bošković, Ž. 2009. 'More on the no-DP analysis of article-less languages'. *Studia Linguistica* 63(2): 187-203.

Bošković, Ž. And J. Gajewski. 2011. 'Semantic correlates of the NP/DP parameter. In S. Lima, K. Mullin and B. Smith (eds) *Proceedings of the Thirty-Ninth Annual Meeting of the Northeast Linguistic Society.* Amherst, MA: CreateSpace Independent Publishing Platform. 121–134.

Brown, H. D. 2000. *Principles of Language Teaching and Learning.* New York: Longman.

Brugman, H., A. Russel and X. Nijmegen. 2004. 'Annotating Multi-media/multi-modal resources with ELAN'. In *LREC Proceedings*. Lisbon. 2065-2068.

Buchstaller, I. and G, Khattab. 2013. 'Population samples. In R. Podesva and D. Sharma (eds) *Research Methods in Linguistics.* Cambridge: Cambridge University Press.

Burnham, K. and D. Anderson. 2002. *Model Selection and Multi-Model Inference*. New York: Springer-Verlag.

Butler, J. 1988. 'Performative acts and gender constitution: An Essay in phenomenology and feminist theory'. *Theatre Journal*: 40(4): 519-531.

Chacón, D. 2012. 'Head movement in the Bangla DP'. *Journal of South Asian Linguistics* 4: 3-25.

Chambers, J. K. 1995. *Sociolinguistic Theory: Linguistic Variation and its Social Significance*. Cambridge, Mass: Blackwell.

Cheshire, J. 2002. 'Sex and gender in variationist research'. In J. Chambers, P. Trudgill and N. Schilling-Estes (eds) *The Handbook of Language Variation and Change. Blackwell Publishers*. 423–443.

Chiu, Y. H. 2004. 'The L2 acquisition of the coordinating conjunction "and" in Taiwanese learners' interlanguage at college level'. *Asian EFL Journal* 6(3): 1-20.

Christians, C. G. 2005. 'Ethics and politics in qualitative research'. In N.K. Denzin and Y.S. Lincoln (eds) *The Sage Handbook of Qualitative Research.* Thousand Oaks: Sage Publications. 139–164.

Chung, T. 2011. 'Variability in the L2 Acquisition of English Articles'. *English Teaching* 66: 169-187.

Cofer, Thomas. 1972. 'Linguistic variability in a Philadelphia speech community'. Unpublished Phd dissertation, University of Pennsylvania.

Cohen, L., L. Manion and K. Morrison. 2007. *Research Methods in Education.* Abingdon, Oxon: Routledge.

Collier, V. P. 1989. 'How long? A synthesis of research on academic achievement in a second language'. *TESOL Quarterly* 23(3): 509-531.

Collins, J. N. 2019. 'Definiteness determined by syntax: A case study in Tagalog'. *Natural Language and Linguistic Theory* 37: 1367-1420.

Colton, D. and R. W. Covert. 2007. *Designing and Constructing Instruments for Social Research and Evaluation*. Hoboken: John Wiley and Sons.

Cornford, T. and S. Smithson. 1996. *Project Research in Information Systems: A Student's Guide*. Basingstoke: Palgrave Macmillan.

Cortes, K., L. Milambiling, I. Paul and P. Caxaj. 2012. 'On Tagalog determiners'. In *Proceedings of the Canadian Linguistic Association*.

Costa, P. and P. Albergaria-Almeida. 2015. 'The European survey on language competences: Measuring foreign language student proficiency'. *Procedia - Social and Behavioural Sciences* 191: 2369–2373.

Crosthwaite, P. and L. L. Y. Choy. 2016. 'A learner corpus investigation of Filipino L2 English article use'. *Asian EFL Journal* 93: 4-23.

Cummins, J. 1991. 'Conversational and academic language proficiency in bilingual contexts'. *AILA review* 8(91): 75-89.

Daly, K. 1992. 'The fit between qualitative research and characteristics of families'. In J.F. Gilgun, K. Daly and G. Handel (eds) *Qualitative Methods in Family Research*. Newbury Park: Sage Publications. 3-11.

Davies, B. 2004. 'The gender gap in modern languages: a comparison of attitude and performance in Year 7 and Year 10'. *Language Learning Journal* 29(1): 53-58.

David, A. 2015. *Descriptive Grammar of Bangla*. In T. Conners and D. Chacon (eds). Berlin: De Gruyter Mouton.

De Bel-Air, F. 2018. Demography, Migration and Labour Market in Saudi Arabia. Last accessed 20 June 2020, from: http://gulfmigration.org/media/pubs/exno/GLMM_EN_2018_05.pdf

DeG`raff, M. 1999. 'Creolisation, language change and language acquisition: An epilogue'. In M. DeGraff (ed) *Language Creation and Language Change: Creolisation, Diachrony, and Development*. Cambridge: MIT Press. 473–543.

Dickson-Swift, Virginia. 2005. 'Undertaking sensitive health research: The experiences of researchers'. Unpublished PhD dissertation, La Trobe University.

Diskin, C. 2017. 'The use of discourse-pragmatic marker 'like' by native and non-native speakers of English in Ireland. *Journal of Pragmatics* 120:144-157.

Dörnyei, Z. 1990. 'Conceptualising motivation in foreign-language learning'. *Language Learning* 40(1): 45-78.

Dörnyei, Z. 2001. *Teaching and Researching Motivation.* Harlow: Longman.

Dörnyei, Z. 2002. *Questionnaires in Second Language Research: Construction, Administration and Processing.* Mahwah, New Jersey: Lawrence Erlbaum.

Dörnyei, Z. 2003. 'Attitudes, orientations and motivations in language learning: Advances in theory, research and applications'. *Language Learning* 53(s1): 3-32.

Dörnyei, Z. 2007. *Research Methods in Applied Linguistics.* Oxford: Oxford University Press.

Drager, K. and Hay, J. 2012. 'Exploiting random intercepts: Two case studies in sociophonetics'. *Language Variation and Change* 24(1): 59-78.

Drummond, Rob. 2010. 'Sociolinguistic variation in a second language: the influence of local accent on the pronunciation of non-native English speakers living in Manchester'. Unpublished PhD dissertation, University of Manchester.

Drummond, R. and E. Schleef. 2016. 'Identity in variationist sociolinguistics'. In S. Preece (ed) *The Routledge Handbook of Language and Identity.* London: Routledge. 50-65 .

Dulay, H. C. and M. K. Burt. 1974. 'Natural sequences in child second language acquisition'. *Language Learning* 24(1): 37-53.

Eaton, S. E. 2010. *Formal, Non-Formal and Informal learning: The Case of Literacy, Essential Skills and Language Learning in Canada*. Calgary, AB: Eaton International Consulting Inc.

Eckert, P. 1989. 'The whole woman: Sex and gender differences in variation'. *Language Variation and Change* 1(3): 245–267.

Eckert, P. and S. McConnell-Ginet. 1992. 'Think practically and look locally: Language and gender as community-based practice'. *Annual Review of Anthropology* 21(1): 461-488.

Edwards, J. G. H. 2011. Deletion of /t, d/ and the acquisition of linguistic variation by second language learners of English. *Language Learning* 61 (4): 1256- 1301.

Elliott, Zuzana. 2018. 'Sociolinguistic variation among Slovak immigrants in Edinburgh, Scotland'. Unpublished PhD dissertation, University of Edinburgh.

Ellis. R. 1992. *Second Language Acquisition and Language Pedagogy*. Clevedon: Multilingual Matters.

Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Epstein, S., S. Flynn and G. Martohardjono. 1996. 'Second language acquisition: theoretical and experimental issues in contemporary research'. *Behavioural and Brain Sciences* 19(4): 677-714.

Epstein, S., S. Flynn and G. Martohardjono. 1998. 'The strong continuity hypothesis in adult L2 acquisition of functional categories'. In S. Flynn, G. Martohardjono and W. O'Neil (eds) *The Generative Study of Second Language Acquisition*. Hillsdale: Lawrence Erlbaum. 61-77.

Ethnologue Languages of the World. Last accessed 4 March 2021 from: https://www.ethnologue.com/

Eubank, L. 1994. 'Optionality and the initial state in L2 development'. In T. Hoekstra and B. D. Schwartz (eds) *Language Acquisition Studies in Generative Grammar*. Amsterdam: John Benjamins. 369-388.

Faltis, C. J. and C.A. Coulter. 2008. *Teaching English Learners and Immigrant Students in Secondary Schools*. Upper Saddle River, NJ: Prentice Hall.

Ferguson, C. A. 1971. 'Absence of copula and the notion of simplicity'. In D. H Hymes (ed) *Pidginisation and Creolisation of Languages*. Cambridge: Cambridge University Press 141-150.

Field, F. 2004. 'Second language acquisition in creole genesis: The role of processability'. In G. Escure and A. Schwegler (eds) *Creoles, Contact, And Language Change: Linguistic and Social Implications*. Amsterdam: John Benjamins Publishing.

Fischer, J. L. 1958. 'Social influences on the choice of a linguistic variant'. *Word*. 14(1): 47-56.

Flaskerud, J. H. and B. J. Winslow. 1998. 'Conceptualising vulnerable populations health-related research'. *Nursing Research* 47(2): 69-78.

Flege, J. E. 1981. 'The phonological basis of foreign accent: A hypothesis'. *TESOL Quarterly* 15(4): 443-455.

Flege, J. E. 1988. 'Factors affecting degree of perceived foreign accent in English sentences'. *The Journal of the Acoustical Society of America* 84(1): 70-79.

Flege, J. E. 1993. 'Production and perception of a novel, second-language phonetic contrast'. *The Journal of the Acoustical Society of America* 93(3): 1589-1608.

Flege, J. E. 2009. 'Give input a chance!'. In T. Piske and M. Young-Scholten (eds) *Input Matters in SLA*. Briston: Multilingual Matters. 175-190.

Flege, J. E., D. Birdsong, E. Bialystok, M. Mack, H. Sung and K. Tsukada. 2006.' Degree of foreign accent in English sentences produced by Korean children and adults'. *Journal of Phonetics* 34(2): 153-175.

Flege, J. E., O. S. Bohn and S. Jang. 1997. 'Effects of experience on non-native speakers' production and perception of English vowels'. *Journal of Phonetics* 25(4): 437-470.

Flege, J. E. and K. L. Fletcher. 1992. 'Talker and listener effects on degree of perceived foreign accent'. *The Journal of the Acoustical Society of America* 91(1): 370- 389.

Flege, J. E., E. M. Frieda and T Nozawa. 1997. 'Amount of native-language (L1) use affects the pronunciation of an L2'. *Journal of Phonetics* 25(2): 169-186.

Flege, J. E. and S. Liu. 2001. 'The effect of experience on adults' acquisition of a second language'. *Studies in Second Language Acquisition* 23(4): 527-552.

Flege, J. E., M. J. Munro and L. Skelton. 1992. 'Production of the word-final English /t/– /d/ contrast by native speakers of English, Mandarin, and Spanish'. *The Journal of the Acoustical Society of America* 92(1): 128-143.

Flege, J. E., G. H. Yeni-Komshian and S. Liu. 1999. 'Age constraints on second language Acquisition'. *Journal of Memory and Language* 41(1): 78-104.

Flick, W. and G, Glenn. 1976. 'Second language learning vs. pidginisation'. Paper presented at the convention of teacher of English to speakers of other languages, March, New York.

Flynn, S. and G. Martohardjono. 1994. 'Mapping from the initial state to the final state: The separation of universal principles and language-specific principles. In B. Lust, M. Suner and J. Whitman (eds) *Syntactic Theory and First Language Acquisition: Cross-linguistic Perspectives: Heads, Projections and Learnability.* Hillsdale: Lawrence Erlbaum. 319-335.

Gal, S. 1978. 'Peasant men can't get wives: Language change and sex roles in a bilingual community'. *Language in Society*. 7(1): 1-16.

Gardner, R. C. 1985a. 'The attitude/motivation test battery'. *Technical Report.* London, ON: University of Western Ontario.

Gardner, R. C. 1985b. *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*. London: Edward Arnold.

Gatbonton, E., P. Trofimovich and M. Magid. 2005. 'Learners' ethnic group affiliation and L2 pronunciation accuracy: A sociolinguistic investigation'. *TESOL Quarterly* 39: 489-512.

Gillham, B. 2002. *Developing a Questionnaire.* London: Continuum International Publishing Group.

Gordon, R. J. 2005. *Ethnologue: Languages of the World.* Dallas: SIL International.

Gorman, K. and D. E. Johnson. 2013. 'Quantitative analysis'. In R. Bayley, R. Cameron and C. Lucas (eds) *The Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press. 214–240.

Guy, G. 1980. 'Variation in the group and the individual: The case of final stop deletion'. In W. Labov (ed) *Locating Language in Time and Space.* New York: Academic Press. 1-36

Guy, G. R. 1993. 'The quantitative analysis of linguistic variation'. In D. R. Preston (ed.) *American Dialect Research.* Amsterdam: John Benjamins. 223–279.

Hall, R. A. 1953. *Haitian Creole: Grammar, Texts, Vocabulary*. Philadelphia: American Anthropological Association.

Haznedar, B. 1997. 'L2 acquisition by a Turkish-speaking child: evidence for L1 influence'. In E. Hughes, M. Hughes and A. Greenhill (eds) *Proceedings of the 21st Annual Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. 245–256.

Hobrom, Anwar. 1996. 'An analysis of the Arabic Pidgin spoken by Indian workers in Saudi Arabia'. MA dissertation, University of Kansas.

Holm, J. 1989. *Pidgins and Creoles: Volume II, Reference Survey*. Cambridge: Cambridge University Press.

Holm, J. 2000. *An Introduction to Pidgins and Creoles*. Cambridge: Cambridge University Press.

Horwitz, E. K., M. B. Horwitz and J. Cope. 1986. 'Foreign language classroom anxiety'. *The Modern Language Journal* 70(2): 125-132.

Huber, M. 1999. *Ghanaian Pidgin English in its West African context: A Sociohistorical and Structural Analysis*. Amsterdam: John Benjamins Publishing.

Hudson, G. 2000. Essential Introductory Linguistics. Wily Blackwell Publishers.

Il-Hazmy, Alayan. 1975. 'A critical and comparative study of the spoken dialect of the Harb tribe in Saudi Arabia'. Unpublished PhD dissertation, University of Leeds.

Islam, Riaz. 2011. 'The morphology of loanwords in Urdu: the Persian, Arabic and English strands'. Unpublished PhD dissertation, Newcastle University.

Ito, R. and D.R. Preston. 1998. 'Identity, discourse, and language variation'. *Journal of Language and Social Psychology* 17(4): 465-483.

Jarrah, Ali. 1993. 'The phonology of Medina Hijazi Arabic: a non-linear analysis'. Unpublished PhD dissertation, University of Essex.

