# Gender Variation in Gulf Pidgin Arabic

Najah Albaqawi

A thesis submitted in partial fulfilment of the requirements of the University of Wolverhampton for the degree of Doctor of Philosophy

April 2020


This work and any part thereof have not previously been presented in any form to the University or to any other body whether for the purposes of assessment, publication or any other purpose (unless previously indicated). Save for any express acknowledgements, references and/or bibliographies cited in the work, I confirm that the intellectual content of the work is the result of my own efforts and of no other person.
The right of Najah Albaqawi to be identified as author of this work is asserted in accordance with ss.77 and 78 of the Copyright, Designs and Patents Act 1988. At this date, copyright is owned by the author.

Signature……

Date: ……

*To the souls of my Mom and Dad, with love and gratitude.*

*To all those who have died after contracting the COVID-19 virus.*

# ACKNOWLEDGEMENTS

All praise and thanks to Allah for helping me overcome what seemed to be insurmountable obstacles.

I would like to acknowledge my indebtedness and render my warmest thanks to my first supervisor, Dr. Michael Oakes, who made this work possible. I deeply appreciate his constant, diligent, and patient guidance as he read every word of my drafts. I would also like to thank my second supervisor, Prof. Ruslan Mitkov, for inspiring my interest in the development of innovative technologies.

The thesis has also benefited from comments and suggestions by Dr. Amanda Bloore, Mrs. Kate Wilson, Mrs. April Harper, and Miss Emma Franklin, all of whom have read through the manuscript. I take this opportunity to thank them. I am equally thankful and grateful to Dr. Almoaily; he was instrumental in defining the path of my research.

Special thanks are due to the participants who volunteered their time and made valuable contributions. Moreover, special thanks to the families in Riyadh who allowed me into their homes to interview their hired help.

Furthermore, I greatly appreciate and acknowledge the support received from the Saudi Cultural Bureau in London, the Ministry of Education in Saudi Arabia, and the Universities of Dammam and Hafr Albatin for their generous funding throughout my postgraduate studies.

Finally, I acknowledge the people who mean a great deal to me, namely, my siblings and friends, for showing faith in me. I salute you all for your valuable prayers, selfless love, care, and support during the past 4 years, all of which contributed immensely to the completion of my thesis.

I owe thanks to a very special person – my husband, *Dr. Mubarak Alotaibi* – for his continued and unfailing love, support, and understanding during my pursuit of this PhD degree that made the completion of the thesis possible. I greatly value his contribution and deeply appreciate his belief in me. I appreciate my dearly loved *children* for the patience they displayed during my thesis writing. Words will never express how grateful I am to all of you. I love you so much, and I am very proud of you.

# ABSTRACT

In the history of pidgins and creoles, many documented contact languages are European-based ones because they arose as a direct result of European colonial expansion between the sixteenth century and the first half of the twentieth century. However, contact languages are developing entirely outside the European context as a result of ongoing international migration and economic integration created by globalisation. One such newly emerging pidgin is known as Gulf Pidgin Arabic (GPA). This unique linguistic phenomenon is a simplified contact variety of the Arabic language used in the Gulf States for communication between native Arabic speakers and foreign workers, as well as among the workers themselves. Pidgin languages have not been studied until relatively recently, since the middle of the last century. Similarly, GPA has received relatively little attention in the literature, apart from a few descriptive works such as Abed (2017), Almoaily (2012), Avram (2014), Næss (2008), Smart (1990), and Wiswal (2002). Importantly, there is an increasing labour market demand for women migrants in the Gulf, and this demand is often more stable than that for men; however, no studies to date have investigated the gender and language variation in Gulf countries conditioned by length of stay or substrate language.

To carry out this research, an integrated research design, combining quantitative and qualitative phases of analysis, is employed to examine data drawn from one-to-one semi-structured interviews. Extensive background research on the Saudi social setting, the Pidgin languages, Gulf Arabic (GA) and GPA, and the major substrate languages of GPA is undertaken to investigate the sociolinguistic and linguistic situations that have resulted in the emergence of GPA. I analyse the influence of the first language of female GPA speakers and the number of years spent in the Gulf as potential factors conditioning language and gender variation in GPA. The dataset for the study consists of interviews with 72 informants from six linguistic backgrounds: Malayalam, Punjabi, Bengali, Tagalog, Sinhala, and Sunda. Interviews were conducted in Riyadh, the capital city of Saudi Arabia. Half of the informants had spent five years or less in the Gulf, while the other half had spent 10 years or more in the area at the time of interview. The analysis is based on 10 morphosyntactic phenomena: free or bound object or possessive pronoun, presence or absence of the Arabic definiteness marker, presence or

absence of Arabic conjunction markers, presence or absence of the GPA copula, and presence or absence of agreement in the verb phrase and the noun phrase.

Regarding the informants' choice of the studied morphosyntactic features, the results of this thesis demonstrate that the length of stay in the Gulf produces more accommodation to standard GA in women than men. However, this shift was significant for only one feature: conjunction markers. For the influence of the first language, a significant adaptation to the system of GA (the lexifier language) was found for two features: conjunction markers and nominal agreement. Furthermore, with years of stay in the Gulf, there was a significant shift for only two features: conjunction markers and definiteness. This finding could be taken to support both universalist theories and substrate theory of the emergence of contact languages. The two theories seem to have effects on the emergence of pidgins and creoles; it is worth noting that neither are separate from each other, and they can be complementary. Thus, my data supports Mufwene's (1993) complementary theory of genesis, which claims that universal as well as substratal factors can contribute to the emergence of contact languages.

# Table of Contents

# List of Tables and Figures

## List of Tables

## List of Figures

# Chapter 1: Introduction

## 1.1. Context and Motivations

*Gulf Pidgin Arabic* (GPA) is a simplified contact variety of the Arabic language used in the Gulf States for communication between native Arabic speakers and foreign workers (FWs), as well as among the workers themselves. In the field of linguistics, the term 'pidgin' can refer to many different things. Sakoda and Siegel (2003: 1) point out that pidgin 'refers to a new language that develops in a situation where speakers of different languages need to communicate but don't share a common language'. Based on their definition, the Gulf countries are the ideal situation for the birth of a contact language. Following the October 1973 'oil boom', the Arab Gulf States experienced radical social, political, and demographic changes in a short time. This led to a rapid increase in the demand for foreign labour, as the Gulf national workforces at that time were too small and without the required skills to execute large projects. Hence, during the 'oil decade' bonanza (1973–1982), the number of foreign labourers in the Gulf countries, especially the Kingdom of Saudi Arabia (KSA), quickly increased, amounting to almost 4.4 million in 1985 – a more than threefold increase within a single decade. According to a Saudi report, 'Migration Information Source',[1] more than 7 million immigrants from Asian countries work in Saudi Arabia (SA; Albakrawi, 2012: 127). Moreover, the kingdom is the largest economy in the Arab world, endowed with the world's second largest proven oil reserves. This makes SA a major hub for population movements and labour immigration (De Bel-Air, 2014: 3).

[1] Retrieved on 29 December 2019 from https://www.migrationpolicy.org/country-resource/saudi-arabia

As stated by Avram (2017: 61), the country has a multilingual setting, as do all Gulf countries; Gulf Arabic (GA) is a form of Colloquial Arabic spoken by the indigenous people of the Gulf Region (see Map 2.1 in Chapter 2). Migrant workers, who come from various linguistic backgrounds and usually do not speak Arabic, come into contact with GA speakers as well as speakers of other Arabic dialects, and there is an urgent need for communication between these two groups – 'Arabic-speaking locals and expats on one hand and non-Arabic speaking expats on the other' (Almoaily, 2012: 1). A simplified form of Arabic, known as 'Gulf Pidgin Arabic' (GPA), has thus developed as a result of this contact. Indeed, GPA and GA are two distinct forms of language, with lexical, phonological, syntactical, and morphological differences. It is worth noting here that, in addition to the indigenous vernacular GA, many other languages, language varieties, and registers are known by these workers and play a role in determining the characteristics of GPA. This situation has resulted in the emergence of various pidginised forms of Arabic across the Arab world, mainly in the Gulf area.

Due to the recently increasing number of FWs, mainly from the Indian subcontinent and South East Asia who travel to these countries to obtain jobs and who usually do not speak Arabic but have contact with speakers of GA and other Arabic dialects, the need for communication among these groups is currently increasing. Some scholars agree that P/Cs could have emerged as a result of imperfect L2 learning (DeGraff, 1999; Mufwene, 1990; Becker and Veenstra, 2003; Winford, 2003, and Field, 2004), while others, such as Siegel (2008b: 208), argue that 'while more creolists today may agree about the involvement of processes of second-language acquisition (SLA) in P/C genesis, there is no consensus about exactly what these processes are and how and when they apply'. More research on the role of language acquisition in P/C genesis must thus be conducted, since GPA is spoken by a non-indigenous workforce, in the Gulf, over a wide geographical area in a multi-ethnic speech community, where language variation seems inevitable.

This is additionally of concern because SA is a conservative society that is controlled by some restrictions due to sociocultural norms, such as gender segregation. The separation of the male and female genders could play a crucial role in the process of acquiring a language (Bakir, 2004). Having to lead gender-segregated lives results in an unequal distribution of women and men across occupations and educational establishments, including foreigners who live in such a conservative community. The variation in speech due to the differences in such a society (e.g. the existence of sexism and social status or power differences), male and female GPA speakers have limited access to the male public or to the female private sphere, thus leading to gender variation, as 'gender is performed in part through "embodied iteration" of particular linguistic acts' (Harrington, Litosseliti, Sauntson, and Sunderland, 2008: 4). Hence, due to the lack of studies involving the quantitative variationist analysis of GPA gender variability, performing such an investigation is essential.

### 1.2. Aims, Hypotheses, and Contributions

This thesis provides original contributions to a) the study of the language variation and change in GPA transcribed spoken documents, b) the emergence of this Arabic-based pidgin contact language variety in the Gulf, and c) methods of compiling and analysing a spoken corpus written in Arabic scripts. This thesis also studies the influence of social norms that have led to the emergence and development of GPA in the Gulf region, particularly in the KSA.

This section presents the aims and the original contributions of this thesis. Section 1.2.1 describes the aims of and the fundamental assumptions behind the research in this thesis. Then, Section 1.2.2 specifies the goals to be achieved, and Section 1.2.3 explains the contributions of this thesis to the knowledge of the subject under investigation.

#### 1.2.1. Thesis Aims and Research Hypotheses

This thesis has three main aims. They are as follows:

**The first aim** is to provide a concise morphosyntactic description for a cross-linguistic comparison of GA and GPA based on an illustration of the five targeted linguistic features resulted from interviews.

**The second aim** is to investigate the first potential factor in language variation in GPA: the first language (L1) of female speakers. This aim is motivated by both the large number of female FWs in the Gulf who come from various linguistic backgrounds and the fact that, to date, no account of language variation in GPA caused by differences in the morphosyntactic systems of the substrate languages of female GPA speakers has been studied.

**The third aim** is to investigate the second potential factor in language variation in GPA: the number of years of residency of female speakers in the Gulf. This aim is motivated by the fact that in spite of some female FWs having been in SA for more than 20 years, no study has yet investigated the effect of the length of stay in the Gulf on female GPA speakers.

Finally**, the fourth aim** is to investigate the third potential factor in language variation in GPA: male and female gender variation. This aim is motivated by the increasing labour market demand for women migrants in the Gulf; although this demand is often more stable than that for men, no studies to date have investigated the gender variation in Gulf countries.

For these aims to be achieved, the investigation is based on the following research hypotheses, which are tested further in the thesis:

**Hypothesis 1**: There are differences between standard GA and GPA. This hypothesis is tested through a qualitative investigation, which involves describing selected aspects of the morphosyntax of GPA (see Chapter 4); the pidgin under investigation in this project; and its lexifier, namely, GA (see Chapter 4). These features are agreement in the verb phrase (VP) and in the noun and adjective phrase (ADJP), definiteness and indefiniteness, pronouns, conjunctions, and copular verbs. I discuss each in turn in the following subsections.

**Hypothesis 2**: There is a difference between the GPA spoken by speakers with different L1s. This hypothesis is tested by conducting a quantitative analysis based on informants' use of the variants of the five selected morphosyntactic phenomena: definiteness, conjunctions, copulas, object and possessive pronouns, and verbal and nominal agreement.

**Hypothesis 3**: Length of stay in the Gulf produces accommodation to standard GA. This hypothesis is tested by conducting a comparative quantitative analysis of the data from newly settled female GPA speakers with that of speakers who have stayed longer in the Gulf. This hypothesis will answer the question of whether female GPA speakers shift to GA after spending some time in the Gulf.

**Hypothesis 4**: Length of stay in the Gulf produces more accommodation to standard GA in women than men. This hypothesis is tested by conducting a comparative quantitative analysis of my female data with that in Almoaily's (2012) male data. Examining the differences in the use of language among the variants of the five selected morphosyntactic phenomena of men and women who come from different linguistic backgrounds (such as Malayalam, Bengali, and Punjabi) based on length of stay in the Gulf will determine whether gender-related issues and social factors influence female GPA speakers' learning of the GA language in SA.

### 1.2.2. Thesis Objectives

The aims specified in Section 1.2.1 are achieved by setting and meeting the following goals:

**Goal 1** is to provide a concise morphosyntactic description for cross-linguistic comparison of GA, GPA, and the major substrate languages of GPA (Bengali, Punjabi, Malayalam, Sinhala, Tagalog, and Sunda) based on an illustration of the five targeted linguistic features (i.e. definiteness, conjunctions, copulas, object and possessive pronouns, and verbal and nominal agreement). This involves an easily accessed cross-linguistic comparison where the hypothesis can be formulated regarding differences in the use of the morphosyntactic features of each language.

**Goal 2** is to design and compile my own corpus from transcribed spoken interviews with female GPA speakers. The collected data is used to identify and analyse five morphosyntactic phenomena.

**Goal 3** is to examine any differences in the use of the five selected linguistic features in the speech of GPA speakers with different L1s (Bengali, Punjabi, Malayalam, Sinhala, Tagalog, and Sunda). The five selected linguistic features are reviewed and examined in the corpus of GA and GPA speakers.

**Goal 4** is to examine and compare the half of the data that was produced by informants who had spent five years or less in the Gulf with the other half from informants who had spent 10 or more years in the Gulf by the time they were interviewed. This involves comparing the data of newcomers to the Gulf area (e.g. GPA Tagalog speakers who had spent five years or less in the Gulf at the time I interviewed them) with that of long-term residents in the Gulf (i.e. those Tagalog who had spent 10 years or more in the Gulf at the time they were interviewed). This allows me to investigate the question of whether female GPA speakers actually shift towards GA after spending more than 10 years in the Gulf.

**Goal 5** is to review the available data produced by male and female speakers with different L1s. This involves noting any differences in the data due to gender variation. Examining the effect of length of exposure to GA, and whether it is greater in men or women, will involve comparing my results of female GPA speakers with those of Almoaily's (2012) male GPA speakers.

**Goal 6** is to examine competing theories – that is, the substratist and the universalist (including imperfect SLA) theories – to account for the genesis of P/C languages based on the evidence found in the data of this study.

**Goal 7** is to identify any weaknesses and limitations of the methodologies proposed in Goal 2 and to identify directions for improvement and future research.

### 1.2.3. Contributions of the Thesis

By achieving the goals set in Section 1.2.2, the thesis makes the following novel contributions to knowledge, which I hope will be positive contributions to the field of P/C studies:

The **first main original contribution** of this thesis is that the investigation is based on non-Indo-European (Arabic) input, whereas most previous studies of contact languages have been based on Indo-European languages. This study thus contributes to the literature of less-described non-Indo-European P/Cs. More must be learnt about non-Indo-European contact languages to discuss universalist theories, which are the most recent views on the origin of P/Cs and the most relevant to the current study. Universalist theories aim to explain structural similarities among the world's P/Cs, with the assumption that all humans are characterised by an innate ability to simplify language (Chomsky, 1965). The theories thus focus more on the role of innate cognitive principles in the process of P/C formation than on the influence of the languages in contact. If the present study reveals that the sampled informants have produced universal features of contact languages which cannot be traced to their L1s, and furthermore that divergent properties of their L1s do not have a significant effect on their production of GPA, then this study would support universalist theories of genesis. Furthermore, this study provides useful insights for researchers interested in language variation and change in general and the evolution of modern Arabic specifically.

The **second main original contribution** is that this thesis is the first quantitative variationist analysis study of female speech in GPA, and it is thus complementary to Almoaily's study of male speech in the GPA spoken in Gulf countries, making it an original linguistic contribution to studies on the relationship between gender and language variation in contact languages.

The **third main original contribution** of this thesis is an examination of how language and culture have an intertwined relationship in a deeply conservative community such as in SA. This study proposes a way to conceptualise how language and social factors interweave to

create a new linguistic system in a language-contact situation among a multi-ethnic speech community (in the case of GPA, the speech community consists of native people from the Arabian Gulf and Asian immigrants working in that region).

The **fourth original contribution** of this thesis is that it is the first variationist quantitative investigation of a) the influence of female GPA speakers' substrate languages (L1) with respect to the selected linguistic features on the production of GPA and b) the effect of number of years of stay in the Gulf on those speakers' choices as regards the studied morphosyntactic features, which make a valuable contribution to the field of (socio)-linguistic variation and change in contact languages.

The **fifth original contribution** of this thesis is the first corpus of a pidgin- or creole-Arabic-based language built from female speakers. This annotated corpus of female speech will be made freely available to all researchers who are interested in studying Arabic contact languages.

**The sixth original contribution** of this thesis is a detailed description and analysis of the morphosyntactic features of all substrate languages (L1 mother tongues) of the female speakers recorded in the corpus.

The **seventh original contribution** of this thesis is a set of suggestions for teaching Arabic as a second language (L2). The findings pertaining to the morphosyntax used by migrant workers and the reasons for this, will help teachers to know how the language of migrant workers should best be adapted to Standard Arabic. The study will also potentially uncover specific areas of difficulty for L2 speakers, thereby enabling teachers to pay more attention to these linguistic phenomena in their teaching practice and when preparing learning materials for their students, as they will have an enhanced understanding on what their students tend to struggle with.

I hope this project will contribute to bringing, to a larger extent, a globalised view on the investigation of contact languages by affording a greater degree of attention regarding this pidgin variety and how it has developed and functioned within its local social context.

### 1.3. Structure of the Thesis

The research study comprises nine chapters. **Chapter 2** provides an overview of the situation in Gulf countries, with a specific focus on the main setting of the research study and the position of women in Saudi Arabian society. **Chapter 3** then presents a critical overview of related work in the field of contact languages and their emergence, with a focus on previous works on GPA and the corpus linguistics field, while **Chapter 4** offers an extensive description of both GPA and its superstrate language, GA. **Chapter 5** is a descriptive account of the five selected linguistic features in the six substrate languages. In **Chapter 6**, the research methodology employed to achieve the aims of this thesis is explained, and the results are listed and discussed in **Chapters 7** and **8**. Finally, **Chapter 9** provides conclusions and recommendations for current and future research.

The contents of each chapter are described more concretely below.

This chapter (**Chapter 1**) introduces the context and motivations for this study and the assumptions behind this research. It also lists and describes the aims and goals to be achieved in this thesis and its original contributions, and finishes by presenting the structure of the thesis.

**Chapter 2** gives an overview of the sociolinguistic and socioeconomic situation of the Arabian Gulf region, where GPA is spoken. It provides a geographic and demographic overview of Saudi Arabia and describes Riyadh City as the main setting of the research study. Most importantly, this chapter will explore the position of women within Saudi culture. It provides information on how the existing Saudi society influences the quality of life of foreign labourers regarding the acquisition of language. In addition, the chapter discusses the impact of gender

roles and gender segregation and highlights the contribution of religion, socio-cultural norms, and traditions.

**Chapter 3** tackles some issues in the literature of pidgins and creoles concerning their definitions and emergence. In addition, it discusses theories of the genesis and development of pidgins and creoles. The chapter provides a sociohistorical overview of pidgins and creoles. It also investigates general linguistic features that characterize pidgin and creole languages based on both Indo-European and non-Indo-European contact languages. Moreover, this chapter discusses contact languages in the field of corpus linguistic and some pros and cons to using a corpus-based method.

**Chapter 4** provides a description of selected aspects of the morphosyntax of GPA, the pidgin under investigation in this project, and its lexifier, GA. The description will be restricted to the five selected morphosyntactic features: agreement in the verb phrase and in the noun and adjective phrase, definiteness and indefiniteness, pronouns, coordination, and copular verbs. With this, Goals 1 and 3 are partially accomplished.

**Chapter 5** provides an extensive description of the main six substrate languages of GPA under investigation: Malayalam, Punjabi, Bengali, Sinhala, Tagalog, and Sunda. The chapter presents a discussion on substrate languages and the methodology in determining the six languages as they have the largest number of speakers in the Gulf region. In addition, Chapter 5 describes the morphosyntactic features relevant to this project and provides a brief sociolinguistic background of the languages under investigation (Malayalam, Punjabi, Bengali, Sinhala, Tagalog, and Sunda). A comparison between GA, GPA, and the substrate languages described in this chapter will be drawn at the end of this chapter. Moreover, Chapter 5 summarizes similarities and differences between the substrate languages and whether they could play a role in the variation between GPA speakers. With this, Goals 1 and 3 are completely accomplished.

**Chapter 6** provides a detailed description of the pilot and main stages of the current study. The

chapter consists of three parts. Sections 6.1 and 6.2 describe the purpose of this study and illustrate the structure of the corpus built to examine it. Sections 6.3 through 6.6 discuss the process of compiling the corpus, i.e. the sampling strategy for the pilot and main study, the data collection instruments, and the transcribing of interviews, as well as the procedures followed in counting and annotating the tokens. Section 6.7 discusses issues and difficulties that the researcher encountered while conducting the pilot and main study. In doing so, it achieves Goal 2.

**Chapter 7** presents the findings of the fieldwork. Each language group was split into two groups based on their length of stay in Saudi Arabia or any other GA speaking country. Those who had been in the Gulf for five years or less are referred to as "New" speakers, while those who had been residents for ten years or more are referred to as "Old" speakers. Chi-squared tests were run to establish the significance of the effect of the informants' L1 and of years of residency in the Gulf on gender and language variation in GPA. In doing so, it achieves Goal 4.

**Chapter 8** discusses the key findings from the GPA speakers' data in the light of the hypotheses listed in Section 4.1.4. It discusses the results of the research in relation to previous theories of genesis on the literature of pidgin and creole languages. This chapter discusses potential universal and substratal factors which have led to the emergence of the patterns evident in the data. Moreover, it discusses gender-related issues and socio-cultural norms in the lives of GPA speakers in Saudi Arabia, and explores how it could have possibly influenced their acquisition of the GA language. In doing so, it achieves Goal 5 and Goal 6.

**Chapter 9** is the final chapter of the thesis. It reviews the extent to which the goals set in Chapter 1 are met and provides details about the original final contributions of this research, as well as a review of the thesis. It also provides set of suggestions for teaching Arabic as a second language and set of possible directions for future work on the basis of the weaknesses

discovered while conducting the research described in the previous chapters. In doing so, it achieves Goal 7.

# Chapter 2: Pidgin Arabic in the Arabian Gulf Region

## 2.1. Introduction to Chapter 2

The aim of this chapter is to introduce the Arabian Gulf region, where GPA is spoken, and to provide an overview of SA, with special emphasis on the city of Riyadh[1] as the main setting of the research study. The chapter offers an overview of the sociolinguistic and socioeconomic situation, as well as the geographics and demographics of the region. Most importantly, it explores the position of women within Saudi culture. The issue of gender-associated variation in language in the Gulf States, particularly in SA, has hitherto been neglected. Relevant literature regarding the influence of gender on SLA is discussed to provide a broader perspective on the conditions and characteristics that affect language use generally and specifically in a non-Western society such as SA. Having now launched Saudi Vision 2030[2] – the new Crown Prince Mohammad bin Salman bin Abdulaziz Al Saud's blueprint for opening up SA to the modern world – the conservative Islamic KSA is undergoing great changes in its culture and economy. Much more important now is how women can be better integrated into the labour market. This deeply symbolic move will transform women's lives in a country where work is seldom a mixed-gender environment. This chapter presents recent information about SA and the social factors that have influenced the situation and role of women in Saudi society.

[1] Riyadh coordinates: 24° 46' 27.3540" N and 46° 44' 18.9096" E
[2] A full description of Saudi Vision 2030 is available in Arabic and English at https://vision2030.gov.sa/en and https://www.alarabiya.net/ar/aswaq/economy/2016/04/25/2030-رؤية-المملكة-العربية-السعودية-تفاصيل.html

## 2.2. Overview of the Arabian Gulf Region

The Arab Gulf States are countries aligned as the Gulf Cooperation Council (GCC) and include SA, Kuwait, the United Arab Emirates (UAE), Bahrain, Qatar, and Oman. Arabic is the official language in all the Arabian Gulf States; however, two forms of Arabic are used side by side: the higher variety, Standard Arabic and the lower variety, GA. On the one hand, Standard Arabic is the language perceived as influential and prestigious, and it is used in writing as well as in formal oral settings (e.g. education, religious, and government areas). GA, on the other hand, is mainly spoken but rarely written. It is used for informal purposes as the everyday language (e.g. at home, with friends).

Standard Arabic is not an L1 for citizens of the Arabian Gulf region. In general, only educated people can speak it spontaneously; other indigenous Gulf people can only comprehend it. As a result, GA is a specific speech community whose speakers can easily understand one another. According to Ferguson (1996: 55), a speech community is a

> social group sharing features of language structure, use and attitudes that functions as a sociolinguistic unit for the operation of linguistic variation and/or change; it may be monolingual or multilingual (Ferguson 1978), and it may be at any level of abstraction for which the definition holds.

Moreover, there are also non-Arabic speakers – mainly temporary FWs who need to communicate with other people in the Gulf region. English and GPA are the most common languages for this purpose, as found by Smart (1990). The geography, demographics, and economy of the region have contributed to shaping the variety of Arabic which is known as GA and, more recently, also GPA.

## 2.3. The Importance of Oil in the Emergence of Gulf Pidgin Arabic

Following the October 1973 ‘oil boom’, the Arab Gulf States (the GCC) experienced radical social, political, and demographic changes in a short time. This led to a rapid increase in the demand for foreign labour, as the Gulf national workforces at that time were too small and without the required skills to execute large projects. Hence, during the ‘oil decade’ bonanza

(1973–1982), the number of foreign labourers in the Gulf countries, especially the KSA, rapidly rose, amounting to almost 4.4 million in 1985 – a more than threefold increase within a single decade. According to a Saudi report, namely, Migration Information Source, more than 7 million immigrants from Asian countries work in SA (Albakrawi, 2012; see Table 5.1). The kingdom is also the largest economy in the Arab world, endowed with the world's second largest proven oil reserves (see Figure 2.1[3]). This makes SA a major hub for population movements (De Bel-Air, 2014), and like all Gulf countries, it has a multilingual setting (Avram, 2017); GA is a form of Colloquial Arabic spoken by the indigenous people of the Gulf Region (see Map 2.1). Migrant workers, who come from various linguistic backgrounds and usually do not speak Arabic, come into contact with GA speakers as well as speakers of other Arabic dialects, and there is an urgent need for communication between the two groups: 'Arabic-speaking locals and expats on one hand and non-Arabic speaking expats on the other' (Almoaily, 2012: 1).

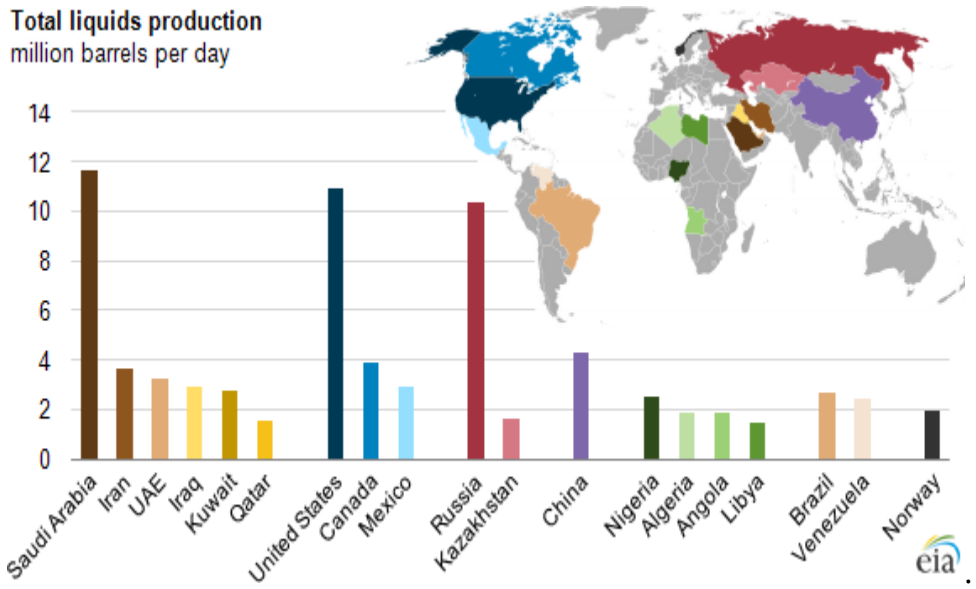



Figure 2.1: Total petroleum production (Source: U.S. Energy Information Administration, International Energy Statistics, and Short-Term Energy Outlook).

[3] Retrieved on 10 March 2016 from https://www.eia.gov/todayinenergy/detail.php?id=10231

A simplified form of Arabic (i.e. GPA) has thus developed as a result of this contact. GPA is a reduced system of language that is used for communication between FWs and native speakers of Arabic. Indeed, GPA and GA are two distinct forms of language, with lexical, phonological, syntactic, and morphological differences. It is worth noting here that, in addition to the indigenous vernacular GA, many other languages, language varieties, and registers are known by these workers and play a role in determining the characteristics of GPA. This situation has resulted in the emergence of various pidginised forms of Arabic across the Arab world, mainly in the Gulf area. The reasons that Asian expatriates working under the same conditions throughout the Gulf States have apparently developed dissimilar pidgins are still to be determined (Bizri, 2014). Recent research has yielded linguistic data on particular Asian varieties of Arabic which, upon examination, report many common features as well as some differences. Throughout the thesis, the varieties under study are referred to as 'pidgins' both because they display processes of pidginisation and because native Arabic speakers recognise them as having a norm different from that of any other variety of Arabic. The migrants' native languages are referred to as 'substrate' languages.

### 2.4. Geography and Demographics of the Region

The Arab Gulf States are located in the centre of the Old World[4] (see Map 2.1). The geographical location of the Arabian Gulf is considered an ideal transit hub for trade ships carrying goods between Asia, Africa, and Europe. As a result, most of the indigenous people already live near the coast or have moved to areas where they are in frequent contact with sailors from various nationalities and linguistic backgrounds, which might possibly explain the large number of loan words in GA from languages spoken in nearby countries, such as Persian, Turkish, and Urdu (Almoaily, 2012).

[4] Some geographers use the term 'Old World' to refer to Asia, Africa, and Europe (see Mundy, Butchart, Ledger, and Piper, 1992).


Map 2.1: Arabian Gulf States (Source: Google Maps[5]).

The demographics of the region have undergone radical changes and developments since the middle of the twentieth century (Feghali, 2004). These changes may have played a significant role in shaping the linguistic scene in the Gulf, especially the arrival of Asian workers from Pakistan, Bangladesh, and India, among other places, who began to arrive in the Gulf States in large numbers. For example, Foulkes and Docherty (2007) demonstrated several social parameters that could also greatly influence linguistic change, namely, geographical area, social class and social network, age, sex and gender, and race and ethnicity, which, for instance, are frequently used in phonological variation studies. In the region of SA, for example, there are 9.7 million foreigners compared to 20.3 million locals, according to the Central Department of Statistics and Information (CDSI) 2013[6]. The geographical proximity of South Asian countries to the Gulf States makes easier to bring Asian workers to the Arabian Gulf region, as

[5] Retrieved on 09 January 2016 from http://maps.google.com/

[6] Kingdom's Central Department of Statistics and Information. Retrieved on 01 August 2016 from http://www.data.gov.sa/dataset/saudi-non-saudi-population/resource/668ffa2a-f536-4e76-9014-05287ddc0df5

this region has had closer historical links with some parts of Asia than with many, more geographically distant parts of the Arab world (Kapiszewski, 2006). Apart from geographical proximity, there are several reasons that explain the high number of Asian immigrants in the Arabian Gulf. First, Asian FWs are less expensive to employ; are easier to lay off; and are believed to be more efficient, manageable, and obedient (Girgis, 2002: 29). Second, Asian governments facilitate their workers' smooth flow to the Gulf and have become more involved in their recruitment and placement, and the workers are able to fully satisfy the needs of the Gulf employers. Third, immigrants are willing to take low-prestige jobs with low income, especially jobs which citizens in Gulf countries do not want (e.g. working as shopkeepers, barbers, tailors, laundry workers, bakers, etc.).

Due to the steady presence of a large number of immigrant workers from various linguistic backgrounds, the situation in the Gulf area has been ideal for the crystallisation of pidginised Arabic language in the region. Demography and economy have played roles in the linguistic development of GA and the emergence of GPA. The next section discusses this issue in more detail.

### 2.5. Economic Factors and Language Contact

As discussed above, the recent significant increase in jobs in the Gulf, after the discovery of oil in the region, and the lack of opportunities to make money in migrants' home countries have influenced GA and have led to the emergence of GPA. Abdeljawad and Abdeljawad (2013: 10) state that 'within the same community, different groups of people go through various, economic, political and cultural processes resulting in competing patterns which may lead to inconsistencies in the application of variation'. According to Feghali (2004), GA varieties in SA have been influenced by the new prosperity. In fact, large groups of Saudi citizens who have moved to Riyadh and to the Eastern Province have promoted cultural diversity and interaction between the dialects of the region, the dialects of Arab workers in SA, and the

languages of non-Arabs (Almoaily, 2012). Over time, this frequent interaction has had an influence on GA, leading to the levelling of the GA varieties at various linguistic levels: the phonology and the morphosyntax, which in turn also lead to the emergence of the unique linguistic features of GPA.

The demand for FWs in Gulf countries at the beginning of the 2000s – after the high oil prices – allowed for a further rapid development of several Gulf states and consequently a large growth in population – the expatriate one in particular. The reality has greatly exceeded earlier predictions. Due to such a vast number of expatriates, the Saudi government sets immigration regulations and policies for FWs in SA, as stated on the website of the Passports Agency[7] of SA. In this regard, someone planning to come to this country must understand the *kafeel* system (a sponsorship system used to monitor migrant laborers, working mostly in the construction and domestic sectors, in Lebanon, Bahrain, Iraq, Jordan, Kuwait, Oman, Qatar, SA, and the UAE). All workers need permission from their sponsor to enter SA and leave the country on a permanent or temporary basis, among many other things, and every expatriate must have an *Iqama* (residence permit) for 2 years.

Some immigrant workers and employers violate these immigration policies, for example by entering the country illegally or with a visitor permit and finding work, while others run their own business and pay a monthly or annual sum to their 'fake employers'. Such violations may have contributed to the emergence of GPA, since pidgins tend to arise in situations where there is a lack of interpersonal integration (i.e. extensive social contact) between the two groups in contact (i.e. locals and immigrants), as described by Bakker (2008) and Almoaily (2012).

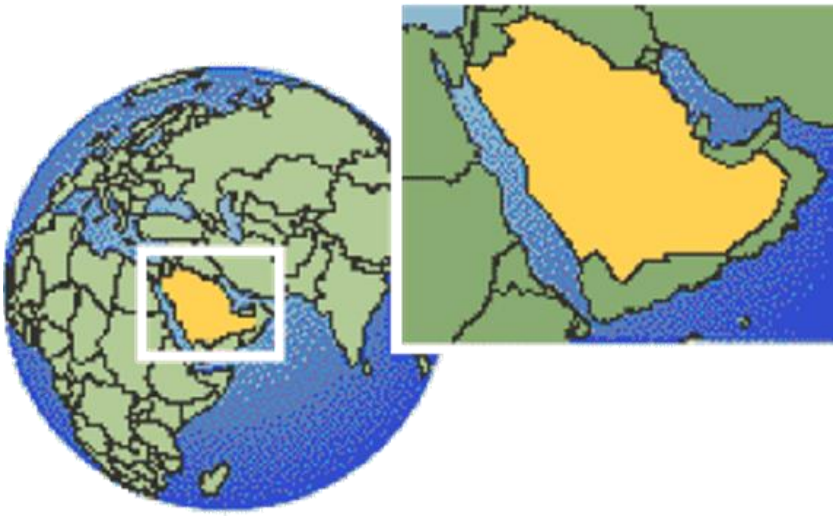
Moreover, another factor may play a major role in the emergence of new pidgins: linguistic simplification. Due to the linguistic complexity of the GA phonology as well as its

[7] Retrieved on 10 August 2016 from http://www.gdp.gov.sa

morphosyntax, most GPA speakers replace most of the typologically less common phonemes in GA, such as pharyngeal phonemes /ʕ/ and/ ħ/ and the fricative voiceless velar phoneme /χ/, with more typologically common phonemes. Indeed, the GPA phonetic inventory does not include these sounds, which have been replaced with /ɑ/, /h/, and /k/, respectively (for a full inventory of GPA and GA phonemes, refer to Albaqawi, 2016; Almoaily, 2008; Naess, 2008; and Smart, 1990). Apart from the above-mentioned factors, it is worth noting that the strategies of language contact were constantly followed by the FWs, which could create a sustainable language that has a new, simplified, and reduced system. For example, the term *Gastarbeiter-Deutsch* (guest-worker-German) is applied to the simplified German of foreign or migrant workers, particularly those who have moved to West Germany and mainly originate from Mediterranean countries of Europe and countries in North Africa.

### 2.6. Overview of the Kingdom of Saudi Arabia

The KSA is the largest country among the six Arabian Gulf nations and covers approximately four fifths of the area of the Arabian Peninsula. It is located in the southwest corner of Asia (see Map 2.2) and is thus at the crossroads of three continents: Europe, Asia, and Africa. The Kingdom is surrounded by the Arabian Gulf in the east and the Red Sea in the west, with an area of approximately 2 million $km^2$ (Siddiqui, 1998; see Figure 2.1). SA is administratively divided into 13 provinces, and each area is characterised by its history, culture, geographical features, and natural resources (Zuhur, 2011: 13). Additionally, the Kingdom is home to the Two Holy Mosques in Makkah and Al-Madinah, considered the most holy spots on Earth, and the direction of the Kaaba (Qibla), to which more than 1 billion Muslims turn during prayer (Murtada, 1996: 32).



Map 2.2: Saudi Arabia.

The current population of SA is 33,993,878, according to the latest United Nations estimates[8]. Arabic is the official language of the country, and Islam is the official religion. Long (2005: 1) considers the KSA to be a 'young nation with an ancient history', noting that its previous communities 'have lived and worked in a symbiotic relationship with nomadic tribes for at least 6,000 years'. However, the average number of nomads in SA, who are commonly known as Bedouins, 'diminished substantially' in the early part of the twentieth century, according to Mufti (2000: 1). Due to the discovery of oil, many Bedouins migrated to developing city centres, such as Riyadh and Jeddah, seeking job opportunities.

Metz (1993) characterises the social life in SA as a religious and conservative one. The SA society thus suffers from a lack of cultural diversity, and its cultural values are essentially a merge of tribal and Islamic principles (Al Lily, 2011). This religious and conservative status has been at its roots since the eighteenth century, as argued by Montagu (2015: 5), who also indicates that around 60% to 70% of Saudi society is comprised of conservatives who refuse to change and adopt reforms that the government has introduced. However, SA expects radical

[8] Retrieved on 03 April 2019 from http://www.worldometers.info/world-population/saudi-arabia-population/

transformation in its economy, and this provides a unique opportunity to improve the human rights of women and the poor since launching the new approach that was reflected in *Vision 2030*. Philip Alston (2019)[9], the special rapporteur on extreme poverty and human rights, reported the following at the end of his official visit to SA:

> Despite the plethora of serious human rights issues in Saudi Arabia, the radically new approach reflected in Vision 2030, the National Transformation Program 2020 and the Fiscal Balance Program recognizes the need to encourage full female participation in the labour market, which will drive the cultural changes needed to enable women to become both more economically productive and more independent.

Moreover, Alston adds:

> Vision 2030 recognizes that Saudi women represent "a great asset" which is currently under-utilized, and the need to recognize women's rights points in the same direction. The 2012 decision allowing women to work in the retail sector transformed the lives of millions of women who were finally able to work. So too should the current economic transformation lift existing restrictions on women's economic and other independence … The driving ban should be lifted, and women should no longer need authorization from male guardians to work or travel.

It is interesting to study the situation of SA after applying Vision 2030, especially in terms of the number of improvements that Saudi women have led in recent years. This situation will also lead to a change in female migrant domestic workers' circumstances. It might, for example, afford them opportunities to work in many different sector jobs where gender segregation is no longer a barrier. Gender segregation is one of the strictest norms that exist in Saudi society and has obstructed the lives of Saudi women and female migrants alike.

### 2.7. The Setting: Riyadh City

The city of Riyadh (plural of *rawdhah*, oasis) was founded around 1740 (Al-Hathloul, 2017), and in 1902, Riyadh was made the capital of modern SA by its independent governor, Abdulaziz Al-Saud (Facey, 1992). The city is located in the Central Province of SA (see Map 2.3), and its history stretches far back to an ancient heritage, dated more than 15,000 years ago.

---

[9] Saudi Vision 2030 could be a Catalyst for Realizing Women's Rights. Retrieved on 13 June 2017 from https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=21099&LangID=E

According to the High Commission for the Development of Al-Riyadh, the golden age of the city began on 15 January 1902, when the city of Riyadh was restored by King Abdulaziz and became the seat of the Government of the KSA and capital of the country. It was therefore the first to attract the attention of authorities. Great changes ensued in Riyadh following the oil boom of the early 1970s, and since then, the size of the city has increased more than 100 times to accommodate a burgeoning population.

Today, Riyadh is both one of the fastest growing cities in the world and one of the largest Arab capitals; it is also the most important from the government's point of view, boasting a population of close to 7 million, including a large expatriate population (64% Saudi citizens vs. 36% expatriates from around the world[10]). Riyadh is the main hub of all national and supreme activities of the country, incubating various government, regional, Islamic, and international institutions and bodies, and it is the first destination for business. As the city is located in the middle of the KSA, it enjoys a unique, strategic position among the continents of Asia, Africa, and Europe. Riyadh Province belongs to the historical regions of *Najd* and *Al-Yamama*. Moreover, Riyadh is famous for its gardens and natural surrounding oases amid the desert, mountains, and plains of *Tuwaiq*. It attracts visitors who wish to learn more about the history and authentic culture of SA.

[10] Riyadh in 2013 (Hijri, 2016) http://www.arriyadh.com


Map 2.3: Riyadh, the capital city of Saudi Arabia.

## 2.8. Saudi Arabian Society and Sociolinguistic Characteristics: Social Relationships with Asian Migrant Workers in the Context of Acquiring the Local Language

Saudi Arabian society has been witnessing massive development and change in all sectors as a result of the 1973 oil boom. This major transformation is reflected in many ways, such as in economic conditions; improvement of the educational, health, and transportation systems; lifestyle changes; the extent of women's work outside the home; and changes in gender roles in the Saudi context (Hamdan, 1990). This thesis examines the Saudi society in terms of the traditional attitudes towards women working and their position in Saudi society; discusses the impact of gender roles and gender segregation; and highlights the contributions of religion, sociocultural norms, and traditions in that society.

As regards Islamic values, Saudi Arabian society is highly influenced by religion socially, culturally, and politically (Salamah, 2018). All areas of law in the kingdom are derived from the religious precepts of Islam, the *Holy Quran*, and the *Hadiths* (noble teachings from the Prophet Mohamed, peace be upon him), and the society is thus conservative in all aspects of life. Jawad (1998: 29) describes the Saudi Arabian society as ‘traditional and ostensibly closer

to Islamic cultural norms than the other Muslim countries', and the two Holy regions (Makkah and Al-Madinah) are located in the country, making it the Centre of Islam.

The local society's traditions and attitudes have been shaped by Islamic values and Bedouin culture. The family bonds in Saudi society are strong, to the extent that they permeate all aspects of life, even the business world. With the economic growth that began in the region due to the discovery of oil, a considerable change in social life occurred in the late 1970s, especially with the significant inflow of many migrants[11] to the region with different cultural, linguistic, and religious backgrounds. Lee (2006) refers to the role of a society's culture in shaping the identity of its migrants. Paiva (2011: 66) adds that the identity construction and language learning processes overlap in human identity. She also states that 'Learning a language involves coping with fractal dimensions of the identity complex system. Besides being a learner, one has other identities, such as gender and social class identities, and additional ones can arise along the SLA process'.