Johnson, D. E. 2009. 'Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis'. *Language and Linguistics Compass* 3: 359–383.

Johnson, K. 1993. 'Acoustic and auditory analyses of Xhosa clicks and pulmonics'. *UCLA Working Papers in Phonetics* 83: 33-45.

Johnson, J. S. and E. L. Newport. 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): 60-99.

Junghare, I. Y. 1983. 'Markers of definiteness in Indo-Aryan'. *Annual Meeting of the Berkeley Linguistics Society* 9: 116-127.

Kachru, Y. 2009. 'Hindi-Urdu'. In B. Comrie (ed) *The World's Major Languages*. New York: Routledge. 399- 416.

Katzner, K. and K. Miller. 2002. *The Languages of the World*. London: Routledge.

Kenstowicz, M. 1994. *Phonology in Generative Grammar*. Cambridge: Wiley.

Kerswill, P and S. Wright. 1990. 'The validity of phonetic transcription: Limitations of a sociolinguistic research tool'. *Language Variation and Change* 2(3): 255-275.

Kheshaifaty, H. 1997. 'Numerals: A comparative study between Classical and Hijazi Arabic'. *King Saud University* 9(1): 19-36.

Klaiman, M. 2009. 'Bengali'. In B. Comrie (ed) *The World's Major Languages*. New York: Routledge. 417- 436.

Koul, O. N. 2008. *Modern Hindi Grammar*. USA: Dunwoody Press.

Kostić, D. and R. S. Das. 1972. *A short outline of Bengali phonetics*. Calcutta: Statistical Publishing Society.

Krashen, S. 1981. *Second Language Acquisition and Second Language Learning.* University of Southern California.

Krashen, S. 1982. *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.

Krashen, S. D., C. M. Jones, S. J. Zelinski and C. Usprich. 1978. 'How important is instruction?'. *English Language Teaching Journal* 32(4): 257-261.

Krashen, S. D. and H. W. Seliger. 1976. 'The role of formal and informal environments in second language learning: A pilot study'. *International Journal of Psycholinguistics* 3:15-22.

Krashen, S., H. Seliger and D. Hartnett. 1974. 'Two studies in adult second language learning' *Kritikon Litterarum* 2/3: 220-228.

Kroskrity, P. 1999. 'Identity'. *Journal of Linguistic Anthropology* 9(1–2): 111–114.

Kusyk, M. and G. Sockett. 2012. 'From informal resource usage to incidental language acquisition: Language uptake from online television viewing in English'. *ASp. la Revue du GERAS* (62): 45-65.

Labov, W. 1966a. *The Social Stratification of English in New York City*. Washington: Centre for Applied Linguistics.

Labov, W. 1966b. 'The linguistic variable as a structural unit'. *Washington Linguistics Review* 3: 4-22.

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

Labov, W. 1981. 'Field methods of the project on linguistic change and variation'. *Sociolinguistic Working Paper Number 81*.

Labov, W. 1984. 'Field methods of the project on linguistic change and variation'. In J. Baugh and J. Sherzer (eds) *Language in use: Readings in Sociolinguistics*. Englewood Cliffs, NJ: Prentice Hall. 28–53.

Labov, W. 1990. The intersection of sex and social class in the course of linguistic change. *Language variation and change* 2(2): 205-254.

Labov, W. 1994. *Principles of Linguistic Change: Internal Factors.* Oxford: Blackwell.

Labov, W. 2001. *Principles of Linguistic Change Volume 2: Social Factors*. Oxford: Blackwell.

Lenneberg, E. 1967. *Biological Foundations of Language.* New York: Wiley and Sons.

Le Page, R.B. and A. Tabouret-Keller. 1985. *Acts of Identity: Creole-Based Approaches to Language and Ethnicity.* Cambridge: Cambridge University Press.

Liamputtong, P. 2006. *Researching the Vulnerable: A Guide to Sensitive Research Methods*. London: Sage Publications.

Littlewood, W. T. 1981. 'Language variation and second language acquisition theory 1'. *Applied Linguistics* 2(2): 150-158.

Llamas, C. 1999. 'A new methodology: Data elicitation for social and regional language variation studies'. *Leeds Working Papers in Linguistics and Phonetics* 7: 95–118.

Long, M. H. 1990. 'Maturational constraints on language development'. *Studies in Second Language Acquisition* 12(3): 251-285.

Loup, G., E. Boustagui, M. El Tigi and M. Moselle. 1994. 'Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment'. *Studies in Second Language Acquisition* 16(1): 73-98.

Lyrigkou, C. 2019. 'Not to be overlooked: Agency in informal language contact'. *Innovation in Language Learning and Teaching* 13(3): 237-252.

MacKenzie, L. 2019. 'Perturbing the community grammar: Individual differences and community-level constraints on sociolinguistic variation'. *Glossa: A Journal of General Linguistics* 4:1-23.

MacIntyre, P. D. 1998. 'Language anxiety: A review of the research for language teachers'. In D. J. Young (ed) *Affect in Foreign Language and Second Language Learning*. Boston: McGraw-Hill. 24-45.

MacIntyre, P. D. and R. C. Gardner. 1989. Anxiety and second-language learning: Toward a theoretical clarification. *Language Learning* 39(2): 251-275.

MacIntyre, P. and T. Gregersen. 2012. 'Affect: The role of language anxiety and other emotions in language learning'. In S. Mercer, S. Ryan and M. Williams (eds) *Psychology for Language Learning: Insights from Research, Theory and Practice* London: Palgrave Macmillan UK. 103-118.

MacIntyre, P. D. and R. C. Gardner. 1994. 'The subtle effects of language anxiety on cognitive processing in the second language'. *Language Learning* 44 (2): 283–305.

Major, Charisse Alaine. 2014. 'The effect of age on second language acquisition in older adults'. MA dissertation, Brigham Young University.

Manessy, G. 1977. 'Processes of pidginisation in African languages'. In A. Valdman (ed) *Pidgin and Creole Linguistics*. Bloomington: Indiana University Press. 129-154.

Marx, N. 2002. 'Never quite a 'native speaker': Accent and identity in the L2-and the L1'. *Canadian Modern Language Review* 59(2): 264-281.

Matras, Y. 2009. *Language Contact*. Cambridge: Cambridge University Press.

McGregor, R. S. 1976. *Outline of Hindi Grammar*. Oxford: Oxford University Press.

Meisel, J. 1975. 'The language of foreign workers in Germany'. *German in Contact with other Languages.* 184-212.

Meisel, J. M. 1977. 'The language of foreign workers in Germany'. In C. H. Molony, H. Zobl and W. Stolting (eds) *German in Contact with other Languages*. Kronberg: Scriptor. 184-212.

Mesthrie, R. 2008. 'Pidgins/creoles and contact languages: An overview'. In S. Kouwenberg and J. V. Singler (eds) *The Handbook of Pidgin and Creole Studies*. Oxford: Blackwell. 263-286.

Meyerhoff, M. and E. Schleef. 2010. 'Sociolinguistic methods for data collection and Interpretation'. In M. Meyerhoff and E. Schleef (eds) *The Routledge Sociolinguistics Reader.* London and New York: Routledge. 1–26.

Miller, C. 2007. 'Arabic urban vernaculars'. In C. Miller, E. Al-Wer, D. Caubet, C. Janet and E. Watson (eds) *Arabic in the City: Issues in Dialect Contact and Language Variation.* UK: Routledge. 1-31.