The situation in SA, with the arrival of high numbers of worker migrants, is that the country's government and the other Gulf states impose strict immigration laws and policies on migrant workers, which might lead to a restrictive learning environment. One of these restrictions, particularly in SA, is that migrant labourers must work on short-term contracts – the duration of a work permit is only 2 years. However, this permit can be renewed an unlimited number of times on condition that the duration does not exceed 2 years each time the work permit is renewed. Furthermore, when coming to SA, those workers leave their families back home, and their main purpose in the country is to financially support their families (Niblock, 2006). Above all, for workers to obtain a work visa, the Saudi government imposes a sponsorship system.

[11] According to the International Migration Report 2015, Saudi Arabia has the third largest number of migrants worldwide (around 10 million).
https://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationReportP2015_Highlights.pdf

This sponsorship system authorises employers to decide whether FWs should stay or leave the country and even to transfer workers' sponsorship to another employer. In fact, this sponsorship system exists not only in the Gulf region, but also in East Asia and South Asia (Raymond and Abdullah, 1999; Piper, 2004; Asis and Piper, 2008). This system has been described by Abed (2017) – as mentioned by Varia (2008) and Fox, Mourtada-Sabbah, and Al-Mutawa (2006) – as one that 'creates a notable difference in social rights between the two groups, whilst also enshrining the economic distance between migrants and nationals' (p. 78). Moreover, these restrictions are also imposed in the civil marriage laws that exist in several Gulf states, such as Saudi, Oman, and Qatar, preventing citizens from marrying outside their Gulf states (Dresch, 2006).

In fact, such an environment creates a dominant native community that divides society into two groups: natives and non-natives. Bakir (2010) highlights that the social order in the Gulf region in general and in Saudi Arabian society specifically, is characterised by a social gap where two groups of communities exist. The first group, the dominant community of the locals, has always kept a distance from the second group, migrant workers; the former does not allow the latter to be a part of their community and culture. The author adds that this social gap has created a restricted and unwelcoming social environment for the migrant labourers' community and restricts them from acquiring the language of the locals. Researchers such as Holm (2000) and Sebba (1997) believe that a social gap between speakers of superstrate and substrate languages is the key factor in the emergence of pidgins. In addition, Abed (2017: 183) concludes that 'there is an apparent social gap between the unskilled/semi-skilled migrant labourers' community and the local population' and that 'migrant labourers in SA live in restrictive social conditions'. He also described the migrant labourers' community in SA as a 'socially marginalised community' whose communications and lives 'revolve mainly around labour activities', with 'very limited access to the dominant local culture'. On the one hand, migrants

display ‘a general tendency to not be involved in social interaction or social activities with the local people outside of labour’. On the other hand, the author adds, ‘the local people do not necessarily hold a direct negative attitude toward the unskilled/semi-skilled migrant labourers, but they did show a preference to not be living next door or amongst the unskilled/semi-skilled migrant labour community’. Finally, he suggests that ‘the social gap between the two groups contributed largely to the emergence of GPA’ (p. 184). His conclusion is supported by a number of researchers in the field of P/Cs, such as Holm (2000), Sebba (1997), Siegel (2008), Todd (2003), Gramley (2012), and Velupillai (2015).

For the most part, in the Saudi social context, GPA, for example, is mostly associated with the migrant community, and it was their target language (TL) once they arrived in the country. However, one of the questions that the current study seeks to answer is whether GPA will always be the TL in the eyes of Asian migrant workers, or whether they will improve their competency in the TL. Table 2.1 and Figure 2.2 provide data for 2000 and 2015 on the estimated number of international migrants, the percentage of migrants in the total population, the percentage of female migrants, and the median age of migrants for the Gulf countries.

| **Country or area** | **Number of international migrants (thousands)** | | **International migrants as a percentage of total population** | | **Females among international migrants (percentage)** | | **Median age of international migrants (years)** | |
|---|---|---|---|---|---|---|---|---|
| | 2000 | 2015 | 2000 | 2015 | 2000 | 2015 | 2000 | 2015 |
| **Bahrain** | 239.4 | 704.1 | 36 | 51 | 31 | 28 | 33 | 36 |
| **Kuwait** | 1127.6 | 2866.1 | 58 | 74 | 32 | 34 | 28 | 34 |
| **Oman** | 623.6 | 1845.0 | 28 | 41 | 22 | 19 | 34 | 30 |
| **Qatar** | 359.7 | 1687.6 | 61 | 75 | 24 | 16 | 33 | 32 |
| **Saudi Arabia** | 5263.4 | 10185.9 | 25 | 32 | 33 | 32 | 32 | 33 |
| **United Arab Emirates** | 2446.7 | 8095.1 | 80 | 88 | 28 | 25 | 31 | 34 |

Table 2.1: International migrant stock by age and sex in the Gulf countries (adopted from International Migration Report 2015).

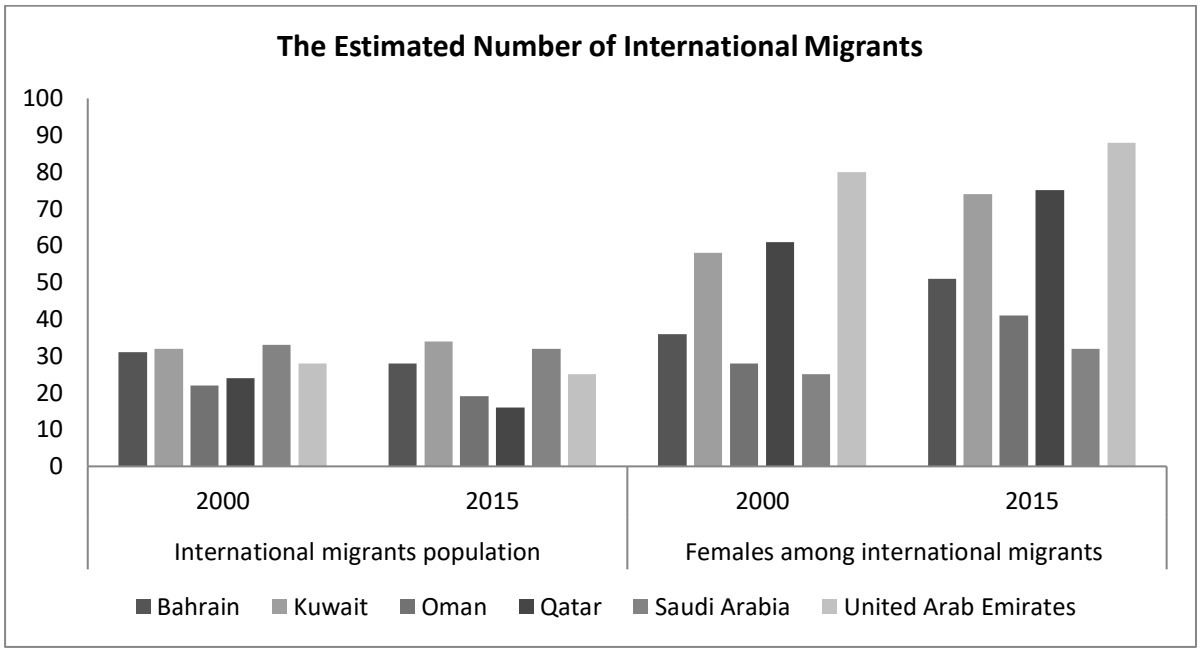



Figure 2.2: The percentage of migrants in the total population and the percentage of female migrants for the Gulf countries.

## 2.9. The Role of Gender in Saudi Arabian Society and its Relationship with Asian Migrant Workers in the Context of Acquiring the Local Language

Examining the position of Asian migrant workers in Saudi society (see Section 2.8) indicates the extent to which the social and labour environments are curtailed and restricted in the context of acquiring the local language. This section examines the role of gender, as gender segregation is part and parcel of the Saudi culture and society (Salamah, 2018). Some countries in the Middle Eastern region continue to implement gender segregation in their institutes, such as schools and other public places (Doumato and Possusney, 2003: 11), while SA imposes strict gender segregation, even on public transport and in public facilities such as restaurants, coffee shops, beaches, and amusement parks (Tamimi, 2012). In fact, having such gender-segregated lives will lead to an unequal distribution of women and men across occupations and educational establishments (Salamah, 2018). As a result of this strict gender segregation, Saudi women have limited access to the public sphere, which in turn limits their social contribution. In SA, the position of men and women differs from that of other Islamic nations, and the consequences

of living in such a conservative society will hence apply to all men and women, including foreigners.

Despite much research having been conducted on the individual differences, such as effect of age, motivation, attitude, aptitude, personality, learning styles, intelligence, personal beliefs, and identity on L2 learning, Rivas (2017) states that gender differences are rarely mentioned within those individual differences. He also points out that '[g]ender as a variable affecting language learning and acquisition is rarely reflected in bibliographic references or taken into consideration in wider or more particular analyses of how languages are learnt' (p. 2). Indeed, neither a singular distinct field of research relating specifically to gender and SLA nor a theory of gender and SLA exists as yet (Feery, 2008). Moreover, a limited amount of research, such as the results in Schmitt (2013), simply mention that females seem to beat males in language learning, without offering further detail. Therefore, additional research on the role of gender in SLA must be conducted.

Since adults learn GPA as an imperfect L2 through interaction with other speakers (locals and their co-workers), '[s]ociocultural theories are concerned with the construction of meaning through negotiation, which links directly with the focus of the latest research in language and gender' (Rivas, 2017). The concept of gendered cultural norms, which focuses on the idea of separation between the male and female genders, could play a major part in the process of acquiring a language. Alqasem (2017) states that this restriction has a great impact on acquiring the GPA language when GPA speakers come into contact with a particular community such as the Saudi one.

Before I examine the gender role in SLA in the contexts of Asian FWs in SA, it is worth presenting a brief overview of the field of language and gender studies (in both Western and Arab societies) and highlight important studies that have dealt with women and gender in Arab

societies. By studying the connection between gender differences in language and the social sources thereof, possible reasons for gender differences in language can be determined.

### 2.10. Language and Gender

Gender was traditionally observed as one sociolinguistic variable, similarly to age, social class, social status, and ethnicity. It was not until the mid-1970s, when Robin Lakoff's *Language and Woman's Place* (1973) was released, that research into language and gender began to be established and continued to become a dynamic and ever-growing area of research today. Traditionally speaking, research into the relationship between language and gender has mainly centred on language *usage* (mostly L1) in any possible relationships, intersections, and tensions between language and gender, as opposed to language *learning* (which for the most part denotes L2) and gender. Feery (2008) examined empirical studies on language and gender and found that they seem to be classified into two categories: variationist studies, which focus on actual linguistic gender patterns, and associated factors and interactional studies, which concentrate on language use in interaction.

The 1970s marked the beginning of the flourishing of research into feminist linguistics, and a feminist critique of language usage gradually began to emerge and opened the door to a major discussion about the relationship between gender and language. Examples of feminist literature are Roszak and Roszak's *Masculine/Feminine: Readings in Sexual Mythology and the Liberation of Women* (1969) and Robin Morgan's edited collection, titled *Sisterhood is Powerful: An Anthology of Writings from the Women's Liberation Movement* (1970). Other work has particularly explored the relationship between language and gender, along with the social construction of woman, which has become a subject of recent discussion in the field of language and gender research, such as Nochlin, Gornick, and Moran's *Women in Sexist Society* (1971) and Firestone's landmark text, *The Dialectic of Sex: The Case for Feminist Revolution* (1970).

As a consequence of the emergence of a feminist critique of language, researchers began to develop and expand their knowledge, seeking more evidence of differences in male and female language use. On the one hand, Lakoff (1973) established the earliest modern theory about 'women's language', which is commonly referred to as the *deficit theory.* It described women's language as deficient compared with men's, essentially labelling women's manner of speaking as a reflection of their insecurity and powerlessness in society (Freed, 2003). On the other hand, the *dominance theory* of language and gender, offered first by Thorne and Henley (1975) – see also Fishman (1983) – is based on the assumption that men have power and dominance. Aslan (2009) suggested that 'men gain and maintain power over women in social interaction by means of interrupting and overlapping women's speech, using a high volume of words, or denigrating women'.

Later, during the 1990s, the *difference theory* was modelled by Maltz and Borker (1982) and Tannen (1990, 1994). It gained substantial ground and moved away from concepts of deficiency and hierarchy towards a new concept that 'women and men used specific and distinct verbal strategies and communicative styles which were developed in same-sex childhood peer groups' (Freed, 2003: 701) but in different subculture environments. In other words, the difference theory proposes that the difference in language usage starts in childhood, where parents use more words about feelings in relation to girls and more verbs in relation to boys, arguing that males and females consequently belong to different subcultures and therefore speak differently. However, critics in academic circles widely criticise works by the proponents of this view, for example the books of Deborah Tannen, as 'they were seen to stereotype and dichotomise women and men even further' (Feery, 2008).

As a result of such criticisms of this kind of approach, a new model began to appear: a speech community-based approach to language and gender that would examine particular groups of men and women. This approach is the community of practice model, which proposes gender

as ‘a transversal category that cannot be analysed in isolation but needs to be connected with identity, ethnicity, social class and nationality. Gender is not the same as biological sex but a contextualised social construct’ (Rivas, 2017: 3). Even though this model was primarily considered a theory of learning, Eckert and McConnell-Ginet (1992) applied it within a language and gender context. Furthermore, a number of theories and assumptions have been drawn from research in language and gender that focused on L1 use and performance:

> It has been found that men prefer to use directives, sentence initial conjunctions and interruptions. In fact, men tend to interrupt other interlocutors’ discourse much more frequently than women, who show a tendency to use more questions, justifiers, intensive adverbs and personal pronouns. Their speech is more abundant in fillers, question tags, empty adjectives, precise colour terms, intensifiers, hyper correct grammar, very polite forms, avoidance of swear words and emphatic stress. These findings are probably changing in the new generations in western civilization with the access of women to more equal social atmospheres and was stronger in the past or in other cultures where women’s equality is still far from being dreamt of (Rivas, 2017: 3)

In terms of performance in SLA, Rivas (2017: 4) mentioned that women are said to learn an L2 faster than men:

> Women have been found to rely more on metacognitive strategies, whereas men show a greater tendency to use translation strategies. In discourse analysis, men tend to interrupt more, dominate and control conversations and they are better at reasoning and analysis, though women are better at memorising, which might explain why girls learn their first language earlier and better than boys. Women are said to learn a second language faster, as well. They tend to perform better in the four skills, their vocabulary is richer and their pronunciation is clearer (Jiménez Catalán: 2003). Males seem to prefer visual and tactile learning strategies and females prefer auditory strategies, formal explicit input and rehearsal.

Rivas (2017: 4) not only explains the reasons behind these differences, which are basically sociocultural, but also connects them to individual differences, such as motivation. He says that women seem to be more motivated and enthusiastic and have a positive attitude towards the language and its culture, both in L1 and L2, whereas

> Males’ motivation tends to be instrumental, a means to obtain material benefits such as better job opportunities or results in exams. Females tend to have a more integrative motivation, which has been claimed to produce better language learning and even native-like proficiency in L2. Women also show more confidence during the acquisition process and have more self-esteem, not rejecting corrections, but more often learning

> from them. They also use more strategies and more often than men. Their aptitude generates more confidence in them and creates positive attitudes to language. They tend to be more interested in social activities, which favours exposure and practice opportunities, and to join in co-operative tasks more often than men, who tend to go for the competitive ones, concentrating more on the result compared to others than on the content of the task.

### 2.10.1. Language and Gender in Arabic-Speaking Societies

Research on gender and language in Arabic-speaking societies 'is still at its beginnings in spite of the fact that Arabic sociolinguistics has attracted the attention of scholars worldwide' (Sadiqi, 2007: 642). Starting in the 1980s, research into gender-based linguistic variation flourished in the Arab world, especially with the most popular topics in the last few years, such as the use of the Islamic veil[12] (Vicente, 2011). Whilst the level of research carried out in this area lagged in comparison to Western countries, some documentation on Arabic dialectology had previously been delivered, such as the work of Bauer (1926) on women in Palestine and the work of Roux (1952) on women in Meknes in Morocco (Vicente, 2011). Later, a number of other studies were conducted: Abdel Jawad (1981), Royal (1985), Haeri (1991), Al-Muhannadi (1991), Al Wer (1991), Dekkak (1979), and Jabeur (1987). Since then, the main legacy of gender and language studies was established with the development of the sociolinguistic and variationist studies of Arabic varieties, although there are still many gaps in comparison with Indo-European languages due to stereotypes which blur facts and tend to generalise. This topic area – studying the relationship between the social category of gender and linguistic change in Arabic-speaking societies – was improved by a number of studies, such as the work of Boucherit (1989), Rosenhouse (1998), Sadiqi (2003), and Bassiouney (2009).

[12] Ahmed (1992), Badran (1995), Khoury and Moghadam (1995), and Heath (2008) provide examples of this topic.

Arabic is considered a gendered[13] language – 'masculine versus feminine' (*muḏakkar* versus *muʔannaṯ*) – with a grammatical gender system that divides all nouns into gendered categories, which requires gender agreement with associated items in the sentence, such as determiners, adjectives, and demonstrative pronouns (Alkohlani, 2016). This is an interesting perspective describing the relationship between gender and language; however, my perspective is quite different. In this study, I consider gender as a sociolinguistic category, not a grammatical one, and this is used as the criterion to analyse the use of language by men and women GPA speakers based on their length of stay in the Gulf. Although Vicente (2011) refers to some similarities between both approaches – considering gender as a sociolinguistic category and considering gender as a grammatical category – male speech in sociolinguistic studies was taken as a reference, 'unmarked' variety. This appears in the words of Joseph Chetrit (cf. 1986: 2) when he talked about the female sociolect of Judeo-Arab women in Morocco 'la parole des hommes sera donc considérée ici arbitrairement, comme non-marquée par rapport à celle de femmes, dans les mêmes contextes d'énonciation dont nous aurons à traiter dans notre analyse' ['the word of men will therefore be regarded here arbitrarily, as unmarked in relation to that of women, in the same contexts of enunciation which we will have to deal with in our analysis']. Observations about language variation by variationist linguistics revealed that in some contexts, gender is a basic element in linguistic change; however, no quantitative variationist analysis of gender variability in contact languages has been provided yet. Thus, a statistical analysis of the relationship between the gender variable (male and female GPA speakers) and other independent variables within a Saudi community (such as length of stay of the speaker in such a community) is perhaps one of the most original contributions to studies on the relationship between gender and language in Arabic-based pidgin languages.

[13] Discussed in more detail in Chapter 5, Section 5.3

Linguists such as Labov (1966), Lakoff (1973), Tannen (1990), Holmes (1990), and Wardhaugh (2010) studied the differences in speech between men and women, and all cooperated in answering one question: Why do women tend to speak differently from men? Labov and Lakoff's studies in the 1970s were of great influence in the early stages of both Arabic sociolinguistics and Western societies. One of Labov's most prominent theories regarding the impact of gender on linguistic change in Western societies is that women rather than men tend to speak like higher classes, and in fact they tend to do so to raise their standards, as this is women's nature[14] (tendency to correct vernacular speech patterns). Labov (2001), in Principle II of his work, *Principles of Linguistic Change*, observed a sociolinguistic phenomenon – called the 'gender paradox' – that 'women conform more closely than men to sociolinguistic norms that are overtly prescribed, but conform less than men when they are not' (p. 281). In other words, women are more likely to use prestige forms and avoid stigmatised variants than men for a majority of linguistic variables, but they are also more likely to lead language change by using innovative forms of variables. His statement was criticised by Haeri (1996), who believes that several social factors might condition linguistic behaviour more than depending only on women accessing the classical-standard rule of the language, such as a speaker's type of occupation speakers. Apart from Haeri (1987, 2000), other authors, such as Ibrahim (1986), Walters (1991), Hachimi (2001), and Bassiouney (2009), found that applying Western theories (non-diglossic societies) to an Arabic-speaking context (diglossic societies) is challenging, since linguistic variation in such Arabic diglossic societies depends more on interdialectal levelling than on the superstratum influence of Classical Arabic (CA) or Modern Standard Arabic (MSA). In this sense, if GA is regarded as a prestige norm operating in Arabic-speaking societies in the Gulf region and having the same role as the variety considered to be

[14] Women's tendency to have the desire to rise to a higher standard was proven by Labov (1966) when 'he carried out a sociolinguistic analysis in New York City focused on the rule of r-dropping…and its use by upper, middle, and lower-class speakers.' ("Dialects", p.431).

standard in Western communities, 'then we can find some similarities between Labov's conclusions and those drawn by Arabic sociolinguistics, in particular regarding the question of women being more sensitive to the prestigious forms of the language than men' (Vicente, 2011: 14).

Vicente (2011) stated that there are two clearly different scenarios regarding Arabic-speaking women and their role in linguistic change in a diglossic society. One is that under certain circumstances, women adapt faster than men to the process of linguistic change, and women are sometimes the ones who start that change and use prestigious forms more frequently than men. The other is that women in rural contexts tend to be considered conservative in their linguistic practices and 'preserve more local productions and keep away from prestigious and generalized forms at a national level, which are then linked to the male world' (p. 14). He also added that both scenarios are entangled with several social factors, such as education, age, and rural/urban context, which, as many researchers believe, play a major role in the process of linguistic change, more than that in Western societies. Vicente (2011: 1) claims that

> It has been established that young, educated and urban women take a more active part in linguistic change and collaborate in the spread of new variations more often than men of any age, while older and illiterate or semi-illiterate women from rural areas preserve the most ancient features which in many cases are actually in danger of becoming extinct and are usually avoided by the rest of the language community because they are stigmatized. It is, therefore, the sociolect of a socially segregated group, with lower access to interdialectal levelling and, due to this, less permeable to innovations (see, for instance, Abu Haidar 1988 and Walters 1991, and Vicente 2002). In these cases, femininity has been identified with linguistic conservatism and also with illiteracy or semi-illiteracy…

Older women, who are usually illiterate or semi-illiterate and stay within private spheres, very much preserve the old dialect and are highly conservative from a linguistic point of view, even in a situation of interdialectal communication. However, amongst the younger population, there are fewer differences between men and women (Abu-Haidar, 1988; Walters, 1991; Vicente, 2002). In this case, the situation is due to a variation based on both age and gender. Although

this study intends not to generalise, it is important to discuss later the role of old people, conservative people, and Bedouin women in the process of GPA language acquisition.

### 2.10.2. Gender and Second-Language Acquisition) in the context of Migrant Gulf Pidgin Arabic Speakers

Some researchers believe that P/C languages could have emerged as a result of imperfect L2 learning, since adults learn pidgins as L2s (see DeGraff, 1999; Mufwene, 1990; Becker and Veenstra, 2003; Winford, 2003; Field, 2004). In this case, I deal with GPA as an L2 learnt by adults (Section 3.1.1.3 discusses the *imperfect L2 learning hypothesis* in more detail).

This section focuses on the differences between females and males in their imperfect L2 learning process based on the concept of the social location of the learner and the learner's social identity. Since most research on language and gender reveals differences between females and males in the use and performance of their L1, it is logical to expect differences in their L2 learning processes as well. However, research related to SLA and gender often includes studies pertaining to similar variables, such as age, race, ethnicity, personality, motivation, aptitude, and individual differences, though research on this specific topic of SLA and gender seems to be scarce. For instance, Larsen-Freeman and Long (1991: 204) affirm that they 'know of no study that has systematically investigated the rate of SLA in females versus males', but they refer to some studies[15] that 'have reported sex-related differences incidental to their main focus' (Larsen-Freeman and Long, 1991). Throughout the previous 10 to 15 years, SLA research relating specifically to the gender variable has gradually emerged, yet theories of SLA seldom tackle the potential role and influence of other variables, such as gender, within SLA (Feery, 2008). The main reference of most research on gender and SLA has exclusively focused on specific learner characteristics (adult, middle-class, Western women) in a classroom

[15] Farhady (1982) and Eisenstein's (1982) studies have been mentioned by Larsen-Freeman and Long (1991).

setting and has neglected learners' other situations outside of the classroom, such as those of ethnic and indigenous minorities, refugees, or migrant communities, where the interaction can take place within a more powerful society in which language is needed in order to survive. The main stream of this research is divided into two main areas: 1) studies focusing on the differences between males and females based on the performance and use of L2 learning strategies, from the perspective of language learning as a female activity, and 2) studies focusing more on the social location of the L2 learner and the social identity using the community of practice method. The second main stream of research takes precedence throughout this discussion.

Researchers seek to understand why women tend to outdo men in SLA. A claim that females exceed males in verbal abilities, accuracy, and memory due to their biological predisposition is one of the misconceptions that has unanimously been rejected by neurology researchers denoting that there is no biological basis for such an assumption (Rivas, 2017). In addition, Sunderland (2000) tackled different factors in her investigation on gender issues in language learning and concluded that gender differences or gender gaps in language learning are basically the result of social factors: 'socially constructed gender differences'.

Other claims, as evident in the studies outlined by Labov and Lakoff in the previous section, are that women are frequently more sensitive to acquiring the prestigious forms of the TL than men, and they therefore avoid the stigmatised variants and pursue the standard rule of the language as a way to make up for their social insecurity. Furthermore, the reasons women tend to acquire standard or prestigious forms in language more than men are likely due to their role in society, which is different from that of men. Men's role in society offers them ample opportunities and exposure to language, thereby enabling them to acquire a better knowledge of language. Indeed, men and women have different exposure, access, or attitudes to the L2, particularly in segregation settings such as in SA where both men and women are completely

socially isolated. Moreover, men – both natives and foreigners – are doubly marginalised from women's society, which affects the SLA process. Those foreigners (GPA male speakers) tend to interact with other men in the same situation who normally have fewer opportunities to access the TL and interact with the rest of society; their deficient acquisition and stigmatised accent are hence used more than the TL as a means of communication to survive within the group. Women (female GPA speakers), in contrast, are mostly accompanied by their employers: They take children to the schools, parks, or restaurants; do the shopping with their employers; and contact and meet other domestic helpers with similar circumstances. Although their interaction is sometimes in GPA, they have more frequent opportunities to interact with other Saudi families of a higher prestige variety in society, and they thus tend to receive a less stigmatised accent and a more progressively precise utilisation of the language.

Regarding expatriates' attitudes towards the conventional culture, Rivas (2017) stated that men and women have different attitudes regarding the mainstream culture, saying that 'men do not usually take a positive attitude to a society that marginalises them, whereas women, already marginalised in their own minority community, tend to be more positive about a wider society that offers more opportunities for equality'. She added that women can easily blend into the more extensive society and dispense with the feeling of being discriminated, which might explain their aim for a non-stigmatised accent. Furthermore, Rivas (2017: 8) still emphasises the idea of women and men's attitudes towards acquiring an L2 and the extent to which their L1 has an impact on them, yet women are marginalised in their own minority community:

> [Women] do not either perceive attrition of their L1 influence on their L2 as negative, as men do, but as another means to progress in society and be respected. If they can hardly overcome racist and sexist reactions, at least they can try to minimise classist situations by masking their speech. These women, away from their cultures of origin and surrounded by a threatening more powerful one, also tend to use language in a more cooperative way, to find and give support within their community, with other women and in the wider society. This, as in L1 acquisition, contrasts with men's use of language in a much more competitive than cooperative way. Whereas female double discrimination in minorities provides less opportunities to access the target language, they still achieve higher levels of proficiency in the dominant language.

In more recent research regarding performance-related differences and differences unrelated to performance, girls attained higher grades in several language tests, but not in all cases. Kettemann and Cossée (1998) characterised this by three different features: (1) the extent to which language subjects are popular amongst boys and girls, (2) the types of learning strategies employed by boys and girls, and (3) the different attitudes that both sexes have towards learning a foreign language. With respect to learning strategies, Ludwig (1983) found no difference between boys and girls in the kinds of strategies that they used, apart from motivational reasons, where he noticed that boys had practical reasons for choosing a foreign language, whereas girls chose it because it seemed interesting. As for Bacon and Finnemann (1992), they found some differences in the types of strategies utilised.

In terms of overall attitudes towards learning an L2, researchers such as Baumert (1992), Kettemann and Cossée (1998), and Schröder (1996) indicated that women have more positive attitudes when compared to men. Additionally, Kettemann and Cossée (1998) also attempted to explain the possible causes of gender-related differences in the first place. They suggest that there are three potential approaches to explaining the differences: biological stance, the cognitive-psychological approach, and the socialisation theory-based approach (Feery, 2008).

Research on gender differences in using an L2 based on learning strategies has found that 'women use more memory, cognitive compensation and affective strategies and that better language acquisition is linked to the fact that women seem to seek more social approval and recognition than men, given their marginalised role in society' (Rivas, 2017). In terms of interaction strategies, Rivas (2017) stated that women use different strategies to interact with others in order to attain maximum input, whereas men use their interaction strategies to produce more output.

Despite researchers not being able to draw consistent conclusions on this, a solid link seems to be exist between gender and SLA and performance in favour of females. Next, Section 2.11 provides the conclusions of this chapter.

### 2.11. Conclusion to Chapter 2

The purpose of this chapter was first to introduce the Arabian Gulf region where GPA is spoken and to explain how this region has gained global significance since the discovery of oil and the region becoming home to a large number of FWs. The second purpose of the chapter was to discuss the issue of gender-associated variation in language in the Gulf States, particularly in Saudi society, reflecting social status or power differences. For this reason, literature on the influence of gender on SLA, together with an overview of the conditions and characteristics that affect language use generally and specifically in the non-Western society of SA, has been provided, since the aim of this thesis is to investigate the impact of male and female gender variation as a potential factor in language variation in GPA, for which the influence of social norms that have led to the emergence and development of GPA in the Gulf region, particularly in the KSA, have been taken into consideration.

In virtually all sociolinguistic studies that include a sample of males and females, there is evidence of a difference in the linguistic behaviour of men versus women, and the conclusion is that 'women use fewer stigmatized and non-standard variants than men from the same social class' (Vasko, 2010a). As has been seen, the phenomenon of gender speech differences is actually still observed, and linguists more or less agree on the reasons that women tend to speak differently. Given that this difference between men and women's use of language is stated in many (albeit non-quantitative) studies, I want to determine whether this difference could also be seen in GPA dialect speech in the Gulf region. Therefore, this thesis focuses on those differences between GPA female and male speakers in their imperfect L2 learning process based on the concept of the social location of the learner and the learner's social identity. The

next chapter presents a critical overview of related work in the field of contact languages and their emergence.

# Chapter 3: Literature Review

This chapter provides a general overview of contact languages, their definition, and their emergence. It outlines the main theories participating in the genesis of P/Cs, investigating the general linguistic features of P/C languages, with a main focus on European-based P/Cs. In addition, this chapter presents an account of the linguistic features based on Arabic P/Cs that have been studied, and it compares them with GPA. The last section highlights the importance of the corpus-based approach in contact languages studies, with a specific focus on the pros and cons of using corpora evidence in sociolinguistic studies.

## 3.1. A Brief Overview of Contact Languages: Definition and Emergence

Contact varieties of a language arise when two or more groups of people from different linguistic backgrounds meet and have an interest in communicating with one another for different purposes (trade, business, plantation work, etc.) but do not have access to a common language. Romaine (1988) and Bakker (2011) wrote that there is still doubt about the minimum number of languages required for the creation of a contact language. Tertiary hybridisation theory, as explained by Whinnom (1971), is a theory that involves at least three languages when P/Cs are created. However, Bakker (2011), in his analysis of Bizri (2010), suggests that pidgins or creoles can result from contact between only two languages. He states that Pidgin Madame – an Arabic-based pidgin spoken in Lebanon – emerged as a result of the contact between only two languages: Levantine Arabic as the superstrate language (higher status language) and Sinhala as the substrate (lower status language; Almoaily, 2012: 9). According to Bakker (2011), Trio-Ndyuka Pidgin and Berbice Dutch are other contact languages which have resulted from contact between a lexifier language (speakers of higher status languages,

the superstrate language) and only one substrate language (speakers of lower status languages, the substrate languages).

Contact languages have various linguistic features. These linguistic features depend on many factors, such as the length of time of the language contact, its intensity, and the languages spoken by the groups in contact (i.e. superstrate language and substrate languages), among many other factors (see DeCamp, 1971; Hymes, 1971; Holm, 1988; Romaine, 1988; Sebba, 1997; Singh, 2000; and Knapik, 2012). The terms 'pidgin' and 'creole' are two of the most relevant forms of outcomes of language contact to the current study. For many creolists, such as Todd (2003) and Sebba (1997), the difference between the two terms is that creoles are native languages, while pidgins are not. Although this is a commonly upheld distinction (see sections 3.1.2 and 3.4.3), creolists still debate the exact definitions and categorisations of P/C languages. In the next section, I discuss in detail the definitions of the types of contact languages, starting with the discussion of how contact languages develop.

### 3.1.1. Genetic Theories of Pidgin and Creole Languages

One of the most intriguing questions in the history of P/C languages is that of how they come about. The emergence of P/C languages has led to a number of theories, which are known as the *theories of genesis* (refer to Holm, 1988, 2000; McMahon, 1994; Todd, 2003; and Singh, 2000). The following sections briefly mention the most common theories in P/C studies, with a significant focus on the theories that are most relevant to the current study: substrate influence, imperfect L2 learning, and universalist theories. They are the most relevant because this thesis aims to discover language variation in GPA resulting from different morphosyntactic structures of the substrate languages on the one hand and language and gender variation from length of exposure to GA on the other.

#### *3.1.1.1. Monogenetic Versus Polygenetic Theories*

Within the literature on P/C genesis, two views have been advocated: the monogenetic and the polygenetic theories of creole languages. The monogenetic approach was first proposed by the Herskovitese[1], a husband and wife anthropological team who had worked in West Africa before moving on to study Caribbean culture and language (Singh, 2000: 46). This theory claims that a common origin of European-based P/Cs is the Portuguese pidgin. This single pidgin emerged during the fifteenth century along the route of Portuguese merchants (Lefebvre, 2004). Later, this original pidgin spread across the world due to European colonisation and became the native language of a first generation of speakers (Stewart 1967: 47). Sebba (1997: 75) referred to the process involved in the creation and development of this pidgin language as a relexification process. Some proponents of this theory are Alleyne (1971), Goodman (1964), Hancock (1968), Stewart (1962), and Whinnom (1956, 1965, and 1971).

The polygenetic theory contrasts with the monogenesis theory of creole genesis, proposing that different pidgins gave rise to different creoles. In this view, all P/Cs did not evolve from a common ancestor but rather developed separately from one another with different starting points**.** For example, French-based pidgin would be the source of French-based creoles, while an English-based pidgin would be the source of English-based creoles, and so on (Lefebvre, 2004). However, this monogenesis approach is no longer advocated in the field of P/C studies, as Lefebvre (2004: 15) points out. Furthermore, Singh (2000: 46) wrote that this theory focuses entirely on P/Cs that exist in the Caribbean region.

#### *3.1.1.2. Substrate Influence Theory*

Some similar structures exist in most contact languages, which are in fact due to the similar structures of most of the substrate European-based languages (Holm, 1988; Hall, 1966; Taylor,

[1] They concluded that 'Negroes [sic] have been using words from European languages to render literally the underlying morphological pattern of West African tongues' (Herskovits, 1936: 131; in Holm, 1989: 37).

1971, 1977) and so many others. The aforementioned scholars have suggested that Atlantic creoles have some linguistic features that could have been carried forward from their substrate languages. Arends et al. (1995) also noted some similarities across contact languages, claiming that in Atlantic creoles, the influence of the substrate can affect all linguistics levels, such as phonology, morphology/lexicon, syntax, and semantics. For instance, in Surinam creole, the syllable structure is different to the European lexifier language and similar to the substrate languages Gbi and Kikongo[2]. In the case of GPA, its substrate languages have divergent segmental phonology and morphosyntactic structures due to an influence from their L1s (this is discussed in detail throughout Chapter 5). It will be interesting to see whether these structures can be linked to language variation within GPA.

#### *3.1.1.3. Imperfect Second-Language Learning Theory*

Some researchers (e.g. Coelho, 1881; Mufwene, 1990; Arends et al., 1995) have claimed that P/Cs could have emerged as a result of imperfect L2 learning, since adults learn them as L2s. The imperfect L2 learning hypothesis was among the first hypotheses that suggested a possible role of the universal aspects of language learning in the emergence of pidgin languages, as proposed by Coelho (1881). This theory claims that the imperfect learning of the slaves' lexifier language causes the emergence of creoles. Coelho's hypothesis is a precursor to what was later known as universalist approaches (discussed in more detail in Section 3.1.1.4). In P/C languages, the first clear statement about the possible role of transfer (i.e. the transmission of elements of a speaker's native language onto the linguistic patterns of the TL) was not mentioned; however, Mufwene (1990: 11) stated that 'research on transfer in SLA and that on the substrate hypothesis in creolistics may benefit one another'. Furthermore, Klein and

[2] Gbi and Kikongo are languages belonging to the Niger–Congo language family which is spoken in northern Liberia. Retrieved from https://www.omniglot.com/writing/kongo.htm on 11 August 2017.

Perdue's study (1997) offered some supporting evidence in favour of the imperfect L2 learning hypothesis. The authors also investigated how learners can develop in acquiring foreign languages, such as English, German, Dutch, or Swedish, outside the classroom. After 30 months, the learners developed 'the basic variety' (i.e. the language necessary for communication), and this term was first coined by Klein and Perdue. In their study, approximately one third of the participants, after acquiring the basic variety, did not display any development except for the acquisition of vocabulary.

Although in the linguistics literature of imperfect L2, similarities exist between P/Cs, such as invariant verb forms and pre-verbal negation markers, Singh (2000) argues that some features of P/Cs are left unaccounted for by this hypothesis. Nonetheless, the similarities between imperfect L2 learning on the one hand and P/Cs on the other might exist by mere chance, as imperfect L2 learning and P/Cs could result from over-simplification and over-generalisation of linguistic rules. Some researchers, including DeGraff (1999), Mufwene (1990), and Becker and Veenstra (2003), prefer the imperfect L2 learning theory of P/C genesis. In addition, Siegel states that 'While more creolists today may agree about the involvement of processes of SLA in pidgin and creole genesis, there is no consensus about exactly what these processes are and how and when they apply' (2008b: 208). Therefore, more research must be conducted on the role of language acquisition in P/C genesis. In Chapter 8, I discuss the potential role of imperfect SLA in the linguistic structure of GPA.

#### *3.1.1.4. Language Universal Theories*

The main feature of these kinds of genesis theories is that they focus on the role of innate cognitive principles in the process of P/C formation rather than on the influence of the languages in contact (superstrate or substrate languages). The history of universalist theories of genesis can be traced back to the late nineteenth century (Almoaily, 2012), and one of their most interesting aspects is the existence of certain similarities in the structures of P/C

languages, which are found irrespective of their different contributing languages. For instance, Arabic-based P/Cs on the one hand and European-language-based P/Cs on the other hand are more or less based on discrete synthetic superstrate languages and on different substrate languages, even though an analytic morphology is attested in Arabic-based P/Cs as well as many Indo-European P/Cs.

Bickerton's (1981) language bioprogram hypothesis (LBH) is one of the most famous controversial theories within the domain of universalism. It was a dominating hypothesis in the 1980s and mid-1990s. According to Veenstra (2008), the LBH was an endeavour to answer the questions of how the human language began a vast number of years ago and how it has advanced since then. Bickerton attempted to generate a connection between L1 acquisition and creolisation; his hypothesis proposes that creoles are the result of an L1 acquisition process. To invent creoles, children use their parental pidgin input. This means that children who grow up in households where parents used pidgins as an input language, use their innate linguistic capacities to convert the limited pidgin input to some similar features which are found in full-fledged languages. According to the LBH, all creoles have similar structures because they were generated by utilising universal human linguistic capacities. The LBH may thus provide insights into the development of creoles when it compares creole languages and the language produced by children. The emergence of both P/Cs has been discussed in this theory. Accordingly, the LBH theory recognises the rise of pidgins as a second dialect learnt by adults and the sudden development of creoles as an L1 made up by infants using parental pidgin input. Since the LBH's emergence in the mid-1980s, it has been the subject of much debate. For instance, McMahon (1994) investigated the ability of children to create complex structures found in creoles out of less complex structures as input found in pidgins. Moreover, this theory was criticised by Seuren (1984) and Siegel (2008a), who stated that the LBH theory fails to explain how certain features have come about in creole languages, such as serial verbs and pre-

verbal TMA marking. Seuren also disputed Bickerton's (1981) description of the serial verbs and pre-verbal TMA (1984), claiming that Bickerton overstated when he described serial verbs and pre-verbal TMA as universals when actually they are not. Furthermore, Eklund (1996), Siegel (2008a), and Veenstra (2008) presented some of the strongest arguments against the LBH that come from Tok Pisin. They stated that Tok Pisin did not emerge suddenly (i.e. as a consequence of L1 acquisition over only one generation). In addition, there are also some proponents of the emergence of P/Cs as universal languages: Ferguson (1971), Todd (2003), Bickerton (1981), and Singh (2000), among others. The influence of universals falls into two types: procedural and constitutive universals. I discuss their role on the genesis of P/Cs next.

- **Procedural universals** are universals related to psycholinguistic strategies devised by speakers of the two mutually unintelligible languages in contact (see Almoaily, 2008). For instance, Kay and Sankoff (1974) and Ferguson (1971) suggested that simplification of language universals (such as replacement of inflections with free morphemes, restricted number of phonemes, and lack of allophones) must have played a role in the creation of creoles worldwide.
- **Constitutive universals** concern the similar structures of contact languages, such as subject–verb–object (SVO) word order, pre-verbal particles, and morphologically complex reflexives (see Arends et al., 1995 and Bickerton, 1984).

This study would support universalist theories of genesis if it revealed that the participants produced universal features of contact languages which cannot be traced to their L1s and furthermore that dissimilar properties of their L1s did not have a significant impact on their production of GPA.

### 3.1.2. Language Birth: Pidgin and Creole Languages

The main focus of this section is on the definition and description of so-called P/C languages. This section also highlights the lack of consensus that still exists in defining and distinguishing

between contact languages. P/Cs are generally the outcome of contact languages, as they arise from contact between two or more existing languages. They develop in areas where speakers with different native languages are in contact with one another even though they do not speak and understand one another's languages, but where a common language is needed. Such necessity for communication is complementarily distributed (e.g. trade, business, plantation work). This continued communication results in the development of a new language, one which did not exist before: A pidgin or jargon (i.e. an unestablished form of communication) is created. The development of these contact languages (i.e. P/Cs) is not exclusively, but still closely based on constant communication between the European colonialist expansion and its accompanying slave trade, and many P/Cs are thus located along former trade routes (Sebba, 1997). In fact, this might explain the existence of a substantial number of European-language-based P/Cs compared to non-Indo-European-language-based ones. This is one of the aims of this thesis, namely, to contribute to the literature of less-described non-Indo-European P/Cs. The next section discusses the earliest attempts to distinguish between the outcomes of language contact: jargon-pidgin-creole.

#### *3.1.2.1. Stages of Development*

According to the life-cycle theory 'jargon-pidgin-creole' (outlined in Figure 3.1), which was developed by Hall (1962), contact languages develop through a shift from an unestablished form of communication (i.e. jargon) to a more standard representation of language (i.e. pidgin), and finally to a nativised, stabilised contact variety (i.e. a creole). Several creolists, such as Holm (1988), Todd (2003), and Singh (2000), favour Hall's model, while others, such as Miyaoka, Sakiyama, and Krauss (2007: 123), claim that not all contact languages follow Hall's theory (1962) of 'jargon-pidgin-creole', since there are some cases where creole languages may have occurred before pidginisation (e.g. Australian koinés, Réunionnais) and other cases where jargons and pidgins do not reach the stage of creolisation (e.g. Pidgin Fijian). Furthermore,

other researchers, such as DeCamp (1991), have made some changes to Hall's model (e.g. adding a *post-creole stage*). Mühlhäusler (1986) also added conceivable situations to the language jargon-pidgin-creole life cycle, which has three possible scenarios:

Type 1: Jargon → creole (e.g. Hawaiian Creole English)
Type 2: Jargon → stabilised pidgin → creole (e.g. Torres Strait Creole English)
Type 3: Jargon → stabilised pidgin → expanded pidgin → creole (New Guinea Tok Pisin)

For a detailed discussion about Muehlhauser's (1986) classification of Hall's life-cycle theory (see Almoaily, 2012).