Milroy, L. 1987. *Language and Social Networks.* Oxford: Basil Blackwell.

Milroy, L. and Gordon, M. 2003. *Sociolinguistics: Method and interpretation*. Malden: Blackwell Publishing.

Miranda, J., F. Azocar, K.C. Organista, R. F. Munoz and A. Lieberman. 1996. 'Recruiting and retaining low-income Latinos in psychotherapy research'. *Journal of Consulting and Clinical Psychology* 64(5): 868-874.

Mirdehghan, M. 2010. 'Persian, Urdu, and Pashto: A comparative orthographic analysis'. *Writing Systems Research* 2(1): 9-23.

Montrul, S. 2000. 'Transitivity alternations in L2 acquisition toward a modular view of transfer'. *Studies in Second Language Acquisition* 22(2): 229-273.

Moore, L. W. and M. Miller. 1999. 'Initiating research with doubly vulnerable populations'. *Journal of Advanced Nursing* 30(5): 1034-1040.

Mougeon, R. and K. Rehner. 2001. 'Acquisition of sociolinguistic variants by French immersion students: The case of restrictive expressions, and more'. *The Modern Language Journal* 85(3): 398-415.

Mougeon, R., K. Rehner and T. Nadasdi. 2004. 'The learning of spoken French variation by immersion students from Toronto, Canada'. *Journal of Sociolinguistics* 8(3): 408-432.

Moyer, A. 1999. 'Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction'. *Studies in Second Language Acquisition* 21: 81-108.

Mühlhäusler. P. 1986. *Pidgin and Creole Linguistics*. Oxford: Blackwell.

Mühlhäusler. P. 1997. *Pidgin and Creole Linguistics*. London: University of Westminster Press.

Naess, Unn. 2008. 'Gulf Pidgin Arabic: Individual strategies or structured variety? a study of some features of the linguistic behaviour of Asian migrants in the Gulf countries'. Unpublished MA Dissertation, University of Oslo.

Naderifar, M., H. Goli and F. Ghaljaie. 2017. 'Snowball sampling: A purposeful method of sampling in qualitative research'. *Strides in Development of Medical Education.* 14(3): 1-4.

Nahedh, Monera. 1989. 'The sedentarisation of a Bedouin community in Saudi Arabia'. Unpublished PhD dissertation, The University of Leeds.

Naro, A. J. 1978. 'A study on the origins of pidginisation'. *Language* 54: 314-347.

Nasrin, M. B. and W. van der Wurff. 2009. *Colloquial Bengali: The Complete Course for Beginners.* London: Routledge.

Newing, H. 2011. *Conducting Research in Conservation: Social Science Methods and Practice*. New York: Routledge.

Nyamathi, A. 1998. 'Vulnerable populations: A continuing nursing focus'. *Nursing Research* 47(2): 65-66.

Obler, L. 1989. 'Exceptional second language learners'. In S. Gass, C. Madden, D. Preston and L. Selinker (eds) *Variation in Second Language Acquisition: Psycholinguistic Issues.* Clevedon: Multilingual Matters. 141-159.

Odlin, T. 1989. *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.

Odlin, T. 2003.'Cross-linguistic influence'. In C. Doughty and M.H. Long (eds) *The Handbook of Second Language Acquisition*. Malden: Blackwell Publishing. 436-486.

Oller Jr, J. W. and E. Z. Redding. 1971. 'Article usage and other language skills'. *Language Learning* 21(1): 85-95.

Omar, M. K. 1975. *Saudi Arabic, Urban Hijazi Dialect: Basic Course.* Washington, D.C: Foreign Service Institute.

Owens, J. 1997. 'Arabic-based pidgins and creoles'. In S. G. Thomason (ed) *Contact Languages: A Wider Perspective*. Amsterdam: John Benjamins Publishing. 125-172.

Oyama, S. 1976. 'A sensitive period for the acquisition of a non-native phonological system'. *Journal of Psycholinguistic Research* 5(3): 261-283.

Parkvall, M. 2017. 'Pidgin Languages'. In M. Aronoff (ed) *Oxford Research Encyclopedia of Linguistics.* Oxford: Oxford University Press. 1-30.

Parkvall, M. 2020. 'Pidgins'. In A. Grant (ed) *The Oxford Handbook of Language Contact*. Oxford: Oxford University Press. 261-281.

Patkowski, M. S. 1980. 'The sensitive period for the acquisition of syntax in a second language'. *Language Learning* 30(2): 449-472.

Patten, M. L. 2001. *Questionnaire Research: A Practical Guide*. Los Angeles: Pyrczak Pub.

Pavia, N., S. Webb and F. Faez. 2019. 'Incidental vocabulary learning through listening to songs'. *Studies in Second Language Acquisition* 41(4): 745-768.

Paul, I., Cortes. K and L. Milambiling. 2015. 'Definiteness without D: The case of ang and ng in Tagalog'. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 60(3): 361-390.

Penfield, W. and L. Roberts. 1959. *Speech and Brain Mechanisms*. New York: Atheneum Press.

Pereltsvaig, A. 2007. 'The universality of DP: A view from Russian'. *Studia Linguistica* 61(1): 59-94.

Phillips, E. M. 1992. 'The effects of language anxiety on students' oral test performance and Attitudes'. *The Modern Language Journal* 76(1): 14-26.

Piller, I. 2002. 'Passing for a native speaker: Identity and success in second language learning'. *Journal of Sociolinguistics* 6(2): 179-208.

Piske, T., I. R. MacKay and J. E. Flege. 2001. 'Factors affecting degree of foreign accent in an L2: A review'. *Journal of Phonetics* 29(2): 191-215.

Poplack, S. 1989. 'The care and handling of a megacorpus: The Ottawa-Hull French project'. In R. Fasold and D. Schiffrin (eds) *Language Change and Variation*. Amsterdam: John Benjamins Publishing. 411-444.

Pratkanis, A. R., S. J. Breckler and A. G. Greenwald. 1989. *Attitude Structure and Function*. Hillsdale, NJ: L. Erlbaum Associates.

Purcell, E. T and R. W. Suter. 1980. 'Predictors of pronunciation accuracy: A reexamination'. *Language Learning* 30(20): 271-287.

Quest, T. and C. A. Marco. 2003. 'Ethics seminars: Vulnerable populations in emergency medicine research'. *Academic Emergency Medicine* 10(11): 1294-1298.

R Core Team. 2012. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Racova, A. 1983. 'On defining the content and form of conjunctions in Bengali'. *Asian and African Studies* 19: 165-184.

Ray, S. P., M. Abdul Hai and L. Ray. 1966. *Bengali Language Handbook*. Massachusetts: Centre for Applied Linguistics.

Raza, A. A., S. Hussain, H. Sarfraz, I. Ullah and Z. Sarfraz. 2009. 'Design and development of phonetically rich Urdu speech'. In *Oriental COCOSDA International Conference on Speech Database and Assessments*. China. 38-43.

Regan, V. 2013a. 'Variation'. In J. R. Herschensohn and M. Young-Scholten (eds) *The Cambridge Handbook of Second Language Acquisition*. Cambridge: Cambridge University Press. 272-291 .

Regan, V. 2013b. 'The bookseller and the basketball player: Tales from the French Polonia'. In D. Singleton, V. Regan and E. Debaene (eds) *Linguistic and Cultural Acquisition in a Migrant Community*. Bristol: Multilingual Matters. 28-48.