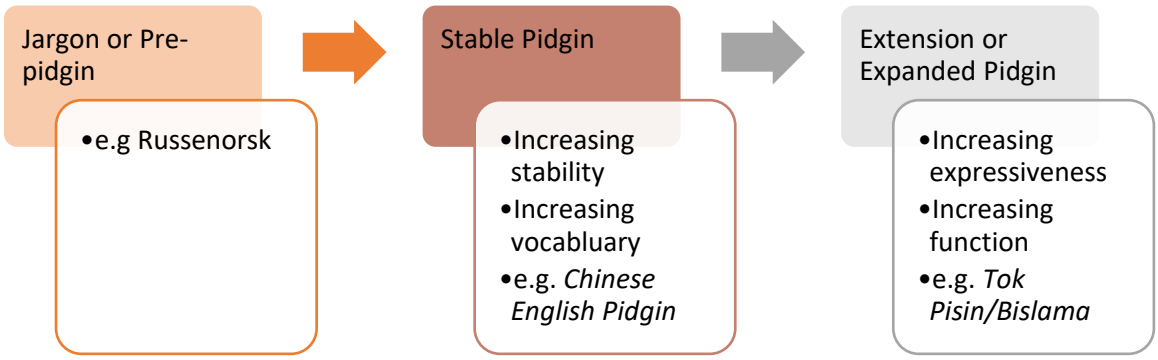



Figure 3.1: The stages of pidgin development (Hall, 1962).

These stages are captured by the complexity and stability of the pidgin's grammar and the degree of communicative functions in serving their purposes. Siegel (2008) indicated that linguistic features and the range of social functions for which they are used (e.g. limitation to trade) have been defining the stages of pidgin languages. This next section illustrates the common stages that pidgins undergo in their development.

#### *3.1.2.1.1.* The Jargon / Pre-Pidgin Stage

At an elementary stage of P/C development, instability is the first feature that characterises pidgin languages, and it is known as 'jargon'. Sebba (1997) defines *jargon* as the most basic and least structured phase in the development of pidgin. Moreover, Serrano, Garzón, and Manzanares (2003: 229) define a jargon as 'an unstable pidgin stage', and Field (2004: 131) refers to it 'as a linguistic variety that is even less developed than a pidgin'.

A jargon emerges in a contact situation between two speech communities that do not understand each other's languages; they subsequently start developing their own way of communication by using their own terms and vocabulary, which they have acquired from other languages, especially from the 'lexifier' (the most familiar language to both groups involved in the contact situation). Siegel (2008) describes jargon or pre-pidgin in a contact situation as individual attempts to communicate with others by using words and phrases from other languages, including the lexifier. In addition, Field (2004: 131) states that 'this variety is greatly influenced by substrate grammar'.

As regards the linguistic features, Jenkins (2003) reported that extended articulation and reiteration, which refer to the tendency to express ideas in lengthy sentences, are some of the linguistic features of jargon or pre-pidgin. Ansaldo and Matthews (2004) emphasised that repeating the same word, or stressing, is also a general linguistic feature in contact languages. In addition, Tarone (1980) distinguished between *foreigner talk* (FT) and jargon in the sense of their usages. Native speakers can use FT to simplify their speech when they speak to non-natives in one group only, while jargon is used by both groups in contact, typically the native speakers of the more prestigious language in the contact situation. Thus, the jargon stage is characterised by a high degree of instability and variability in its grammar and lexicon among its speakers. Sebba (1997) reported that this variation is based on the influence of the speakers' native languages on the structure of the jargon.

According to the model in Figure 3.1 above, the jargon may turn into either a pidgin or a creole depending on the duration of time that passes between the two speech communities. Moreover, Todd (2003) revealed that speakers of pre-pidgin are able to communicate with one another in a limited social function. For example, from the eighteenth until the early twentieth century, Russenorsk, an extinct pidgin language, was used between Russian sailors and Norwegian fishermen for the purpose of trade (Sebba, 1997).

Mesthrie (2008) believes that when a jargon is continuously used in a contact situation, this variety may develop into a new stabilised stage, which is called stable pidgin. The terms *pidgin* and *creole* are discussed next.

#### *3.1.2.1.2. The Stable Pidgin Stage*

The following period of the *pidgin-creole lifecycle* is the Stable Pidgin phase, where language is used not only for communication but for self-expression also. At this stage, Sebba (1997: 105) suggests that a *stable pidgin* may emerge with its norms of grammar, pronunciation and lexicon from an unstable jargon. In the same manner, there is a stronger sense of linguistic complexity as both simple and complex sentences are applied. According to Todd (2003: 1), pidgin is a 'marginal language that develops as a means of communication for limited communication purposes between people who do not have a language in common'. To distinguish between a pidgin and a jargon, Holm (1988: 5) stated that pidgin is 'more stable and has certain norms of meaning, pronunciation, and grammar'. Sebba (1997) pointed out that the most common feature of stable pidgin is a reduction of its linguistic variability. Stable pidgin has more established linguistic norms, which develop its own norms of grammar, pronunciation, and vocabulary (e.g. Chinese Pidgin English). These norms are often independent and distinct from their source languages and could be described as *pidgin* rather than jargons (Romaine, 1988). Tok Pisin and Fanakalo are examples of stable pidgins; however, these spoken languages have already moved beyond the stage of stable pidgins and have become *expanded pidgins.*

#### *3.1.2.1.3. The Extension or Expansion Pidgin Stage*

Once the stable norms of pidgin languages begin emerge, the process of development does not necessarily stop, as Sebba (1997) suggests: '[w]hat happens at this point largely depends on the uses to which the pidgin's speakers decide to put it' (1997: 106). In this sense, he means that if speakers of pidgin decide to narrow their use of the language to limited purposes (e.g.

trading from different countries), then ‘it is unlikely to develop significantly once it has stabilised’. However, this pidgin remains in its status unless its users decide to expand its communicative function and develop it into an extended or expanded pidgin. According to Sebba (1997), this scenario of pidgin development was initially postulated by Todd (2003: 5):

> An extended pidgin is one which, although it may not become a mother tongue, proves vitally important in a multilingual area, and which, because of its usefulness, is extended and used beyond the original limited function which caused it to come into being.

If pidgin enters this expansion phase with a high range of communication functions, and if its overall structure is elaborated in both grammar and vocabulary, then it potentially becomes linguistically sophisticated as a full-fledged language.

#### *3.1.2.1.4. Beyond Extension: From Pidgin to Creole*

As discussed above, pidgin languages are more stable than jargons, which are considered to be at an earlier stage in the *life cycle*. According to the *life-cycle* theory, some creolists, such as Hall (1966) and Mühlhäusler (1986), claim that the *life cycle* is supposed to progress ‘from Jargon, to Pidgin, to Creole, to Post-Creole by progressive structural expansion, stabilisation, and closer approximations of the lexifier–the language which contributed the largest part of a Creole's lexicon’ (Mufwene, 2002: 135). In addition, Mufwene (2001) emphasises that ‘Creoles and pidgins developed in separate places, in which Europeans and non-Europeans interacted differently – sporadically in trade colonies (which produced pidgins) but regularly in the initial stages of settlement colonies (where Creoles developed)’. The term ‘Creole’ was first used in the sixteenth century, in reference to non-indigenous people born in American colonies and descended from French, Spanish, or Portuguese settlers living in the West Indies, and later in English by the early seventeenth century (Bush-Caver and Williams, 2009, and Mufwene, 2002). Moreover, there is general agreement that the term Creole derives from

*crioulo*, the Portuguese word which means ‘a slave born in the master's household’ (Tureaud Jr, Trammell, Mark, Tureaud Sr, and Tureaud, 2018).

Until the late eighteenth century, creole languages may not have applied widely to language varieties. Mufwene (2002) stated that ‘[i]t is not clear how the term became associated only with vernaculars spoken primarily by descendants of non-Europeans…several speakers of Creoles (or pidgins) actually believe they speak dialects of their lexifiers.’ Furthermore, Bloomfield’s (1933: 474) argument was among the earliest claims that creoles developed from pidgins: ‘when the jargon [i.e. pidgin] has become the only language of the subject group, it is a creolized language’. Later, Hall (1962) reinterpreted Bloomfield’s claim and linked creolisation with nativisation. Thus, according to Mufwene (2002), creoles have been defined inaccurately as ‘nativized pidgins’. In this case, pidgin languages have acquired native speakers and have become the L1s of a speech community, and they have therefore expanded both their structures and functions and have stabilised.

However, some creolists (such as Alleyne, 1971; Arends, Muysken and Smith, 1995; and Bakker and Muysken, 1995) dispute this connection between creolisation and nativisation. Alleyne (1971) was the first creolist who argued that in Haitian Creole, the inflectional morphology was fossilised and that this is evidence that Europeans did not communicate with Africans using *baby talk* or *FT*, which would have fostered pidgins on the plantations. As for Muysken and Smith (1995), they mentioned that some extended pidgins (e.g. Tok Pisin and Nigerian Pidgin English) have actually gained native speakers.

To date, the debate in defining creoles relies mainly on two assumptions. First, creoles have been seen as nativised pidgins (as in Hall’s 1966 *pidgin-creole life-cycle* discussed above). The second assumption suggests that creoles separate languages from their lexifiers and are not necessarily preceded by a pidgin stage (e.g. Jourdan, 1991; Bakker and Muysken, 1995; Mather, 2001). Sebba (1997) is among the researchers who propose a definition of the term

creole that relies on Mühlhäusler's (1986) previously mentioned scenarios. In his definition, he endeavours to address the researchers' perspectives on the rise of creoles and distinguishes between two types of creoles. Sebba (1997: 136) defines a creole as 'a language with native speakers which results from contact without normal transmission'. Then, he suggests that the first set of creoles resulted from 'abrupt creolisation through a sharp break in the transmission of language in some community' (i.e. not preceded by stabilised pidgin), while the second set of creoles evolved as a result of 'nativisation'. In Sebba's definition, there are no standards for classifying a newfound contact language as either a pidgin or a creole.

Some creolists (e.g. Markey, 1982) have adopted an alternative approach to defining creoles, as the jargon-pidgin-creole model is still debated in the literature pertaining to contact languages. To classify creoles, those creolists have focused on the common linguistic features of creoles rather than on whether they have become nativised. The use of reduplication as a word formation process, a strict SVO order, and the use of adverbs to mark TMA are some of the components that have been offered as characteristics of creoles. However, McWhorter (2000: 85) critiqued this list of components as being insufficient; he claimed that many or some of these linguistic features are absent, and there are also 'non-creole languages which combine many of them'. However, if the definition of creoles as nativised pidgins is not accepted, and if there are no common linguistic features of creoles, then the following question arises: By what method would we be able to distinguish between P/Cs?

There are numerous accepted views in the literature on P/Cs regarding their emergences and definitions and whether jargons can turn into pidgins through the process of stabilisation, while pidgins turn into creoles via the process of nativisation. However, most of the challenging views concern the emergence of pidgins (i.e. gradual versus abrupt creolisation). I hence provide a brief discussion on their emergence and some alternative views.

### 3.1.3. Is the Emergence of Pidgin Gradual or Abrupt Creolisation?

As has been mentioned in Section 3.1.1.4, the LBH theory is possibly one of the most popular theories about abrupt creolisation. This theory was highly dominant in the 1970s and 1980s, albeit criticised by some researchers, such as Alleyne (1971) and Hancock (1980). Indeed, Tok Pisin and other contact languages, such as Sranan, Saramaccan, and Haitian, provide evidence for the gradual emergence of creoles and pidgins, since their historical developments do not correspond with the abrupt creolisation theory of the LBH. For example, Tok Pisin has changed through the process of its expansion over several generations, as reported by Arends et al. (1995) and Sankoff and Laberge (1974). Moreover, the authors demonstrated that the linguistic structures of Tok Pisin do not change during the process of nativisation within one generation. Regarding whether the development process from a pidgin to a creole is an abrupt or a gradual one, Bakker and Muysken (1995) claim that Hawaii Creole English is the only strong evidence in history showing that a creole was preceded by a pidgin. Therefore, they emphasise that due to the structural differences between P/Cs (for example, nearly all creoles have an SVO word order, while pidgins vary), it is problematic to assume that all creoles are derived from pidgins. They thus suggest that if creoles developed from pidgins, then all P/Cs should have similar structures. Other researchers take intermediate positions (e.g. Mühlhäusler 1986, Sebba 1997, and Siegel 2008a); they state that although some creoles have developed directly from pidgins, other creoles have actually emerged without a preceding pidgin.

In the next section, I provide historical background on the development of the study of P/C languages, which were once considered marginal languages in the academic research. I also discuss the attitudes of scholars and the public towards contact languages before and after P/Cs were perceived as an academic field of study. This will enable the identification of some of the significant patterns in the history of the study of P/Cs.

### 3.2. Contact Languages as an Academic Field of Study

Holm (1988: 13) stated that the investigation of P/Cs has been ignored for centuries in spite of the fact that 'language contact seems likely to be nearly as old as language itself'. Pidgins were previously considered to be broken, low-status mixtures of languages that did not merit formal study, with Holm (1988: 1) and McArthur (1998: 161) stating that P/Cs have been assigned humiliating or derogatory names, such as *nigger French*, bastard Portuguese, broken English, cookhouse lingo, and coolie language. This demonstrates how P/Cs were treated before being recognised as an academic field of study.

Apart from their lack of esteem, pidgins are also not as linguistically sophisticated as full languages (Singh, 2000). It is true that in the past, people viewed contact languages as primitive languages due to their tendency to have minimal or simplified structures. In fact, contact languages remained marginal and were only recognised as a field of linguistics in the late 1950s and early 1960s, following the works of Robert Hall and Douglas Taylor (Holm, 1988 in Almoaily, 2012: 19). Since then, an extensive number of researchers have gradually recognised the value of the investigation of P/Cs as linguistic subdisciplines. According to Bickerton (1974), researchers were afforded significant opportunity to examine and develop theories within the field of general linguistics by conducting creole studies. In this regard, the study of P/C languages in the field of historical linguistics can help in the confirmation or rejection of many theories. Lefebvre (2004: 7) thus described P/C languages as 'a goldmine for historical linguistics'.

I briefly outline some eras in the history of P/Cs. According to Holm (1988), the history of P/C languages can be ordered into three major historical eras. The first era is before the sixteenth century (i.e. the period before European interference). The second is during the European expansion (i.e. the period from the sixteenth century to the first half of the twentieth century), and the third era is after the establishment of P/Cs as a scholarly field of study. This last era

extends from the middle of the last century to the present. These three historical periods are discussed below, focusing on some significant turns in the discovery and documentation of contact languages.

According to Hall (1966) and Holm (1988), among others, contact languages are believed to have existed as early as groups of people with different languages began to communicate with one another. Moreover, slavery, migration, trade, and colonisation are believed to be the reasons behind the emergence of contact languages at that time. Despite the scarcity of documented P/C languages, Holm (1988) reported that there were only two documented contact languages prior to the European expansion (i.e. before the sixteenth century). Maridi Arabic is the first known text of a contact language; it is a trade pidgin language thought to have been spoken in either Mauritania (Thomason and Eljibali, 1986) or Sudan (Owens, 1996). Almoaily (2012) suggested that this restricted Arabic variety might date back to the eleventh century AD, and the substrate language of this pidgin is possibly Nilo-Saharan. However, Souag (2006) added that clear evidence for this hypothesis and for the place where Maridi Arabic used to be spoken is lacking. Furthermore, Al-Bakri[3], an Arab geographer who cited some text from this restricted Arabic variety, described Maridi Arabic as a deterioration of the Arabic language in all of his writing (Almoaily, 2012).

In addition, Sabir was a Lingua Franca of the Mediterranean that was attested as another early contact language. According to Holm (1988), its lexicon is mainly based on a mixture of Southern Romance languages and also contains a mixture of other substrate languages, such as Arabic, Berber, and Turkish. The first available text of Sabir was recorded in the fourteenth century in Djerba and Tunisia (Holm, 1988).

[3] The geographer of Muslim Spain, Abu ‘Ubayd ‘Abd Allah, b. ‘Abd al-‘Aziz al-Bakri (d. 487/1094).

Holm (1988) stated that due to the growing number of researchers, missionaries, and anthropologists who became interested in contact languages in the era of the European expansion (i.e. from the seventeenth to the twentieth century), the documentation of contact languages flourished. This documentation has contributed greatly to the study of contact languages, even before pidginisation and creolisation were established as an academic field of study in the 1960s. Naro (1978) stated that in 1516, Negro Portuguese Pidgin was the first documented language of a European-language-based contact language.

In the 1730s, the Moravian Church's attempt to communicate in Dutch with slaves in Suriname was considered the first serious study to treat a creole as an independent language. When the Church's first attempt to communicate with slaves failed, Moravians began to learn Negerhollands, the language spoken by the slaves (see Holm, 1988).

During the nineteenth century, contact languages were generally perceived to be trivial and rudimentary forms of speech until Greenfield (1830), who suggested that creole languages are not broken forms of language or degraded tongues but rather rule-governed languages (Almoaily, 2012). Therefore, Greenfield's position can be considered a major turning point in the attitudes towards P/C languages (Holm, 1988). In the 1960s, P/C languages were recognised as a field in linguistics for the first time. DeCamp (1977) reported that the number of researchers working on P/C languages had grown from a dozen to a hundred by the end of the 1970s. Today, the number of European-language-based P/Cs is higher than non-Indo-European-language-based ones (see the discussion in Section 3.4.1) despite major advances in the field of P/Cs over the past 50 years. This could be a result of the lack of research and documentation on non-Indo-European-language-based contact languages. In fact, it is highly possible that a large number of P/Cs worldwide are still undiscovered, which thus calls for more extensive documentation and investigation of these languages, especially non-Indo-European ones such as Chinese, Arabic, and Indonesian.

In the next section, I review the typological features of contact languages and investigate the extent to which these features are in fact evidenced in the atypical (non-Indo-European) contact languages. Then, in the subsequent subsections, I review the literature of the pidgin under investigation.

### 3.3. General Linguistic Features of Pidgin and Creole Languages

In this section, I briefly investigate the common characteristics which have been reported across P/C languages. Simplification of linguistic structure is a common feature of pidgins that concerns all aspects of grammar: lexicon, phonology, syntax, semantics, and morphology. Accordingly, the grammar of pidgins is characteristically less complex than that of their source languages. In this section, I concentrate on the morphosyntax, as this linguistic category is the focus of this thesis, and I briefly discuss the segmental phonology, as this linguistic feature plays a major role in the process of pidginisation and creolisation. I also focus on pidgin languages, whereas creole features are discussed in less detail, since the majority of researchers have classified GPA as a pidgin (see the next sections for a detailed discussion about the classification of GPA).

It is important to note that the majority of research reported in the literature on P/Cs tends to be based on Indo-European superstrate and West African substrate P/Cs. Therefore, the resulting general features might be affected by the superstrate and substrate languages involved in the contact situation and thus not be a reflection of the systemic features of all P/Cs spoken worldwide (as discussed later in Section 3.4). However, some researchers, such as Bakker, Daval-Markussen, and Parkvall (2011), have conducted empirical research examining 29 Indo-European-based and five non-Indo-European-based pidgins, and they have concluded that Indo-European and non-Indo-European contact languages do not behave differently in terms of linguistic features (e.g. highly simplified grammatical structure compared to the structures of their source languages). In general, reduction is the main characteristic which distinguishes

P/Cs from other 'normal languages'. For example, P/Cs are believed to have a reduced, if not absent, inflectional system, reduced derivation, and a small inventory of function words. Some researchers, such as Bakker and Muysken (1995), Sebba (1997), and Siegel (2004), indicated that the amount of reduction in pidgins is higher than in creoles. The reduction in segmental phonology and in the morphosyntax of P/C languages is discussed in more detail next.

### 3.3.1. Segmental Phonology

Phonological universals could have played a role in the pidginisation and creolisation process, as Holm (1988) and Bakker and Muysken (1995) have suggested. For instance, the phonemes found in the majority of the world's languages, such as /t/, /d/, and /m/, are more easily transmitted into P/Cs than less common phonemes such as /θ/, /v/, and /ð/, since phonemes exist in Arabic that are typologically marked, such as /ħ/, /ʕ/, and /χ/ (see the *World Atlas of Language Structures Online*[4]). Therefore, GPA speakers could be expected to use a simplified phonological system replacing such phonemes with more commonly used phonemes (see Bakker and Muysken, 1995). For instance, Akan speakers of Ghanaian Pidgin English have been found to replace the phoneme /v/ with /b/ or /f/ due to the absence of /v/ in the Akan inventory (Huber, 1999).

Reduction has taken place on the phonological level; the consonant inventory in pidgins is usually reduced. Similarly, McMahon (1994) found that vowels are usually fewer than their lexifier counterparts, and length distinction is lost (see Albaqawi 2016 for a more detailed account of GPA phonology). Furthermore, the consonantal system adopted by Asian migrants (Table 3.1) seems to be consistent within GPA, independently of the differences in the phonologies of the respective GA dialects that may have served as a TL for Asian migrants in different parts of the Gulf.

[4] It surveys the typology of the phonology and morphosyntax of most of the currently known languages (edited by Dryer and Haspelmath, 2001).

| | Labial | Inter-dental | Alveolar | Alveo-palatal | Palatal | Velar | Uvular | Pharyngeal | Glottal |
|---|---|---|---|---|---|---|---|---|---|
| Plosive | p | | t | | | k | | | ʾ |
| | b | | d | | | g | | | |
| Fricative | f | | s | š | | | | | h |
| | | | z | | | | | | |
| Affricate | Voice-less | | | ĉ | | | | | |
| | Voiced | | | j | | | | | |
| Liquid | | | l r | | | | | | |
| Nasal | m | | n | | | | | | |
| Glide | w | | | | y | | | | |

Table 3.1: Consonantal phonemes of GPA in linguistic symbols (Albaqawi, 2016).

### 3.3.2. Morphosyntax

Since the grammar of pidgins is characteristically less complex than that of their lexifier languages, Sebba (1997) refers to this reduction system by general or principal features ascribed to pidgin grammar. Some of these features are also valid for creoles, even though they are geographically widely separated and are based on different lexifier languages.

#### *3.3.2.1. Lack of Morphological Complexity*

In the field of pidgin typological studies, the consensus among researchers, such as Todd (1980), Hudson (1996), Holm (2000), and many others, is that pidgins are characterised by a lack of morphological complexity, which is due to either partial or complete lack of inflections (i.e. number [singular/plural], case [nominative/accusative], tense, gender, or grammatical agreement). Pidginisation 'may involve a shift from synthetic to analytic morphology' (Roberts and Bresnan, 2008: 270-271); the authors arrived at this conclusion after examining the retention of inflection in 29 languages that reflect a history of pidginisation in their development, comparing the morphological richness of pidgins with their respective lexifiers (i.e. having morphological systems which use free morphemes instead of inflections).

Analytic morphology also seems to be typical of creole languages. Bakker and Muysken (1995), for example, argue that the only common feature between restricted P/Cs is reduced inflection, and Romaine (1988) stated that contact languages are reported to have a reduction in agreement markers. Almoaily (2012) took an alternative perspective when he examined some common traits in the morphosyntax of 10 non-Indo-European pidgin languages, as listed in Table 3.2, to assess their compliance with the proposed typological features of contact languages. The purpose of his investigation was not to describe the pidgin under comparison but to show the existence or absence of affixes. The sign + indicates that the affix(es) marking the linguistic feature in question exists in the pidgin, while - indicates that affixation is missing.

| P/C / AFFs for | Fanakalo | Kituba | Lingala | A-70 | Restr. Sango | Restr. Swahili | Pidgin Madame | Hiri-Motu | Naga Pidgin | Pidgin Fijian |
|---|---|---|---|---|---|---|---|---|---|---|
| Tense | +(12→6)[5] | + | + | - | - | - | - | - | + | - |
| Mood | - | - | - | - | + (irrealis tone only) | - | - | - | - | - |
| Aspect | - | + | + | + | - | + | - | - | - | - |
| showing S-V Agreement | + | + | - | + | + | - | - | - | + | - |

Table 3.2: Inflection in some non-Indo-European pidgin languages (adapted from Holm, 1988; Roberts and Bresnan, 2008; Bakker, 2011).

Below, I briefly define each of these non-Indo-European-language-based pidgins as Almoaily (2012) defined them by providing the region in which the pidgin is spoken and the languages in contact during its creation:

**Fanakalo** is a pidgin spoken in South Africa. Its superstrate language is Zulu, and the substrate languages are English and/or Afrikaans (Mesthrie, 1989).

[5] The number of affixes for tense in Fanakalo is reduced from 12 in the superstrate language to only six in the pidgin.

**Kituba** is spoken in the Congo. Its lexifier language is Kikongo, and the substrate languages are French and Lingala (Ethnologue, 2011).

**Lingala** is an expanded pidgin spoken in the Congo (Smith, 1995). The lexifier is Bobangi, and other languages in contact are Lusengo and Bangala (Ethnologue, 2011).

**A-70** is a pidginised variety of the Bantu languages Ewondo and Bulu, spoken in Cameroon (Sebba, 1997).

**Restructured Sango** is a pidgin spoken in the Central African Republic. Its lexifier language is Ngbandi, and other languages in contact are French and English (Thomason, 2001).

**Restructured Swahili** is a Swahili-based contact language spoken in the Congo (Holm, 2000). According to Wardhaugh (2009), the substrate – and possibly adstrate – languages are English and some other African pidgins.

**Pidgin Madame** is an Arabic-based pidgin spoken in Lebanon. The superstrate language of this pidgin is Lebanese Arabic, and the substrate language is Sinhala (Bakker, 2011).

**Hiri Motu** is a pidgin spoken in Papua New Guinea. The lexifier language is Motu, with Pidgin English and other Papuan languages also involved in the language contact (Thomason, 2001).

**Naga Pidgin** is spoken in Bangladesh. Languages involved in the contact are Assamese and Bengali (Holm, 1989).

**Pidgin Fijian** is spoken in Fiji as a lingua franca between Indians, Chinese, and the indigenous people of Fiji (Tryon and Charpentier, 2004).

Table 3.2 above indicates that 4 of the 10 non-European pidgins listed in the table use affixes to mark for tense. Similarly, to mark for aspect, affixes are used in four pidgins in the sample, and to mark for pluralisation, affixes are used in half of the sampled pidgins. Thus, with the exception of the morphosyntactic feature of mood – because it is the only grammatical information that is consistently not marked in this set of non-Indo-European Pidgin languages – no generalisation can be made about the use of inflectional morphology to mark for any of

the categories listed in Table 3.2. Therefore, Almoaily's (2012) findings run somewhat contrary to what has been commonly stated in the literature of P/Cs, namely, that pidgins lack affixation. Bakker (2003) tested the common notion that morphological inflection is always either reduced or absent in pidgin languages. After examining the data of more than 30 pidgins, including some less studied pidgins (such as Asmara Pidgin Italian and GPA), he concluded that pidgins are richer in morphology than creoles. Half of the pidgins he surveyed contained a form of inflection; for example, the suffix *-ato* in Asmara Pidgin Italian is used to mark for past participle ending (with -o further marking grammatical gender – male). Likewise, the prefix *ta-* in Kenyan Pidgin Swahili is used to mark for future tense. However, Bakker's claim was criticised by Roberts and Bresnan (2008) when they highlighted that in Bakker's (2003) list, none of the pidgin languages had more inflectional morphemes than their input language. In fact, this criticism of Bakker (2003) is also highlighted by Almoaily (2012). Accordingly, this suggests the need for further analysis of atypical P/Cs, since the claim that inflection is reduced as compared to the source language has not been convincingly refuted.

In summary, reduced inflectional morphology in contact languages is a widespread belief within P/C typologies. This conclusion might be subject to revision, since it has been made on the basis of Indo-European / West African pidgins. As verified by Bakker (2003) and Almoaily (2012), taking non-Indo-European P/C languages into typological account reveals some difficulties with this assumption. Thus, more extensive work that includes as many pidgin languages as possible is required to check the validity of Almoaily's claim that the morphosyntactic systems of pidgin languages contain fewer affixes than full languages. The next section addresses the claim that pidgin languages have a reduced inventory of function words.

### *3.3.2.2. Lack of Grammatical Complexity*

Pidgins have traditionally been defined as languages with limited or highly reduced lexicon and grammar and lacking the range of grammatical devices compared to 'full' languages (Van Craenenbroeck et al., 2008). In the early stages of the development of pidgin languages, they are 'stripped of everything but the bare essentials necessary for communication' (Romaine, 1988: 24). This indicates that pidgin languages have few, or even totally lack, function words, particularly at their inception. In this section, I discuss some common features of function words in P/C languages, as these linguistic features are later examined among other features as part of the linguistic features under investigation: definite and indefinite articles, copulas *be*, and pronouns.

1. **Lack of Definite or Indefinite Article**

Researchers such as Sebba (1997) and Samarin (2000) have stated that pidgins in general do not have overt definite or indefinite articles (i.e. 'the' or 'a'). This is the case in GPA: The definiteness marker in GA is normally dropped in GPA (i.e. there are no indefiniteness markers in the superstrate language – neither GA nor GPA, see Chapter 4). Definiteness is also a grammatical feature that is conspicuously absent in GPA (Al-Azraqi, 2010). In addition, Almoaily (2012) stated that GA definiteness markers are usually dropped in GPA. However, with regard to this view that pidgins do not have markers for definiteness or indefiniteness, Versteegh (1984) argued that those pidgin languages create definite articles out of demonstratives. Regarding creoles, Holm (1988) argued that definiteness markers are used commonly after the noun and mark the end of the noun phrase (NP). The claim that pidgins lack definite articles is tested in Table 3.3. The table shows that markers are utilised for either definiteness or indefiniteness in Lingala, Restructured Sango, Hiri Motu, and Pidgin Fijian, whereas Fanakalo has markers for both definiteness and indefiniteness.

| P/C / Feature | Fanakalo | Kituba | Lingala | A-70 | Restr. Sango | Restr. Swahili | Pidgin Madame | Hiri Motu | Naga Pidgin | Pidgin Fijian |
|---|---|---|---|---|---|---|---|---|---|---|
| Definite articles | + | - | + | - | + | - | - | - | - | + |
| Indefinite articles | + | - | - | - | - | - | - | + | - | - |

Table 3.3: Definiteness and indefiniteness markers in some non-Indo-European pidgins (adapted from Holm 1988, Roberts and Bresnan 2008, Bakker, 2011).

Among the contact languages that Almoaily investigated, such as the 10 non-Indo-European languages above, it seems that the claim of the non-existence of definiteness or indefiniteness markers is not a defining characteristic of pidgin languages. These results only reflect a small, select sample of pidgin languages; therefore, the non-existence of those markers cannot be a generalised feature of pidgins. It thus appears that the non-existence of definiteness or indefiniteness markers is not a describing feature of pidgin languages, since they can be found in a number of the contact languages investigated in Almoaily's study. The next part of this section discusses copulas in P/C languages.

**2. Lack of Copula 'Be'**

Ferguson (1971), Romaine (1988), McWhorter (1995), and Sebba (1997) agree with the traditional view that both P/C languages lack copulas. However, full agreement among researchers, such as Arends, Muysken, and Smith (1995), has not been reached regarding whether copulas are absent, particularly for creole languages. Indeed, some creole languages have created forms of the verb *be*, which Holm (1988: 174) referred to as 'more complex than their lexical source languages'. By contrast, in pidgin languages, there seems to be an agreement that copulas are uncommon and rare. Siegel (2008b: 26) commented on this assumption as a defining diagnostic of restricted pidgin languages, stating that 'Pidgin Fijian also differs from other restricted pidgins in its use of a copula'. In addition, Rickford (1998) argued that the absence of the copula in African American Vernacular English is evidence that the use of copulas originates from a pidgin. Moreover, most non-typical pidgin languages, such

as China Coast Pidgin (Ansaldo, Matthews, and Smith 2011) and Romanian Pidgin Arabic (Avram, 2010), have an absence of the copula. In contrast, Baker (1987), Faraclas (1988), and Williams (2000) reported that copulas exist in some pidgin languages, such as Chinese Pidgin English, Nigerian Pidgin English, and Yimas-Arafundi Trading Pidgin, respectively. A comparison of the existence versus absence of copulas in the selected 10 non-Indo-European language pidgins is presented in Table 3.4 (Almoaily, 2012).

| P/C / Feature | Fanakalo | Kituba | Lingala | A-70 | Restr. Sango | Restr. Swahili | Pidgin Madame | Hiri Motu | Naga Pidgin | Pidgin Fijian |
|---|---|---|---|---|---|---|---|---|---|---|
| Copula | - | + | + | N/A | + | +[6] | + | + | + | + |

Table 3.4: Copulas in non-Indo-European pidgin languages (adapted from Holm; 1988; Roberts and Bresnan, 2008; Bakker, 2011).

The table shows that copular verbs exist in eight languages except for Fanakalo (Asher and Simpson, 1994), while Pidgin A-70 was excluded due to a lack of available data. This evidence is in line with the fact that the copula exists in some Arabic-based pidgins, such as GPA and Pidgin Madame, although in Arabic, the copula is not used in the present tense. This statement 'could be an argument against the claim that pidgins typically have null copula' (Almoaily, 2012: 31).

### 3. Pronouns Lack Inflections

The number of varieties in the pronominal systems of pidgin languages is limited (Mühlhäusler, 1986). As an example, the pronouns in Melanesian Pidgin are not marked for the accusative case, gender, and plural. For instance, *em* stands for both 'he' and 'him', as well as 'she' and 'her'; likewise, *ol* means 'they' and 'them' (Sebba, 1997). Moreover, the dropping of pronouns is another universal property in the pronominal system of pidgin languages (see Mühlhäusler,

[6] The copula occurred once in a short, translated text by Holm (1988) in his collection of texts.

1986; Romaine, 1988; Schumann, 1986). Romaine (1988, 1990) claimed that the pro-drop feature is an unmarked form in pidgin languages, while Bresnan (2000) indicated that there is a prevalence of free pronominal morphemes in pidgin languages: 'Pidgins prefer free pronoun forms to bound ones' (Mühlhäusler and Harré, 1990: 262). Moreover, for creole languages, their pronominal systems have been claimed to be reduced (see Bailey, 1966 and Valdman, 1978). However, Holm (1988: 202) proposed that there is 'evidence that creolisation does not necessarily lead to extreme morphological simplicity in pronominal systems'. For instance, Portuguese-based creoles spoken in the Gulf of Guinea used independent, strong pronominal systems, which are as complex as those used in Portuguese.

Almoaily (2012) tested the claim that pidgins have free rather than bound pronouns, as indicated in Table 3.5. The table shows that 5 of the 10 non-Indo-European pidgin languages have bound pronouns.

| P/C / Feature | Fanakalo | Kituba | Lingala | A-70 | Restr. Sango | Restr. Swahili | Pidgin Madame | Hiri Motu | Naga Pidgin | Pidgin Fijian |
|---|---|---|---|---|---|---|---|---|---|---|
| Copula | - | - | + | + | + | - | - | - | + | - |

Table 3.5: Existence vs. absence of pronominal clitics in some non-Indo-European pidgins (adapted from Holm, 1988; Roberts and Bresnan, 2008; Bakker, 2011).

To sum up this section, within P/C typology, there is a widespread belief that contact languages have a considerable lack of grammatical complexity. The assumption that copulas and definite articles are common in creoles but not in pidgins seems to be invalid, certainly amongst the small sample of non-Indo-European-language-based pidgins used in this study. For example, 8 of the 10 non-Indo-European pidgins polled in this study have copulas, in contrast to the traditional view that pidgin languages lack a copula. The next section addresses the claim that pidgin languages have a reduction in word formation.

### *3.3.2.3. Reduction in Word Formation*

Hymes (1971) stated that pidgins have limited lexical stock. The lexicon of P/Cs has been characterised by reduced derivation, semantic transparency, and reduplication. These proposed lexical features across P/C languages are discussed in this section.

**A. Reduction in Lexical Items**

One of the most widely cited characteristics which distinguish pidgins from other 'normal languages' is that pidgin languages have a limited lexicon compared with non-contact languages, as noted by Samarin (1971), Romaine (1988), McMahon (1994), Bakker and Muysken (1995), Sebba (1997), and many other researchers. In fact, the smaller number of lexemes in pidgin languages is nonetheless still able to express most of the semantic functions which can be communicated in full languages (Romaine, 1988 and 1992).

Sebba (1997) mentioned that there are so-called *all-purpose prepositions* which shape the lexicon of a pidgin language. For example, Tok Pisin makes use of the preposition 'long', which can be translated into English as, inter alia, 'in, on, at'. Moreover, he explained that synonymy is almost always absent in pidgins, and words in pidgins are also often highly polysemous. Bakker and Muysken (1995) presented a prime example of divergent polysemy in Chinook Jargon, where the word *muckamuck* has many equivalents in English, including 'eat, drink, and bite'. Romaine (1988) reported that the term *kato* in Pidgin Fijian can express four lexical items in the Fijian language.

Furthermore, the multifunctionality of lexical items in pidgin languages enables one word in a given pidgin to function as a noun, a verb, and an adjective. This multifunctionality, among other things, also enables the lexical items in pidgin languages to express all the semantic functions found in full languages. For example, *sik* ('sick') in Tok Pisin can function as a noun as well as an adjective. Mühlhäusler (1987) adopted the concept of 'maximum use of a minimum lexicon' which reflects these two phenomena (i.e. multiple meanings and functions

for a single item). In contrast with pidgin languages, creole languages are characterised by lexical expansion. Alleyne (1980) and other researchers have stated that the lexical expansion in creole languages is due to the intensive implementation of morphological processes, such as reduplication (see D below) and compounding (Alleyne, 1980; Crowley, 2008; Hancock 1980).

**B. Reduction in Derivational Morphology**

Bakker and Muysken (1995), Sebba (1997), Smith (2008), and Versteegh (2008) reported that analytic morphology is the most common feature in both the inflectional and derivational morphology of P/C languages. Mühlhäusler (1986) indicated that in some pidgin languages, compounds are used to create new meanings from one word instead of using derivational affixes. For example, the words for 'work' and 'man' in Tok Pisin are compounded to create a new meaning: *wroko.man*, or 'work.man' (i.e. worker). To overcome the shortage of derivational affixes in the lexicon of contact languages, there are some word formation processes, which are discussed below: compounding with semantic transparency and reduplication.

**C. Semantic Transparency**

Baayen and Schreuder (2003) use the term semantic transparency to refer to compounds whose meaning is easily guessed by the meanings of its constituents. Sebba (1997) argued that in non-contact languages, such as English, Arabic, or Spanish, the relationship between form and meaning is weaker than in pidgin languages. Simplicity is the main feature of pidgin languages, and according to Seuren and Wekker (1986) and Sebba (1997), semantic transparency occurs to achieve simplicity in pidgin languages. Sebba (1997) presented an example from Tok Pisin where gender is indicated by the use of a separate word: *man* for the male and *meri* for the female, as in the words *kakaruk* ('chicken'), *kakaruk man* ('rooster'), and *kakaruk meri* ('hen'). In another example by Sebba (1997), lexemes can take on derivational status, since they are routinised as agentive morphemes, as in *kam.man* (come.man) 'new arrival' and *mas.man*

(march.man) ‘marcher’. As illustrated in these examples, semantic transparency is an alternative way in which to use derivational affixes, such as the English *-er* (e.g. *Backer*) and *-ess* (e.g. *Waitress*). Moreover, to create a new meaning in pidgin languages, repetition of a morpheme is another word formation process which is frequently reported for pidgin languages, as discussed in further detail next.

**D. Reduplication**

Holm (1988: 88) defines reduplication as a ‘[...] mechanism for forming new words. It involves the repetition of a word (or part of a word) resulting in a distinct lexical item slightly different in meaning’. Reduplication is a distinguished process that is different than iteration, with the former employed to devise a new meaning, and the latter utilised for emphasis. Bakker and Muysken (1995: 33) stated that ‘[t]he morphological process of reduplication is common (but not universal) in creole languages, but, strangely enough, rare in pidgins’. An example of meaning-forming reduplication in P/C languages is in the Haitian Creole French language: The word *yun* (one) is reduplicated to *yunyun* to create a new meaning: ‘distribute’. Another example is *vroevroe* in Negerhollands Creole Dutch, where the word *vroe* (‘early’) is reduplicated to express a new meaning: ‘morning’. In addition, reduplication is used in a range of pidgin languages, such as Pidgin Maori (Bell and Holmes, 1990) and Indian Pidgin English (Mehrotra, 1997). Bakker (1995), Mühlhäusler (1997), Bakker (2003), and Bakker and Parkvall (2005) claim that reduplication represents a diagnostic feature which distinguishes creoles and expanded pidgins from jargons and stable pidgins, as also is evidenced by Avram’s (2011) finding.

This section concludes that derivation is less used in P/C languages. Moreover, to overcome the shortage of such reduction, compounding and reduplication have been adopted – whether to use a single word to cover many meanings or to create many meanings out of limited lexicon.

In summary, the discussion of the proposed common morphosyntactic features of P/C languages in Section 3.3.2 revealed that the assumption that morphosyntactic complexity is reduced in pidgin languages needs to be revised. For instance, the tendency of P/C languages to be analytic rather than synthetic. Many typological studies on P/Cs suggest that creoles have copulas while pidgins normally do not, creoles restrict word order to SVO, whereas pidgins have free word order, and finally TMA preverbal elements, definite articles, and reduplication are common in creoles but not in pidgins. However, conducting a wider typological account of the features of P/C languages would be interesting and might reveal that P/Cs contain a number of typologically less or more features than they are thought to have. Indeed, the findings of this thesis suggest that it might be worthwhile to revisit many current assumptions about the structure of pidgin languages using a larger and more typologically diverse sample. Refer to Section 8.3.2 for a discussion on the compliance of GPA with the proposed general features listed above.

In the following section, I shed light on some limitations that reveal from the above discussion in Sections 3.3.2.1, 3.3.2.2, and 3.3.2.3 that the literature of P/C languages still suffers from some insufficiency: 1) lack of agreement in defining P/Cs and 2) and lack of consensus on the features which define P/Cs. The purpose of the discussion is to raise the scholarly awareness of the current limitations in the literature of P/Cs and, if possible, how they might be resolved or minimised.

### 3.4. Limitations in the Literature of Pidgins and Creoles

In this section, I shed light on some limitations which, from the above discussion in Sections 3.3.2.1, 3.3.2.2, and 3.3.2.3, reveal that the literature of P/C languages still suffers from some insufficiency and appears to have experienced three main limitations, as stated by Almoaily (2012). The first is that the literature has been built up by a European-centric view of contact languages, which might have led to a biased view in defining P/Cs – both in classifying their

typological features and proposing theories about their origin and emergence. The second source is the deficiency of available data for P/Cs. Finally, there is a lack of consensus in defining P/Cs and hence difficulty in making a clear distinction between these two types of contact languages.

The purpose of the discussion is to raise scholarly awareness about the current limitations in the literature on P/Cs and, if possible, about how they might be resolved or minimised. These three aforementioned factors are discussed in more detail in the following subsections.

### 3.4.1. European-Centric View

Most recent studies on P/C languages are the result of the investigation of contact languages based on European lexifiers, such as English, French, Dutch, Portuguese, and Spanish (cf. Holm, 1988, 2000; Todd, 2003; Arends et al., 1995; and many others, with a few exceptions such as Bakker, 2003; Versteegh, 2008). However, Western cultural biases and geographical boundaries hinder the production of a more encompassing account of the world's P/Cs (Almoaily, 2012). Although scholarly awareness of these limitations has grown since the 1980s (see Holm, 1988; Romaine, 1988), a European-centric bias still seems to pervade this field of linguistics. Almoaily (2012) explained the reason behind this bias: the pioneering European scholars who documented most of the currently known contact languages. Another reason for the European-centric view is the insufficient investigation of non-European contact languages, which could in turn have had a considerable effect on the current conception of P/Cs. For instance, the definitions provided in Stanford University's reference guide for P/C languages (2005) state that

> By definition Pidgins and Creoles involve language mix, and currently spoken Creole languages arose as a direct result of European Colonial expansion. Between 1500 and 1900, there came into existence, on tropical islands and in isolated sections of tropical littorals, small, autocratic, rigidly stratified societies, mostly engaged in monoculture, which consisted of a ruling minority of some European nation and a large mass of (mainly non-European) laborers, drawn in most cases from many different language groups. (Stanford University Research Guide: Pidgin and creole languages Introduction, paragraph: 2)

Similarly, Samarin (1982) claimed that non-Indo-European contact languages can only evolve under the influence of Europeans. Moreover, alternative names for those languages have been used; for example, pseudo-pidgin, creoloid, semi-pidgin, and secondary hybrid have deliberately been used to designate contact languages which were not based on African or Atlantic languages, because these contact languages were not intelligible to the early European researchers who first documented the varieties spoken in the European colonies between the sixteenth and the first half of the twentieth century (see Stewart, 1962 and De Granda, 1968). These terms, as Versteegh (2008: 161) suggests, have been given to non-Indo-European, non-Atlantic P/Cs 'in order to avoid assigning true pidginhood or creole status to cases outside the restricted corpus of Atlantic creoles'.