Ringbom, H. 1986. 'Crosslinguistic influence and the foreign language learning process'. In M. S. Smith and E. Kellerman (eds) *Crosslinguistic Influence in Second Language Acquisition.* New York: Pergamon Press. 150-162.

Roberts, Nicholas. 2015. 'A sociolinguistic study of grammatical variation in Martinique French'. Unpublished PhD dissertation, Newcastle University.

Romaine, S. 1988. *Pidgin and Creole languages*. London; New York: Longman.

Romaine, S. 2003. 'Variation'. In .C. Doughty and M. H. Long (eds) *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell. 409-435.

Rosenhouse, J. 2006. 'Bedouin Arabic'. In K. Versteegh, M. Eid, M. Woidich and A. Zaborski (eds) *Encyclopaedia of Arabic Language and Linguistics.* Leiden: Brill. 259-269.

Ruane, J. M. 2005. *Essentials of Research Methods: A Guide to Social Science Research*. Malden, MA: Blackwell Publishing.

Russell, C. 1999. 'Interviewing vulnerable old people: Ethical and methodological implications of imagining our subjects'. *Journal of Aging Studies* 13(4): 403-417.

Rutter, E. 1931. 'The Hejaz'. *The Geographical Journal* 77(2): 97–108.

Sabourin, L., L. A. Stowe and G. J. De Haan. 2006. 'Transfer effects in learning a second language grammatical gender system'. *Second Language Research* 22(1): 1-29.

Salem, A. A. 2013. 'Linguistic Features of Pidgin Arabic in Kuwait'. *English Language Teaching* 6(5): 105-110.

Sankoff, G. 1974. 'A quantitative paradigm for the study of communicative competence'. In R. Bauman and J. Sherzer (eds) *Explorations in the Ethnography of Speaking.* Cambridge: Cambridge University Press. 18–49.

Schachter, J. 1974. 'An error in error analysis'. *Language Learning* 24(2): 205-214.

Schachter, P. and L. A. Reid. 2009. 'Tagalog'. In B, Comrie (ed) *The World's Major Languages*. London: Routledge. 833-855.

Schachter, P and F. T. Otanes. 1983. *Tagalog Reference Grammar*. Berkeley: University of California Press.

Schachter, P. 1976. 'The subject in Philippine languages: Topic, actor, actor topic, or none of the above'. In C. Li (ed) *Subject and Topic.* New York: Academic Press. 493-518.

Schilling-Estes, N. 2007. 'Sociolinguistic fieldwork'. In R. Bayley and C. Lucas (eds) *Sociolinguistic Variation Theories, Methods and Applications*. Cambridge: Cambridge University Press. 165- 236.

Schmidt, R. W. 1990. 'The role of consciousness in second language learning'. *Applied Linguistics* 11(2): 129-158.

Schmidt, R. L. 1999. *Urdu: An Essential Grammar*. New York: Routledge.

Schreier, D. 2008. *St Helenian English: Origins, Evolution and Variation*. Amsterdam: John Benjamins Publishing.

Schumann, J. H. 1976. 'Second language acquisition: The pidginisation hypothesis'. *Language Learning* 26(2): 391-408.

Schumann, J. H. 1978. 'The relationship of pidginisation, creolisation and decreolisation to second language acquisition'. *Language Learning* 28(2): 367-379.

Schwartz, B. D. and R. A. Sprouse. 1994. 'Word order and nominative case in non-native language acquisition: A longitudinal study of (L1 Turkish) German interlanguage'. In T. Hoekstra and B. D. Schwartz (eds) *Language Acquisition Studies in Generative Grammar*. Amsterdam: John Benjamins.

Schwartz, B. D. and R. A. Sprouse. 1996. 'L2 cognitive states and the full transfer/full access model'. *Second Language Research* 12(1): 40-72.

Schwartz, J. 1998. 'Testing a proposed universal'. In S. Gass and J. Schachter (eds) *Linguistic Perspectives on Second Language Acquisition.* Cambridge: Cambridge University Press. 73-88.

Schwarz, M. 2013. 'Learning with Lady GaGa & CO: Incidental EFL vocabulary acquisition from pop songs'. *VIEWS* 22: 17–48.

Scovel, T. 1988. *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech*. Rowley, MA: Newbury House.

Sebba, M. 1997. *Contact Languages: Pidgins and Creoles*. London: Macmillan International Higher Education.

Selinker, L. 1972. 'Interlanguage'. *International Review of Applied Linguistics* 10: 209–231.

Selinker, L. 1974. 'Interlanguage' In J. C. Richards (ed) *Error Analysis: Perspectives on Second Language Acquisition*. London: Longman. 31- 54.

Seright, L. 1985. 'Age and aural comprehension achievement in francophone adults learning English'. *TESOL Quarterly* 19(3): 455-473.

Siegel, J. 2008. *The Emergence of Pidgin and Creole Languages*. Oxford: Oxford University Press.

Sieny, M. E. 1973. *The Syntax of Urban Hijazi Arabic*. Lebanon: Librairie du Liban.

Silverstein, M. 1972. 'Chinook Jargon: Language contact and the problem of multi-level generative systems 1'. *Language* 48: 378-406.

Sin, C. H. 2005. 'Seeking informed consent: Reflections on research practice'. *Sociology* 39(2): 277-294.

Singh, I. 2000. *Pidgins and Creoles: An Introduction*. London: Oxford University Press.

Skehan, P. 1986. 'The role of foreign language aptitude in a model of school learning'. *Language Testing* 3(2): 188-221.

Smart, J. R. 1990. 'Pidginisation in Gulf Arabic: A first report'. *Anthropological Linguistics* 32: 83-119.

Sockett, G. 2014. *The Online Informal Learning of English*. Basingstoke, Hampshire: Macmillan Publishers.

Stone, T. H. 2003. 'The invisible vulnerable: The economically and educationally disadvantaged subjects of clinical research'. *Journal of Law, Medicine and Ethics* 31(1): 149-154.

Suter, R. W. 1976. 'Predictors of pronunciation accuracy in second language learning'. *Language Learning* 26(2): 233-253.

Syed, Saurov. 2017. 'The structure of noun phrases in Bengali: What it tells us about phases and the universality of DPs'. Unpublished PhD dissertation, University of Southern California.

Syed, S. and A. Simpson. 2017. 'On the DP/NP status of nominal projections in Bangla: Consequences for the theory of phases'. *Glossa: A Journal of General Linguistics* 2: 1-24.

Tagliamonte, S. 2006. *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Tagliamonte, S. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation.* Oxford: Wiley-Blackwell.

Tarone, E. and B. Parrish. 1988. 'Task-related variation in interlanguage: The case of articles'. *Language Learning* 38(1): 21-44.

Thompson, I. 1991. 'Foreign accents revisited: The English pronunciation of Russian immigrants'. *Language Learning* 41(2): 177-204.

Thompson, H. R. 2012. *Bengali*. Amsterdam: John Benjamins Publishing Co.

Todd, L. 1990. *Pidgins and Creoles.* London: Routledge.

Trask, L. 2000. *The Dictionary of Historical and Comparative Linguistics.* Edinburgh: Edinburgh University Press.

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

Trudgill, P. 2009. 'Contact, isolation, and complexity in Arabic'. In E. Al-Wer and R. de Jong (eds) *Arabic Dialectology*. Leiden: Brill. 173–186.

Ulbrich, C. 2021. 'Everyone is different, so everyone is the same–intra-individual variation in second language acquisition'. In A. Werth, L. Bülow, S. E. Pfenninger and M. Schiegg (eds) *Intra-Individual Variation in Language*. Mouton: De Guyter. 47-86.