There are two fundamental ways in which to challenge European-centric views. First, some studies, such as those by Bakker (2003), Avram (2010), Bakker, Daval-Markussen, Parkvall (2011), and Almoaily (2012) – see also the description of GPA in Section 4.2 – suggest that European- and non-European-based P/Cs are structurally relatively similar. Therefore, classifying them as qualitatively different from European-based ones seems to be a typologically false notion (Almoaily, 2012). Second, Almoaily asked, 'why do some linguists feel the need to terminologically distinguish between the two?', since both European and non-European contact languages have evolved as a result of language contact between at least two groups of people speaking different languages, similar to the circumstances of L2 learning by adults in the case of pidgins.

Indeed, further investigation into P/Cs with non-European input and a comparative account with European-based contact languages will allow for further understanding of the nature of P/Cs and how more accurate theories can be formulated about their emergence, as well as how more precise typologies of the typical features of these languages can be established. For instance, Lefebvre's (1998) claim is a prime example of the European-centric view on the

emergence of P/Cs. He said that the emergence of pidgins can typically result in communities of contact speech, where the substrate language is spoken by the majority of a population, and the superstrate language is the minority language. This seems to have been the case for many European-language-based P/Cs, but not necessarily for all contact languages. However, the case is completely the opposite in the development of GPA (see Almoaily, 2008, 2012) and Pidgin Madame (see Bizri, 2010). In such Arabic-based pidgin contact languages, the superstrate languages are the languages spoken by the majority of the speech community, and the substrate languages are the contact languages only spoken by the minority. This might show that some pidginisation and creolisation contact languages actually evolve in different circumstances than what researchers in the field have considered to be the norm, based solely on data from Indo-European languages.

Importantly, a more extensive examination of lesser-known P/C languages might lead to more accurate inventories of the typological features of contact languages (Almoaily, 2012). Some considerable efforts have been made, for example by Bakker (2003) and Bakker, Daval-Markussen, and Parkvall (2011), to compare Indo-European- with non-Indo European-based P/Cs. As I have argued above, further comparative work is required because many non-Indo-European-based P/C languages are still under-researched. Thus, one of the aims of this thesis is to determine the extent to which a less studied Arabic-lexifier pidgin, namely, GPA, complies with the proposed typological features of P/Cs. This leads us to the next limitation in the literature of contact languages: the deficiency of available data for non-Indo-European P/Cs.

### 3.4.2. Insufficient Data on Pidgins and Creoles

One of the most common limitations that researchers in the field of P/C studies face is the shortage of data on describing some particular language or the difficulties in interpreting the available data for a contact language. This limitation might lead researchers to make hard decisions in refuting or verifying theories about the evolvement and typology of P/Cs

(Almoaily, 2012). Testing the accuracy of the available data of a given pidgin or creole is considered one of the most well-known restrictions for researchers. For instance, Reinecke (1975) argued against Zyhlar's (1932) suggestion that the ancient Egyptian language of hieroglyphic symbols might have been a creole that grew out of a pidgin spoken in the Nile Valley. That language was a language of traders, a pidgin which developed among several Afro-Asiatic languages for improved communication around the Nile ports. I agree with Almoaily that it is difficult to check the validity of this claim due to the controversies surrounding the interpretation of the Egyptian hieroglyphic symbols (see Bae, 2004).

Another limitation regarding what Almoaily mentioned in his thesis is the difficulty, if not impossibility, of collecting more data or checking accuracy if a given text is the only one available for that particular contact language, such as analysing old and scarce texts of extinct P/Cs, as in the case of the Maridi Arabic script (mentioned earlier in Section 3.1.1). Souag (2006) questions the accuracy of the script of this contact language, Maridi Arabic, which dates back to the eleventh century. He claims that this old script might have been altered or contains omissions by the copyists to make it easier to print, especially the features relevant to the phonology, and written scripts, particularly those using conventional spellings, 'often fail to accurately represent the phonological features of a language' (Almoaily, 2012). It is also difficult to verify the accuracy of the phonological features of contact languages that vanished or became extinct before they were either phonetically transcribed or audio-recorded. Sebba (1997: 244) writes about two methods of transcribing P/Cs:

> There are basically two approaches to orthographic (spelling) systems for pidgin and creole languages: phonemic and etymological. The phonemic approach involves treating the pidgin or creole as a language in its own right, without historical connections to any other, and producing a spelling system which has one, and only one, symbol per phoneme of that language... The etymological orthography treats the pidgin or creole as a dialect of the lexifier, and uses the conventional spelling of the lexifier for words which identifiably originate from the lexifier. Other words are spelt using the conventions of the lexifier, with modifications if necessary.

The advantages and disadvantages of these two approaches are discussed later in Section 6.6.5.

### 3.4.3. Pidgins and Creoles: A Typologically Discrete Class

This subsection discusses one of the most controversial issues in the literature of P/Cs, namely, the lack of consensus in defining and distinguishing between pidgins and creoles. The precise boundaries between the two are practically impossible to draw when discovering a new contact variety, and there is still no single accepted, unified definition which all researchers can agree on; '[c]reolists agree neither about the precise definition of the terms pidgin and creole, nor about the status of a number of languages that have been claimed to be pidgins and creoles' (Muysken and Smith, 1995: 3). Many other researchers, such as (Knapik, 2012), place import on defining the phenomenon of P/Cs, as they are an attempt to capture the nature of the terms, *pidgin* and *creole*, indicating different aspects of, inter alia, the languages' functions, the characteristic of their structures, and the role the languages play in the group. In this regard, Miller (2009) mentioned in her paper that the joining of various linguistic features and non-linguistic factors (historical context of emergence, type of contact, processes of acquisition, etc.) may help to define P/Cs. This was also discussed by Jourdan (1991: 190-91), who argues that 'no structural characteristics seem to exist that would help discriminate creoles from pidgins apart from the sociohistorical circumstances of their genesis'. It generally seems difficult to clearly distinguish pidgins from creoles based only on their linguistic features or to provide means of classifying contact languages into either a pidgin or a creole. Therefore, more precise definitions of P/Cs are required, which would help researchers classify a newly discovered contact variety as one or the other. Alternatively, the distinction between these two terms should be abandoned.

For three decades, a substantial amount of literature has been written on this topic, and the criteria used for the definition of P/Cs have evolved. The most important point is that linguistic criteria alone do not suffice to define P/Cs. As Winford (1997: 1) pointed out, 'the identification of pidgins and creoles is based on a variety of often conflicting criteria including function,

historical origins and development, formal characteristics, or a combination of these'. As mentioned earlier, the convergence of a number of linguistic features and non-linguistic factors may help to define P/Cs. However, in many instances, there is no clear-cut delimitation between pidgins and creoles; the precise boundaries between the two are practically impossible to draw. The discussion above opens doors for further research on the nature of P/C languages. As for the current contact language, GPA, it can be classified as a pidgincreole, according to Almoaily (ibid).

Most earlier and recent studies on this variety, such as those by Smart (1990), Naess (2008), Almoaily (2008, 2012), Alghamdi (2014), Abed (2017), and Alshammari (2018), have qualified GPA as a pidgin and not just 'learner Arabic' (e.g. Alfaifi and Atwell, 2013). This can be justified because none of the previous works claimed that GPA has native speakers as opposed to Arabic learners. GPA speakers are mainly adult Asian immigrants who only stay temporarily in the Gulf region and leave their families in their home countries, though it is even difficult to conceive that GPA would gain native speakers in the near future. Therefore, based on the previous discussion about the definition of P/Cs, first, creoles having native speakers would exclude the possibility of GPA ever becoming a creole, and second, pidgins are actually not native languages. These reasons are sufficient to qualify GPA as a pidgin, and not just 'learner Arabic'.

### 3.5. Gulf Pidgin Arabic: Previous Related Studies

This section presents some related studies that have investigated pidginised forms of the Arabic language in Gulf countries over the past 20 years, in chronological order. In this review, I focus on some of the linguistic features as regards the structural pattern of GPA, as well as the advantages and limitations of methodological choices in the previous works. Most of these studies thus far have been carried out in a descriptive way and/or are theoretical in nature. To date, only Almoaily's (2012) study has provided an in-depth quantitative analysis of substrate-

language-based variation or variation caused by the duration of stay in the Gulf area in the morphosyntax of GPA. My study is different from Almoaly's in that it investigates the gender and language variation in GPA resulting from morphosyntactic differences in the speakers' L1s and length of stay in the Gulf region, thereby filling a research gap.

Smart's (1990) study of 'pidginization in Gulf Arabic' was the pioneering report on GPA – the term 'Gulf Pidgin' was coined by Smart (1990) – describing the variety of pidginised forms of Arabic in the Gulf States. He described what he perceived as an emergent pidgin in the interaction between foreign expatriates and the local population in the Gulf countries. Smart's study focused on the phonological (vowels, consonants), morphological (bound forms, free forms), and syntactic components of pidginised GA. His investigation was based on printed materials (cartoon captions imitating the language of workers) selected from newspapers in the UAE, Qatar, and Oman between 1986 and 1987. Smart claims that at that time, these materials were the only published examples of the form that he was aware of. This suggests that there was no mention of any Arabic-based pidgin or creole outside the African continent prior to his article. The two other Arabic-based pidgins spoken outside Africa – Romanian Pidgin Arabic and Pidgin Madame – were only reported later by Avram (1993) and Bizri (2005), again suggesting that Smart's claim is true. In his study, he descriptively analysed a number of key linguistic elements in the phonological, morphological, syntactic, and lexical systems of GPA. Smart's study was based on written material, which is not commonly found in GPA or even non-Standard Arabic, and these materials were written in GPA by native speakers for the sake of humour.

This methodology raises the questions of whether his description of GPA is authentic or valid, and whether it reflects what this system really is. For instance, this source can provide insufficient and misleading data because 'it is hard to establish the phonology of Arabic or one of its lexified pidgins or creoles on the basis of written scripts' (Almoaily, 2012: 57). Some

letters in GA can be pronounced differently. For example, the letter ج can be pronounced as a /dʒ / in some dialects of GA and /j/ in other dialects. Another example is that the instances of the long vowel [u:] are found only in Smart's data: دوكان *du:kan* (shop) and كورة ku:rah (ball). Note that the short vowel /ʊ/ is dropped in the first word, and the long vowel [u:] is dropped in the second when they are spelled in Standard Arabic. Smart explains the rare occurrence of the long vowel [u:] by suggesting that the difference between the short /ʊ/ and long [u:] vowels is not phonemic in GPA. The two occurrences of the vowel /u: / could thus be a result of typographical errors in the written material: Typists could have misspelled the word for shop (i.e. دوكان *du:kan* instead of دكان dukan) and also misspelled the word for ball (i.e. كورة ku:rah instead of the proper spelling كرة kurah; Almoaily, 2012: 57). The long vowel [u:] in GPA is an allophone rather than a phoneme.

Smart (1990: 87) himself demonstrated awareness of this issue: 'Since, however, we are dealing here with a written source that is only partially representational, I have reduced the vowel system to that of S[tandard] A[rabic]'. Smart's article thus illustrates the problems faced by P/C research that is based on written material. Therefore, to avoid such problems, spontaneous spoken data is needed when describing a contact variety. It is generally concluded that although Smart's description was based on cartoon captions in newspapers which were written by native Arabic speakers who were imitating the language of FWs, he still considered this variety as a pidgin and not as FT:

> It seems to me fairly clear that, if material (albeit humorous) can be published in this language, it must be sufficiently developed and systematized to be familiar and intelligible to a reasonably large sector of the general Arabic-reading public. (Smart, 1990: 83)

Wiswall's (2002) work is the first on actual Gulf pidgin usage. In this work, he criticised Smart's failure to acknowledge that his corpus should be analysed as FT and challenged his view that some features are simply due to substratic influence, instead of the results of native Arabic stereotypes (superstrate speakers' influences). He examined the influence of native

speakers on the local language when using a pidgin with foreign pidgin users by conducting a comparative analysis of lexical borrowing in GPA. He specifically examined the use of the copula *fi*, the use of the command verb *sawwi* (do) instead of the GA inflected verb, and the possessive *mal* to replace the GA clitic possessive pronouns *-i* (my) and *-na* (our) in the speech of locals and expats speaking GPA.

Wiswall's work is perhaps the first and only one thus far which examines the variations in linguistic production of GPA use between local Arabic speakers and Indian expats. The participants in his study were divided into two groups: 10 native Arab speakers (i.e. Kuwait, the UAE, Qatar, and eastern SA) and nine Indian workers (IWs), all of whom were working in the UAE. All participants were asked to translate eight sentences from English to GPA, and Wiswall then compared the use of three linguistic features (the copula *fi*, compound verb structures, and the *mal* possessive) in the two groups. He noticed significant differences in this 'foreigner talk' used by the native speakers of GA (first Arabic language speakers) versus IWs (L2 of immigrant speakers) in the Gulf countries. He found that the three morphosyntactic features were used much more by the local native GA speakers than by the IWs. Almoaily (2012) hypothesised that this might be due to *hypercorrect speech*, as indicated by Labov (1966), where the native GA informants exaggerated their usage of GPA features in trying too hard to reach the typical GPA speech. Moreover, Wiswall stated that the choice of borrowed words was much more frequent in the translation of IWs, for example *taksi driver*, *seem seem* (the same; English lexical borrowing) and *sawa sawa* (together; Persian/Urdu). This may support my hypothesis regarding a substratic influence on GPA speakers.

Furthermore, a descriptive analysis was conducted by Naess (2008), who descriptively analysed the phonology and some grammatical features (possession, negation, the copula *fi*, and the VP of GPA) to define the GPA structures and compare them to GA. The aim of her study was to determine whether GPA constitutes a 'true pidgin', or whether it should merely

be considered as a means of communication among Asian immigrant co-workers or with local citizens. Naess' report was based on interviews with GPA speakers conducted in the Omani border town of Buraimi and in the Emirati city Al-Ein. She divided her sample into two different social classes (eight mid-class and eight low-class participants) and according to sex (eight males and eight females). The variability in linguistic and social background in Naess' informants is one of the great interests of this study. One noteworthy aspect is that Naess is not a native speaker of Arabic, let alone GA, which might have influenced the nature of her data. Naess (2008: 10) herself was also aware of this problem: 'As a non-native speaker of Levantine Arabic, initially unfamiliar with the Gulf Arabic dialect, my speech might have influenced my consultants. For a majority of these, though not for all, English would have been the natural mode of communication with an English-speaking foreigner'. This hindrance could led her participants to code-switch to English unconsciously. As Buchstaller (1999: 14) states, 'Employing an interviewer from within the speech community is thus the most efficient means to gather data the least affected by code-switching'. Therefore, in Naess' case, it would have been better if she had utilised local people or expatriates to lead the interviews.

Almoaily (2008) submitted his MA dissertation, *A Data-based Description of Urdu Pidgin Arabic* pidgin and creole: *a socio-historical and a data-based linguistic*, where he described the phonology and morphosyntax of GPA. He conducted his study by interviewing six Asian FWs from Pakistan, Bangladesh, and India who were working in central SA, in Alkharj City. The participants in Almoaily's study were all male and came from lower-class occupations, unlike Naess' sample where the informants were equal in terms of both gender (equal number of males and females) and class status (equal number of low- and middle-class participants). A description based on interviews between a local speaker of GPA and expats. The interviewees who share similar backgrounds certainly offers an optimal testing ground to determine the effect of ethnicity on the patterning of GPA syntax. In his study, Almoaily also attempted to

determine the sociolinguistic status of GPA (attitudes of locals towards GPA, frequency of use, ease of communication in GPA, etc.) in SA by distributing a questionnaire to 77 Saudi students in the UK, where the researcher was conducting his study. Almoaily's study is the first quantitative variationist analysis of GPA that uses data collected via interviews between native speakers who speak GA and foreign Asian workers. The results revealed that ethnicity (the nationality of the GPA speaker) had little impact on the speech production of the GPA speakers. Indeed, the small sample size (two speakers from Bangladesh and two from Pakistan who produced only 4,000 words each) in Almoaily's (2008: 61) work brings into question whether the selected sample linguistically represents the whole speech community. Almoaily himself was aware of this problem:

> the corpus on which this description of UPA is based is relatively small (1 hour, 19 minutes), which is much smaller than the corpora used in other pidgins and creole studies…due to the small number of participants and non-negligible inter-ethnic variability, more individuals need to be sampled in order to make definite statements.

To this end, Almoaily (2012) increased the number of his informants to 16 and built up a corpus of 12,000 words (discussed in detail at the end of this section). Regarding the questionnaire, the responses revealed that almost half of the participants did not mind using GPA as a means of communication with Asian foreigner workers. However, the majority held negative attitudes towards GPA. Hence, these findings support Smart's (1990) claim that GPA is a low-prestige variety and that some of its speakers endure the absence of admiration by natives.

Bakir (2010) began his paper with a clear description of the pidgin status of GPA. This was followed by a descriptive account of the verbal (morphological characteristics of the verbs) system of GPA based on field work data collected in Doha, Qatar. He made a comparative analysis of the properties of GPA with the general characteristics of some Arabic-lexified P/Cs. He conducted his study by interviewing 10 Asian FWs from different linguistic backgrounds: three Bengali, two Malayalam, one Tagalog, two Sinhala, one Tamil, and one Hindi. There was an equal distribution of gender, just like Naess' (2008) study above. The time that each

informant in Bakir's sample spent in the Gulf varied; some of them had lived and worked in the Gulf region for 20 years, while others had stayed for just 2 years. However, Bakir (2010: 204) did not investigate the influence of the importance of the following social factors on language production: '[t]he transience of the working conditions, the mobility of the working force, and the social and racial differences'. He made a general statement (without any statistical support) that '[t]he settlement of some of these expatriates in the region motivates them to move up the linguistic ladder in the sense of moving closer to Gulf spoken Arabic' (Bakir, 2010: 204). His statement was, however, refuted by Almoaily (2012) when he statistically investigated the relevance of length of stay on linguistic production. His results did not support Bakir's claim, and this is discussed in more detail at the end of this section. In my thesis, I conduct a further investigation into the effect of duration of stay in the Gulf on female GPA speakers' choice between GPA linguistic variants to determine whether long-term female speakers make a significant shift to GA after spending 10 or more years in the Gulf compared with newly settled female GPA speakers.

The sixth piece of research on GPA is Alshammari's (2010) MA dissertation, which provides a detailed morphosyntactic (inflections, determiners, quantifiers, word order, negation, copulas, and verb form) description of GPA based on interviews conducted with nine male expats who had spent between five and 15 years in SA, specifically in two cities of the north of SA, namely, Hail and Sakaka, and who came from three linguistic backgrounds: four Bangladesh (Bengali), two Indian (Malayalam), and three Afghani (Pashtu). The author investigated whether the characteristics of GPA are in harmony with the universal features of pidgins. Alshammari is a native GA speaker who could participate positively in collecting spontaneous data in this study. Moreover, Alshammari offers phonetic symbols of Arabic consonants and their descriptions in an appendix, which are helpful for a researcher who is interested in documenting and investigating GPA. Alshammari's findings revealed that GPA

supports the universal features of pidgin languages in most of the selected grammatical features except the copula and numeral system. Thus, on the one hand, his results contradict Ferguson's (1971) claim that dropping the copula is one of the universal characteristics of pidgins, and on the other hand, they support Smart (1990) insofar as the morphosyntactic system of GPA agrees with the universal features of Indo-European lexified pidgins in features such as analytic morphology and indication of tense via adverbs. Indeed, despite all the informants coming from different linguistic backgrounds, Alshammari (in alignment with other aforementioned researchers) did not investigate possible variation patterns across the informants.

My study is closely linked to Almoaily's (2012) PhD thesis as the gender variation is investigated by conducting a comparative quantitative analysis of the morphosyntax of my female GPA speakers with that of Almoaily's male GPA speakers. To my knowledge, Almoaily's (2012) PhD contains the first quantitative variationist analyses of GPA. He begins his thesis by discussing some common theories on the emergence of P/Cs, and he concludes with a suggestion that his data supports Mufwene's (1993) complementary theory of genesis, which claims that universal as well as substratal factors can contribute to the emergence of contact languages. Mufwene (2006: 320-21) writes,

> Few creolists subscribe nowadays to one exclusive genetic account, as evidenced by the contributions to Mufwene (1993). The 'complementary hypothesis' (Baker and Corne 1986, Hancock 1986, and Mufwene 2001) seems to be an adequate alternative, provided we can articulate the ecological conditions under which the competing influences (between the substrate and superstrate languages, and within each group) may converge or prevail upon each other.

In his PhD thesis, Almoaily (2012) provides a quantitative study analysing the influence of speakers' L1s and the number of years of their residency in the Gulf as potential factors conditioning language variation in GPA. His data was based on interviews with 16 informants from three linguistic backgrounds (Malayalam, Bengali, and Punjabi), and interviews were conducted in two cities, namely, Riyadh and Alkharj, SA. The informants, like Alshammari's (2010) participants, were all male, and all held a low status. Indeed, it is difficult for a male

researcher to conduct interviews with females face-to-face in SA because Saudi society is gender-segregated. To answer the question of whether GPA speakers actually shift towards GA after spending more than 10 years in the Gulf, the researcher purposefully chose half of his informants who had spent five or less years in the Gulf, while the other half had spent 10 or more years in the area.

The analysis in Almoaily's (2012) study was based on 10 morphosyntactic phenomena: free or bound object or possessive pronoun, presence or absence of the Arabic definiteness marker, presence or absence of Arabic conjunction markers, presence or absence of the GPA copula, and presence or absence of agreement in the VP and the NP. His results revealed that there is little impact on the informants' variable choices of the selected morphosyntactic features based on their L1s and their length of stay in the Gulf. Among all 10 morphosyntactic features above, only conjunction markers showed a significant relation between the informants' L1 and their choice of GPA variables. Almoaily's (2012) contribution to the literature is less-described non-Indo-European P/Cs. He concludes that there are some fallacies in the literature of P/C languages, resulting from an Indo-European-centric view and a lack of consensus in defining different forms of language contact. For example, when discovering a new contact variety such as GPA, classifying it into one particular category (jargon, pidgin, or creole) is difficult, since the literature still lacks consensus in defining and distinguishing between them. Almoaily thus redefines Bakker's (2008) definition of the term pidgincreole[7] as contact languages carrying features typical of both pidgins and creoles.