Uritescu, D., R. Mougeon, K. Rehner and T. Nadasdi. 2004. 'Acquisition of the internal and external constraints of variable schwa deletion by French immersion students'. *International Review of Applied Linguistics in Language Teaching* 42(4): 349-364.

Vainikka, A. and M. Young-Scholten. 1994. 'Direct access to X'-theory: Evidence from Korean and Turkish adults learning German' In T. Hoekstra and B. D. Schwartz (eds) *Language Acquisition Studies in Generative Grammar*. Amsterdam: John Benjamins. 265-316.

Vainikka, A. and M. Young-Scholten. 1996a. 'The early stages in adult L2 syntax: Additional evidence from Romance speakers'. *Second Language Research* 12(2): 140-176.

Vainikka, A. and M. Young-Scholten. 1996b. 'Gradual development of L2 phrase structure'. *Second Language Research* 12(1): 7-39.

Van Liempt, I. and V. Bilger. 2012. '21 Ethical challenges in research with vulnerable migrants'. In C. Vargas-Silva (ed) *Handbook of Research Methods in Migration*. Cheltenham: Edward Elgar Publishing. 451-466.

Van Liempt, I. and V. Bilger. 2018. 'Methodological and ethical dilemmas in research among smuggled migrants'. In R. Zapata-Barrero and E. Yalaz (eds) *Qualitative Research in European Migration Studies.* Cham: Springer. 269-285.

Velupillai, V. 2015. *Pidgins, Creoles and Mixed Languages: An Introduction*. Amsterdam: John Benjamins.

Versteegh, K. 1984. *Pidginisation and Creolisation: The Case of Arabic*. Philadelphia: John Benjamins Company.

Wenger, G. C. 2002. 'Interviewing older people'. In J.F. Gubrium and J.A. Holstein (eds) *Handbook of Interview Research: Context and Method.* Thousand Oaks: Sage Publications. 259–278.

Weston, K. 2004. 'Fieldwork in lesbian and gay communities'. In S. Hesse-Biber and M. Yaiser (eds.) *Feminist Perspectives on Social Research*. Oxford: Oxford University Press. 198-205.

Whinnom, K. 1971. 'Linguistic hybridisation and the special case of pidgins and creoles'. *Pidginisation and Creolisation of Language*. 91: 115-134.

White, L. 1989. *Universal Grammar and Second Language Acquisition*. Amsterdam: John Benjamins.

White, L. 1991a. 'Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom'. *Second Language Research* 7(2): 133-161.

White, L. 1991b. 'The verb-movement parameter in second language acquisition'. *Language Acquisition* 1(4): 337-360.

White, L. 1992. 'Long and short verb movement in second language acquisition'. *Canadian Journal of Linguistics* 37(2): 273-286.

White, L. 1996. 'Clitics in L2 French'. In H. Clahsen (ed) *Generative Perspectives on Language Acquisition: Empirical Findings, Theoretical Considerations, Crosslinguistic Comparisons*. Amsterdam: John Benjamins. 335-368.

White, L. 2000. 'Second language acquisition: From initial to final State'. In: J. Archibald (ed) *Second Language Acquisition and Linguistic Theory.* Oxford: Blackwell. 130-155.

Willim, E. 2000. 'On the grammar of Polish nominals'. In R. Martin, D. Michaels and J. Uriagereka (eds) *Step by Step*. Cambridge: MIT Press. 319-346.

Winitz, H., Gillespie, B and J. Starcev. 1995. 'The development of English speech patterns of a 7-year-old Polish-speaking child'. *Journal of Psycholinguistic Research* 24 (2): 117-143.

Wiswall, Abdul-Qadir. 2002. 'Gulf pidgin: An expanded analysis'. Unpublished pro-seminar paper. University of Ohio, Linguistic Department.

Wittenburg, P., H. Brugman, A. Russel, A. Klassmann and H. Sloetjes. 2006. 'ELAN: A professional framework for multimodality research'. In *Proceedings of LREC*. Genoa. 1556-1559.

Wolf, C. and E. Jimenez. 1979. 'A sound change in progress: Devoicing of Buenos Aires /Z/'. MS: University of Pennsylvania.

Wolfram, W. and R. W. Fasold. 1974. *The Study of Social Dialects in American English*. Englewood Cliffs: Prentice-Hall.

Woods, Howard. 1979. ‘A socio-dialectology survey of the English spoken in Ottawa: A study of sociological and stylistic variation in Canadian English’. Unpublished PhD dissertation, University of British Columbia.

Young, R. 1991. *Variation in Interlanguage Morphology.* New York: Peter Lang.

Young, R. and R. Bayley. 1996. ‘VARBRUL analysis for second language acquisition’. In R. Bayley and D. R. Preston (eds) *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins. 253–306.

Zlatic, Larisa. 1997. ‘The structure of the Serbian noun phrase’. Unpublished PhD dissertation, The University of Texas at Austin.

Zobl, H. and J. Liceras. 1994. ‘Functional categories and acquisition orders’. *Language Learning* 44(1): 159-180.

Arabic Reference:

السوسوة، ع. ٢٠٢٢. " اللغة الهجين: دراسة لسانية اجتماعية (جنوب المملكة العربية السعودية أنموذجا)". في غ. جريس (محرر) *القول المكتوب في تاريخ الجنوب*. الرياض: مطابع الحميضي. ٣٠١-٣٥٨.

Appendices

Appendix A: The Sample Distributed across particular Predictors.

Table 43: The number of participants categorised by formal Arabic instruction in their country, L1 and sex

L1	Sex	Formal Arabic instruction in their country	
		Mostly yes	Mostly no
L1 Bengali speakers	Male	3	2
	Female	4	1
L1 Hindi-Urdu speakers	Male	1	3
	Female	2	3
L1 Tagalog speakers	Male	2	3
	Female	zero	5

Table 44: The number of participants categorised by their identity, L1 and sex.

L1	Sex	Identity of the migrants	
		Temporary migrants	Permanent migrants
L1 Bengali speakers	Male	4	1
	Female	1	4
L1 Hindi-Urdu speakers	Male	4	1
	Female	4	1
L1 Tagalog speakers	Male	5	zero
	Female	3	2

Table 45: The number of participants categorised by their educational level in L1, L1 and sex.

L1	Sex	Educational level in L1		
		Primary school	High school	University and above
L1 Bengali speakers	Male	4	zero	1
	Female	3	2	zero
L1 Hindi-Urdu speakers	Male	2	1	2
	Female	4	1	zero
L1 Tagalog speakers	Male	zero	zero	5
	Female	zero	4	1

Appendix B: Interview Questions

1. What do you do every day? When do you sleep? What time do you get up here in Saudi Arabia and in your own country? What do you do at the weekend in Saudi Arabia and in your country?
2. Do you have any friends or relatives here in Saudi Arabia? In which city? Do you visit them? How often?
3. How did you find out about your current job? Who told you about it? How did you apply for it? How much did your ticket cost to come to Saudi? Do you like your job or would you like to change it?
4. Have you performed Hajj? If you have performed it, how many times? Tell me about your experience. If you have not performed it yet, would you like to perform it? When? How about Umrah?
5. Do you have a driving licence? If yes, when did you pass? Who taught you? How long did it take you to learn to drive? How did you learn? Is petrol cheaper or more expensive in your country compared to Saudi Arabia?
6. Where do you live? Do any of your friends or relatives live with you? If a friend lives with you, what is their nationality? How many rooms does your house have?
7. What is the most common food in your country? How do you prepare it? Do you know any restaurant here in Saudi Arabia that makes this particular food? Is it cheap or expensive?
8. What is your favourite food? How do you prepare it?
9. What do you eat for breakfast, lunch and dinner?
10. Do you like spicy food? Do your children eat spicy food? If my kids eat spicy food, they will cry. How about your kids?
11. Do you like Saudi food? What do you like? Can you make it or do you have to buy it?
12. Do you have kids? How many? Do they go to school? What year are they in? What time does school start and what time does it finish in your country? What do children do after school? When do they go to sleep?
13. How did you learn Arabic? How long did it take you to learn Arabic?
14. Do you have siblings? How many? How old are they? What do they do?
15. What is your job in your country? Tell me about your job? How many hours do you work?

16. As you live here in Saudi Arabia, how do you contact your family? Which app do you use to contact your family? WhatsApp or Line? How often do you call them? Do you call them every day? For how long?

17. How much do you top up your mobile phone with each month? Do you have unlimited gigabytes?

18. Are you planning to buy a house or anything else in your country in the future? Have you bought a house? A farm? Any animals such as cows?

19. Have you ever seen someone steal something? Have you ever heard about a family or someone whose house was burgled and a thief stole something? Please tell me the story. What did they steal? Where and when? Did they call the police?

20. If I would like to visit your country, where would you tell me to visit? How many hours does it take from Saudi Arabia to your country by plane?

Appendix C: English Version of the Questionnaire

School of English Literature,
Language and Linguistic
Percy Building
Newcastle University.
Newcastle upon Tyne
NE1 7RU

Migrant workers' questionnaire

Project title	Migrant workers' experience of living in Saudi Arabia
Researcher	Abrar Bazerbay

Thank you for taking your time to complete this questionnaire. I really appreciate your participation. Please answer the following questions as honestly as possible. Your answers will be used for research purposes only and will be confidential and pseudonymised.

Section one: Personal information:

1. How old are you?…….
2. Are you …….
   a. male b. female c. non-binary d. prefer not to say
3. What is your first (native) language? ……
4. When did you arrive in Madinah (yyyy)? ……..
5. How long have you lived in Saudi Arabia?.....
6. How long have you lived in Madinah?.....
7. What is your job?…….

8. Has your employment changed since coming to Saudi Arabia? Yes\No..... If yes, what was it before?.....
9. Have you worked in any country other than Saudi Arabia? Yes\No…… If yes, which country?.....
10. Have you worked in Makkah, Jeddah or Taif before? Yes\No……., if so which city?..... How long did you stay there?……..
11. Have you worked in any city in Saudi Arabia other than Madinah, Makkah, Jeddah and Taif? Yes\No …….. If yes which city?..... How long did you stay there? ……..
12. What is your highest academic qualification? (ex. Primary school, did not complete high school, high school, Bachelor's degree, etc)……..
13. Do you watch TV in Arabic? Yes\No……. If yes, how many hours every day?.....
14. Do you listen to the radio in Arabic? Yes\No…… If yes, how many hours every day?.....
15. Do you listen to Quran every day? Yes\No……. If yes, how many hours every day?.....
16. Can you read the Quran? Yes\No.....

Section two: Arabic Language ability:

Rate your ability to speak Arabic on a scale from 1 = I cannot speak Arabic at all, to 10= I speak Arabic like a Saudi native speaker.

1. When you arrived in Saudi Arabia, how did you rate your Arabic speaking level?

   1 . 2. 3. 4. 5. 6. 7. 8. 9. 10

2. How would you rate your Arabic speaking level now?

   1 . 2. 3. 4. 5. 6. 7. 8. 9. 10

3. When you were <u>at school</u>, how many hours of Arabic classes or Quran classes did you have per week?
   a. When I was between 7 and 10 years old, I had …… hours per week.

b. When I was between 11 and 14 years old, I had …… hours per week.
c. When I was between 15 and 18 years old, I had …… hours per week.

4. When you were outside school, how many hours of Arabic classes or Quran classes did you have per week?
   a. When I was between 7 and 10 years old, I had …… hours per week.
   b. When I was between 11 and 14 years old, I had …… hours per week.
   c. When I was between 15 and 18 years old, I had …… hours per week.

5. When you left school, did you have any Arabic or Quran classes? Yes\No……If yes, between leaving school and coming to Saudi Arabia how many hours of Arabic or Quran classes did you have a week?.....

6. Since coming to Saudi Arabia, how many hours of Arabic or Quran classes have you had approximately?.....

7. Since coming to Saudi Arabia, how regularly you have attended Arabic or Quran classes?.....
   a. Nothing
   b. Every day
   c. Once a week
   d. Once a month

Which choice best describes how much you speak Arabic and your first (native) language on average each day in the following situations: Please choose one option.

8. When you are at home.
   a. 100% your first language, 0% Arabic.
   b. 75% your first language, 25% Arabic.
   c. 50% your first language, 50% Arabic.
   d. 25% your first language, 75% Arabic.

e. 0% your first language, 100% Arabic.

9. When you are socialising with your friends.
    a. 100% your first language, 0% Arabic.
    b. 75% your first language, 25% Arabic.
    c. 50% your first language, 50% Arabic.
    d. 25% your first language, 75% Arabic.
    e. 0% your first language, 100% Arabic.

10. When you are at your place of work.
    a. 100% your first language, 0% Arabic.
    b. 75% your first language, 25% Arabic.
    c. 50% your first language, 50% Arabic.
    d. 25% your first language, 75% Arabic.
    e. 0% your first language, 100% Arabic.

11. Which of the following bests describe you? Please choose one option.
    a. In the next year, I plan to return to my country of origin forever.
    b. In the next two years, I plan to return to my country of origin forever.
    c. In the next five years, I plan to return to my country of origin forever.
    d. In the next ten years, I plan to return to my country of origin forever.
    e. I plan to remain in Saudi Arabia forever.
    f. At the moment, I have no plans to return to my country of origin.

Section three: Attitude

On a scale from 1 to 7, please rate your agreement with the following sentences (1= strongly disagree; 7= strongly agree) Please choose one option

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

1. Madinah is a good place to live.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

2. I could get a more interesting job if my Arabic language skills were better.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

3. The Saudi people in Madinah are honest and trustworthy.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

4. Saudi people in Madinah are kind and friendly.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

5. I like the Arabic variety of Saudi people in Madinah.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

6. Good Arabic language skills are important for my job.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

7. The ability to write in Arabic is important for me.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

8. When I have to speak Arabic, I feel nervous in case people cannot understand me.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

9. If I do not speak Arabic accurately, I can still be part of the local community.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

10. I do not worry about the way I speak Arabic in public.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

11. I would like my Arabic language skill to be more like that of the Saudi people in Madinah.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

12. Being proficient in Arabic makes Saudi people respect me.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

13. Being proficient in Arabic will help me to communicate more easily with people.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

14. I am worried that people will not understand my Arabic when I am on the telephone.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

15. Being proficient in Arabic can help me to be successful in my life.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

16. Speaking Arabic will help me to better understand native speakers' way of life.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

17. I am self-conscious about my Arabic.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

18. Being able to read in Arabic is not important for me.

Strongly disagree 1. 2. 3. 4. 5. 6. 7. Strongly agree

Thank you for completing this questionnaire. If you have any questions or would like to withdraw from the study do not hesitate to contact me or my supervisors. Our contact details are on the participant's information sheet

Appendix D: English Version of the Consent Form.

Consent Form

Title of Study: Migrant workers' experience of living in Saudi Arabia

Researcher: Abrar Bazerbay

Thank you for your interest in taking part in this research. Please complete this form after you have read the Participant Information Sheet and listened to an explanation about the research study. You will be given a copy of this consent form.

Please add a tick if you agree with each statement.

1.	I confirm that I have read the participant information sheet for the above study. I have had the opportunity to consider the information, ask questions and I have had any questions answered satisfactorily.	
2.	I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. I understand that if I decide to withdraw, any data that I have provided up to that point will be omitted.	
3.	I understand that I have the right to refuse to answer any question or to stop the recording without providing a reason.	
4.	I understand that my recording may be used in future research by the named researcher; pseudonymised and de-identified data extracts may be quoted in published work and used for teaching purposes or played at conferences.	
5.	I understand that my data will be kept completely confidential and pseudonymised.	

6.	I understand that my pseudonymised research data may be published as a report.	
7.	I understand that only the researcher and her supervisors will have access to my data. Access to recordings will be password-protected.	
8.	I consent to being *audio* recorded and understand that my personal information and my recording will be transcribed and securely stored under a pseudonym as an electronic file on Newcastle University file storage. Hard copies of my data will be kept in a securely locked cabinet. All of the data will be used for research purposes only.	
9.	I understand that the interview will be recorded on multiple electronic devices and will be deleted from these devices as soon as it has been transferred to secure file storage. I will also delete the recording immediately from my mobile phone after sending it to the researcher.	
10.	I voluntarily agree to participate.	

Participant	Name	Signature
Interviewer	Name	Signature
Date		

Thank you for your participation!

Appendix E: English Version of the Participant Information Sheet.

Participant Information Sheet

Title of Study: Migrant workers' experience of living in Saudi Arabia

Invitation and brief summary

I invite you take part in this research study. Before you decide whether or not you wish to take part it is important that you understand why the research is being done and what it will involve. Please read this information carefully and discuss it with others if you wish. Take time to decide whether or not you wish to take part. If you do decide to take part, you will be asked to sign a consent form. However, you are free to withdraw at any time, without giving any reason and without any penalty.

Purpose of the study

The current study will investigate how migrant workers in Hijaz talk about their experience of living in Saudi Arabia

Participant selection

You are invited to participate in this project because you are: an adult migrant worker in Madinah or the Hijaz Region; you have not worked in any other Arabic cities or countries other than Madinah or the Hijaz Region before; and your first language is Hindi, Bengali or Tagalog.

What is involved?

I will conduct the interviews with females, whilst my male assistant will conduct interviews with males. We will talk to you for approximately one hour about different topics that are related to your life in Madinah or the Hijaz Region, for instance your job, your friends etc.

This interview will be carried out at a time that you prefer. Due to Covid-19 and in adherence with Saudi government guidelines, interviews will be conducted on-line via Zoom, a cloud-based video communications app which allows us to set up virtual audio and video calling. To participate in

this study, you need to have any of the following devices: a smartphone, tablet, computer or laptop which has speakers and a microphone. We will not be able to remember what you said exactly; therefore, we ask your permission to record the conversation on the researcher's mobile phone. We will ask you to record the conversation with your mobile phone too in case our recording is not clear. We will give you the instructions on how to use Zoom and how to delete the recording from your phone too. You will upload the recording to a specially designed folder on my Newcastle University cloud storage (OneDrive) after the interview and then delete it immediately from your mobile phone. I will also record the conversation on Zoom. I will set up the Zoom meeting so the record will be in my machine and not on the cloud storage in Zoom. We will never use the video-recording which Zoom automatically produces and we will delete it immediately.

At the end of the interview, I will send you a questionnaire via email or postal mail if you do not have an email that you will need to complete. It will take approximately 20 minutes. You will then scan it and send it back to me. If you do not have a scanner, you can use mail.

Anonymity and confidentiality

Your data will be kept completely confidential. In any report or presentation that may arise from this study, your data will be pseudonymised. Your real name and identity will never be used or revealed. To refer to you, I will use a pseudonym, your non-identifying demographic information, such as (gender, age, etc) or an entire group membership. If you mention the names and identities of any people while you are talking, these names and identities will be pseudonymised.

Confidentiality and access to the data

Only the researcher and her supervisors will have access to your research data. The access to recordings will be password-protected. Your recording will be used for the project above and securely retained for future research by the named researcher.

Storage of data

Your personal information and your recordings will be transcribed and stored in a pseudonymised format as electronic files on OneDrive, which is a secure cloud storage at Newcastle University, UK. When the recordings are transferred to OneDrive, they will be

deleted from the researcher's mobile phone immediately. Similarly, Zoom videos will be deleted. I will also have a physical backup for the transcription; a written document of your speech during the conversation, which will be held in a securely locked cabinet in my home. If you would like to use mail to send the questionnaire and the consent form, I will save these hard copies in the same cabinet. However, soft copies will be saved on OneDrive. All of the data will be used for research purposes only.

Risks, benefits and voluntary participation

There are no expected risks related to your participation. You have the right to end the conversation and withdraw from the study at any time without giving any reason and without any penalty. If you decide to withdraw, any data that we have provided up to that point will be omitted. During the conversation, you have the right to decline to answer certain question, stop the recording, not to record part of what you say, or to delete part of it. Your participation contributes to our understanding of migrant workers discourse about their experience of life in Saudi Arabia.

Dissemination of the results

The recordings and the result of the study may be used in future research by the named researcher. Pseudonymised and de-identified data extracts may be quoted in published work and used for teaching purposes or played in conferences.

If you would like to receive the results of my study, please contact me. I would be happy to send you a summary at the end of the research.

Further information and contact details

If you have any questions, do not hesitate to contact me or my supervisors.

Researcher's contact details

Abrar Bazerbay

a.a.a.bazerbay2@newcastle.ac.uk.

0500055291

+44 (0) 7846944204

P.O. Box 3568

Almatar Street

Bani Haritha district

Madinah

Supervisors' contact information

Dr. Heike Pichler

Heike.Pichler@newcastle.ac.uk

+44 (0) 191 208 3519

School of English Literature, Language and Linguistics

Percy Building (Room 2.05A)

Newcastle University

Newcastle upon Tyne

NE1 7RU

Dr. Daniel Duncan

Daniel.Duncan@newcastle.ac.uk

+44 (0) 0191 208 5879

School of English Literature, Language and Linguistics

Percy Building

Newcastle University

Newcastle upon Tyne

NE1 7RU

Thank you in advance!

Appendix F: English Version of the Debriefing Sheet

Debriefing Sheet

Title of Study: Migrant workers' variety in Hijaz; Madinah.

Thank you for taking part in this study. Your participation is beneficial and contributes to our understanding of the Arabic language of migrant workers.

The main aim of this study is to establish what kind of communication system the Arabic language spoken by migrant workers is: a new variety of Arabic influenced by workers' first language, or a new language that combines elements of multiple different languages. We also want to establish whether all migrant workers' in Hijazi speak the same variety of Arabic, or whether there are differences depending on migrant workers' first language, how often they use Arabic, etc. Investigation of these questions will provide new insights into how language works, how languages vary and how they change.

If you have any complaints, concerns or questions about this research, or you would like to obtain a copy of the results once the study is complete, please feel free to contact me.

Researcher's contact details

Abrar Bazerbay
a.a.a.bazerbay2@newcastle.ac.uk.
0500055291
+44 (0) 7846944204

Thank you very much for participating!