Albakrawi (2012) conducted a qualitative study to investigate the linguistic effect of Asian foreign expatriates on the Saudi Arabic variety of Arabic Pidgin in SA. Albakrawi's study focused on workers from Asian origins rather than other workers. His data collection was based

[7] Bakker (2008) defines *pidgincreole* as a restructured language which is the primary language of a speech community, or which has become the native language for some of its speakers.

on tape-recorded and planned interviews with foreign Asian workers who spoke Arabic Pidgin and worked in Tabuk city, SA. Reporting the existence of GPA in Tabuk, a city located in the far north-west of SA, is further evidence of the wide geographical distribution of this contact language. Albakrawi provided a clear account of GPA by examining several morphosyntactic features (lack of inflections, determiner, quantifier, word order, negation copula, and verb) and comparing the similarities and dissimilarities between GPA and some of the universal characteristics of pidgin. The findings revealed that GPA is linguistically influenced by these features and that they are compatible with pidgin universals.

Salem (2013) qualitatively examined the linguistic features of GPA in Kuwait, which he literally refers to as 'Pidgin Arabic in Kuwait'. He interviewed 40 Asian informants who had been living and working in Kuwait and had spent between six and 18 years in the Gulf region. His study aimed to provide a descriptive account of the phonology, syntax, and lexicon of this pidgin. However, similar to other descriptive studies, Salem did not investigate any potential linguistic variation patterns across the informants from different linguistic backgrounds.

Another study presented by Potsdam and Alenazi (2014) proposed a unified analysis of the syntactic function of the morpheme *fi* in GPA based on previous descriptions (Smart, 1990; Næss, 2008; Bakir 2010). However, the authors extracted more authentic examples of GPA from recorded conversations with foreigners who lived and worked in Riyadh, SA, while GA examples were elicited from native GA speakers. They started their paper with a full discussion of the form *fii(h)* in GA and in GPA. This was followed by a unified structural analysis of GPA *fi* as a verb that introduces a non-verbal predication, and they argued against suggestions that it is an expletive subject, a possessive verb 'have', or a TMA marker, as proposed by Smart (1990), Næss (2008), and Bakir (2010). They concluded that GPA speakers use GPA *fi* only as a preposition or a predicational copula verb and that the other suggested uses of *fi* can and should be reduced to the predicational copula use. Moreover, the morpheme *fi* is a verbal head

that selects for a non-verbal predication. This analysis relied on Freeze's (1992) unification of existential, possessive, and locative clauses, which also supported their analysis. In addition, their results oppose Ferguson's (1971) claim that simple contact languages, such as pidgins, uniformly lack copulas.

Alghamdi (2014) also provided a descriptive and quantitative account of GPA morphology, syntax, and lexicon. Alghamdi's study was preceded by a theoretical discussion about the emergence of pidgin languages, in which he emphasised the major role of FT in the formation of GPA. His data collection was based on interviews conducted with 10 participants: seven males who were Bengali, Urdu, or Indi and two females – one Sinhala and one Indonesian – all of whom were GPA foreign speakers working in Jeddah, located in the western province of SA. A description based on interviews with female and male informants who come from various linguistic and social backgrounds is one of the great advantages of this study. However, in some cases, for example where female maids live with a local family that mostly uses GA when communicating with them, GA can be the TL more than newly arrived GPA speakers who are receiving GPA input with a limited amount of GA input, which appears to be the case for the vast majority of GPA speakers. The author's finding revealed that morphological inflection is rarely utilised in GPA nouns and pronouns, particularly in distinguishing grammatical cases, numbers, and genders. He provides the following two examples when GPA speakers did not add affixes to distinguish singular nouns from plural nouns (1) or masculine nouns from feminine nouns (2) in GPA:

(1) ana rūh Makkah tanēn marah-Ø (Alghamdi, 2014)
1SG went Makkah two time
'I went to Makah twice'

(2) hāda siyāra hagg- Ø baba mafi hagg anta (Alghamdi, 2014)

this car-F POSS father NEG POSS 2SG

'This car is my boss' car; it is not yours'

Moreover, he stated that there is no dependable verbal declension for verbal tenses, nor is there declension for person and number. Instead, GPA speakers use masculine imperative and imperfect forms of verbs to refer to all verbal forms, and the only negation marker, *mafi*, is used in GPA. The syntactic structures of GPA are simple and lack complex sentences, and it has been shown that an SVO sentence structure is used over other sentence structures and patterns. Through examples, Alghamdi explains that GPA speakers use GA adverbials such as *ams* ('yesterday'), *awal* ('before'), *bādēn* ('then'), *alhēn* ('now'), and *bukrah* ('tomorrow') to indicate tense, aspect, and modality (TAM), as shown below:

(3) huwa ?amis mafi yakul gadā (Alghamdi, 2014)

3SG yesterday NGT 2SG.M.IPF-eat lunch

'He did not eat his lunch yesterday'

(4) lazm ana bukrah ruuh bank jib fuluus (Alghamdi, 2014)

should 1SG tomorrow IMP-go bank IMP-get money

'Tomorrow, I should go to the bank to get some money'

GPA lexicon is the last linguistic feature that Alghamadi examined. He stated that the GPA lexicon consists of small vocabularies that are repeatedly used with many different meanings and functions, as illustrated in the below example:

(5) huwa mafi kalam ana fi kalam (Alghamdi, 2014)

3SG NEG speech 1SG TAM speech

'He [did, do, will not] speak [when] I talked to him'

A number of studies on GPA have been carried out by Avram (2013, 2014, 2015, 2017, and 2018). His earlier studies provide an analysis of the main structural features of the phonology, morphology, syntax, and vocabulary of GPA attested across speakers and regions. The linguistic description is preceded by a discussion of the sociolinguistic situation in the Gulf. Avram's investigation is based on two GPA sources, namely, published and unpublished ones, and unlike Smart (1990) and Al-Azraqi (2010), his corpus does not include representations of GPA in, for example, the media, cartoons, TV series, and films. The first source contains transcripts of interviews, answers to questionnaires, and translations of test sentences (Wiswall, 2002; Almoaily, 2008, 2012; Næss, 2008; Al-Azraqi, 2010; Bakir, 2010; Almoaily, 2012; Salem, 2013). The second one consists of online sources: internet discussion lists (involving participants with different L1s, not including Arabic), songs, and poems. He concludes that GPA is not only used between native Arabic speakers and Asian workers but also 'among immigrant workers of various linguistic backgrounds, i.e. when no native speakers of Arabic are involved'. The general conclusion of Avram's studies, drawn from an overview of the key linguistic structure of the language as well as the substantial variability between and within speakers, is that GPA is currently a pre-pidgin/minimal pidgin with significant fluctuations (as in Mühlhäusler, 1986). His conclusion is in parallel with Næss (2008) that GPA is 'an incipient pidgin variety' that is yet to achieve conventionalisation. However, this is unlike Almoaily (2012), who argued that GPA 'has a mix of pidgin and creole features' and has already achieved stabilisation.

Avram (2017) examined the impact of FT as one of the potential factors conditioning language variation in GPA. The analysis is based on GPA data from general descriptions of GPA as well

as from analyses of GPA attested across specific regions and some online sources. The comparison is based on the following morphosyntactic features: 'two' + singular noun, dropped definite article, independent pronouns, the invariant masculine singular form of adjectives and demonstrative, the invariant forms of verbs, light verb constructions with 'make', time adverbials to express tense and aspect, *fi* + adjective, *fi* + verb, variable word order, lexical items from English and from other languages, lexical polysemy, and circumlocutions. The linguistic description is preceded by a discussion of the superdiversity of the linguistic situation in the Arabian Gulf, and the analysis of the data revealed that these structural features are also evidence in Arabic FT. Avram thus suggests that 'Arabic Foreigner Talk might be one of the sources of Gulf Pidgin Arabic structures'.

Avram's (2018) latest work compares the morphosyntactic and lexical features of the Arabic FT register to those of four Arabic-lexifier pidgins, namely, Pidgin Madame, Jordanian Pidgin Arabic, Romanian Pidgin Arabic, and GPA. In this research, he found a significant number of features which Arabic FT shares with some, if not all, of these Arabic-lexifier pidgins. His study suggests that a relationship exists between Arabic FT and Pidgin Arabic, which reinforce each other in the occurrence of these features. In addition, Avram himself is not a speaker of GA, which could have affected the quality of the data analysis.

Furthermore, Al-Zubeiry (2015) provides a descriptive account of the three major morphosyntactic structures of Saudi Pidgin Arabic (SPA) – sentence word order, VP form, and NP form – based on interviews with 30 Asian foreign expatriates who lived and worked in Al-baha, the southern region of SA, and came from various linguistic backgrounds – Urdu, Hindi, Bengali, Tagalog, Indonesian, and Malayalam. Al-zubari (2015: 49) commented on the length of exposure to Saudi Arabic, saying that '[i]t is worthy to mention that most of the informants who were interviewed have been working in the country for a period ranged between 4–10 years', but he did not make any attempt to tease apart the importance of social and

developmental factors on language production. He concluded that the selected morphosyntactic features of SPA are compatible with the universal characteristics of SA, such as reduction and simplification.

At the end of his study, Al-zubari draws some implications of SPA which could be disproven or supported if a variationist analysis could be carried out on his data. For example, 'substrate languages play an essential role in the formation of the SPA'. This claim could have been strengthened if Al-zubari had attempted to demonstrate a significant effect of SPA speakers' L1 on their choice between SPA linguistic variants. Similarly, despite the fact that GPA has been frequently highlighted in a large number of Avram's works, no attempt has been made to provide any variationist quantitative analysis. In addition, as mentioned earlier, Avram himself is not a speaker of GA, which could have affected the quality of the data analysis.

Al-Ageel (2015) conducted a linguistic and sociolinguistic analysis of GPA, examining one of the aspects of politeness behaviour in making requests among two generations of Saudi female speakers making requests to Asian females working in the cafeteria of a governmental institute in the Saudi capital, Riyadh. The analysis of SPA in Al-Aqeel's study focuses on how the use of pidgin was shaped by social factors such as power, status, social distance, and the age of the individual making the request. The report of the linguistic and grammatical systems of the elements (the NP, the VP, affixation, the copula *fii*, pronouns and demonstrative pronouns, and lexical features) is based on interviews with 16 SPA Asian participants, half aged between 20 and 39 years and the other half aged between 40 and 60 years, who were staff and trainees at a governmental institute. Furthermore, unlike other studies on GPA (such as Almoaily, 2012; Alshammari, 2010; Bakir, 2010; Al-Zubeiry, 2015), where informants were all male, the informants in Al-Aqeel's study were all females among whom SPA might be influenced by the social variables due to the conservative social life in SA (e.g. social gap between the local population in SA and foreign Asian workers). Although her social sample comprised two

different generations of Saudi women, she did not provide any variationist quantitative analysis to determine whether age had an influence when using SPA to interact with non-native Arabic speakers. On the one hand, Al-Aqeel (2015) found that the use of SPA, to a great extent, is limited to the most straightforward verb forms and NPs utilised in the local variety (i.e. Najdi Arabic). On the other hand, she suggests that younger generations tend to use more English expressions in SPA as a result of the growing impact of English on these generations. Another observation made by Al-Aqeel regarding Saudi informants' behaviour is that the use of SPA reflects the wide social distance between Asian migrant workers and their employers in SA. This finding might support my results, as the social dimensions between locals and FWs may contribute to shaping GPA.

My prior study (Albaqawi, 2016) provides the first phonological comparative study across Asian migrant Arabic pidgins in the Gulf region. The linguistic description is preceded by a discussion of the geographic, economic, and sociolinguistic situation in the Gulf region. My investigation is based on a data collection employing different methods and for different purposes, and it consists of three types of linguistic material: (i) FT material collected from the media (e.g. newspapers, as in Smart, 1990; TV scripts, as in Al-Azraqi, 2010; and online sources such as poems and songs, as in Avram, 2012), (ii) data elicited for descriptive purposes (e.g. elicited translation data, as in Wiswall, 2002, and other interviews' elicited data, as in Al-Azraqi, 2010), (iii) spontaneous interactive data which are either published as a whole (e.g. interviews by Almoaily, 2008 and Salem, 2013) or cited only in the illustration of a linguistic analysis – i.e. taken out of their interactive context (e.g. Næss, 2008 and Bakir, 2010). The analysis of the available GPA-related linguistic data revealed that its segmental phonology seems to comply with universal features of P/Cs, either reduced or simplified (i.e. some consonants are either lost or have undergone a shift from the lexifier GA).

My findings suggest that both factors (linguistic and sociolinguistic) are responsible for unity and diversity within spoken varieties, such as GPA. These findings support Bizri's (2014: 385) claim that in the case of Asian Migrant Arabic Pidgins, 'mobility across the region is the major factor for homogenising both native Arabic-speakers' foreigner talk and migrants' pidgin Arabic'. Therefore, GPA can be assumed to have developed due to regular contact between native Arabic speakers/employers in the Gulf area and Asian migrant workers. Native speakers consequently tend to interact with migrant workers in the pidgin rather than the superstrate language (Almoaily, 2012). Here, native Arabic speakers have already adopted pidginised forms of Arabic in their FT – whatever pidginised variety migrants develop – so only what is recognised in native Arabic speakers' FT will be transmitted to newcomers. By doing so, they validate the pidginised system and totally impede further access to the superstrate.

In spite of their high degree of unity in both linguistic structure and social context, GPA varieties do display aspects of diversity. This diversity is attributed to differences between the individual linguistic and non-linguistic backgrounds in which migrants navigate. On the one hand, each individual GA country has a unique set of foreign substrate languages. For example, SA has a highly complex multilingual setting with a variable number of substrate languages, each represented by a sizeable speech community. On the other hand, within individual speech communities, there are distinct groups defined by typical features and a peculiar context for language acquisition, determining what they have in their linguistic repertoire, which in turn determines the structure of the incipient variety. These groups are determined, for example, by participants' age and gender, the range of the Arabic norm to which the migrants are exposed, and the personal motivation of the migrants throughout the Gulf countries (Bizri, 2014). Notably, however, as in the other descriptive studies, my study does not investigate potential variation patterns across the informants coming from different linguistic backgrounds.

Abed (2017) recently wrote a PhD thesis on GPA, in which he attempts to explore GPA from a sociolinguistic rather than a linguistic stand point. He provides a comprehensive picture of the social circumstances which have resulted in the development of GPA, and he evaluates the social attitude towards this language depending on its speakers' points of view. The description of GPA is based on the analysis of sociolinguistic data by using two data collection instruments, namely, interviews and a survey questionnaire. The first dataset included the speech of 38 unskilled/semi-skilled Asian workers who come from seven linguistic backgrounds – Urdu, Bengali, Tagalog, Malayalam, Tamil, Malay, and Maranao. The informants interviewed in Abed's study were from both genders and had spent between five and 30 years in the Gulf region. The interviews were conducted in Jeddah, the largest city in the western region of SA located in Makkah, which is one of the two holiest cities in Islam, making Jeddah home to a diverse and multi-ethnic population with varied social classes. The data collected from this group was transcribed word for word and then analysed in detail. The data of the second group, the Saudi participants, was collected through a number of random participants (all Saudi citizens and residents of SA for no less than eight years) who answered a public online questionnaire. Abed's data from this group was analysed statistically. The fact that Abed is a native speaker of GA could have aided in collecting spontaneous data. The findings from Abed's interviews revealed that the social atmosphere for Asian workers in relation to the local population in SA plays a major role in the emergence of GPA. By contrast, the data collected from the questionnaire suggested that the social attitude towards GPA is that the language is generally associated with the low social status of unskilled/semi-skilled Asian workers; this attitude was held by the local population. Abed's study affords readers better awareness and knowledge of GPA from a sociolinguistic standpoint, but does not link social factors and the choice of linguistic variables, which can make a valuable contribution to the field of (socio)-linguistic variation and change in contact languages.

Another recent analysis study on GPA is Alshammari's (2018) PhD thesis, which provides a descriptive and quantitative analysis of the verbs and pronouns of GPA based on frequency, detectability, markedness, substratal influence, accommodation, and FT. The linguistic and sociolinguistic analysis is based on two sources of data: sociolinguistic interviews and a web-based Arabic corpus. The first dataset was composed of sociolinguistic interviews with 30 male participants: Seven were native Saudis, and 23 were FWs who lived and worked in two cities, namely, Ḥā'il and Sakaka (in the north of SA), and came from three linguistic backgrounds – Pashtu, Bengali, and Malayalam. The second dataset used in the study was an existing Arabic corpus, *arTenTen*, available via *Sketch Engine*. The approximate length of the corpus of sociolinguistic interviews was five hours (equating to 301 minutes), yielding roughly 39,500 words, which comprised roughly 130 single-spaced pages of transcription, while the total number of words in the *Corpus of the Arabic Web*[8] is 8,322,097,229 words and represents a mix of both dialectal Arabic and MSA. The comparative analysis of these two corpora on the basis of competing variants of verb form and pronoun form revealed that the differences between GPA speakers and Saudis are statistically significant. However, both groups tend to not use native-like verb and pronoun forms and instead show a preference for selecting agreed-upon solutions, pointing to a degree of conventionalisation within the speech community and a preference for select morphologically free pronoun forms over bound pronoun forms. Furthermore, length of stay in SA and substratal effects on GPA speakers' choices among the selected linguistic features seem to have little influence on form selection. Examining both the social factors and linguistic factors in speech communities where the factors are interwoven to create a new linguistic system allows readers to better identify the forces that control language learning, such as GPA in a naturalistic setting.

[8] The Sketch Engine interface is available at https://www.sketchengine.co.uk/arTenTen-corpus/. It was crawled by SpiderLing in January 2012, then encoded in UTF-8, cleaned, and duplicated. Tagged by Stanford Arabic Parser in August 2015, it provides easy access to tools for visualizing different aspects of word frequency.

To my knowledge, Alshammari's study (2018) is the first to present systematic attempts to quantitatively combine the cognitive and social factors that contribute to the emergence of GPA. However, the data collection methodology for the *web-based Arabic corpus* implemented by Alshammari raises some concerns. For instance, although this corpus includes a mix of both MSA and dialectal Arabic from various native Arabic-speaking countries, the written Arabic scripts do not show diacritics (e.g. [ ◌́ ], [ ◌ِ ], [ ◌ُ ]), which correspond to short vowels in English, such as /a/, /i/, and /u/ in transcription, respectively. Therefore, it is impossible to differentiate orthographically between some Arabic forms and their derivatives, unless diacritics are employed. For example, the verb *nasʿaħa* (third Masc.Sg.Pas.; he advised) and its verbal noun form *nusʿħ* ('advice') are both written in the same orthographic. Similarly, the third Masc.Sg.Pas. *kataba* ('he wrote') and its plural noun form *kutub* ('books') cannot be differentiated.

Furthermore, Arabic script differentiates a geminate from a singleton by means of a gemination diacritic [ ◌ّ ]; however, Alshammari's current corpus lacks the gemination diacritic, which complicates the issue. For example, *fikir* ('thought'), *fakkara* ('he thought'), and *fakkir* ('you think') are all read the same, unless the meaning is inferred contextually. In fact, Alshammari (2018: 124) was aware of this impediment: 'As a result, the ability to look at the frequency of different verb forms was limited'. In addition, Alshammari's findings were based on certain linguistic features (only the use of pronouns and verbs) in a small sample size (only 23 FWs and seven native Saudis) and for a small number of substrate languages (only three linguistic backgrounds, namely, Pashtu, Bengali, and Malayalam).

Finally, my recent co-authored paper (Albaqawi and Oakes, 2019) explores the methodology used to compile and analyse a transcribed spoken GPA corpus, with a specific focus on the influence of length of stay in the Gulf on foreign expat female speakers of GPA. The study provides the first quantitative analysis of language variation in female GPA speakers based on

five morphosyntactic features that are related to the length of stay in the Gulf: definiteness and indefiniteness, coordination, copular verbs, pronouns, and agreement in the VP and in the NP and ADJP.

The linguistic analysis is based on digital recorders and planned interviews, which were collected for this PhD thesis. All the interviews were conducted in SA, Riyadh to test one of the potential factors, namely, the influence of the length of stay on female GPA speakers' language variation. The database consists of interviews with 72 GPA-speaking female informants from six linguistic backgrounds (Malayalam, Punjabi, Bengali, Tagalog, Sunda, and Sinhala), as these substrate languages have the largest number of speakers in SA, based on the results of the Population Census from De Bel-Air (2018). Half of the data was produced by informants who had spent five years or less in the Gulf, while the other half had spent 10 years or more in the area at the time of interview. This paper sought to investigate whether long-term residents have actually shifted towards GA after spending over 10 years in the region. The structural patterns of GA that were collected from the newcomers of each language group were compared with those of long-term residents (e.g. newly settled Tagalog speakers vs. Tagalog speakers who had spent more than a decade in the Gulf).

Evidence from this corpus' data indicates that length of stay in the Gulf seems to have little effect on informants' choice between GPA linguistic variants. Chapter 8 investigates the relevance of length of stay on linguistic production, supporting Bakir's claim that '[t]he settlement of some of these expatriates in the region motivates them to move up the linguistic ladder in the sense of moving closer to Gulf spoken Arabic' (Bakir, 2010: 204).

Furthermore, based on this study's findings, some factors could have had an effect on the informants' choice between the selected features' variants. This could potentially be because most of the informants were female maids living with a local family that mostly used GA when communicating with them, which could have played a major role in the process of acquiring a

language. This in turn led them to rapidly learn the language of the host community and effortlessly adopt the system of GA (the TL). Another effect on the informants' choice between the selected feature variants is that it may depend more on the amount of GA input that GPA speakers receive during their stay in the Gulf (rather than the language of origin), different learning abilities to learn a language, and motivation. However, this thesis seeks to test other factors, such as the impact of female speakers' L1 and the number of years of residency in their location in the Gulf as potential factors conditioning gender and language variation in GPA. In the next section, I discuss the issues of corpus-based studies with a specific focus on contact languages.

### 3.6. Corpus-Based Approaches to the Study of Contact Languages

Corpora have been put to many different uses in fields such as natural language processing, critical discourse analysis and applied linguistics (e.g. forensic linguistics), and language pedagogy, and they have been used extensively in nearly all branches of linguistics, including lexicographic and lexical studies, grammatical studies, language variation studies, contrastive and translation studies, diachronic studies, semantics, pragmatics, stylistics, sociolinguistics, and discourse analysis,. Despite the fact that corpora still occasionally receive hostile criticism, they have gained widespread popularity over time (e.g. Widdowson, 1990, 2000, and 2003).

In this section I focus specifically on the principles of corpus linguistics as a research methodology. I review some debates over the use of corpus data in linguistic analysis and discuss some of the pros and cons of using corpus data. I also explore the implications of this specific approach to the study of language for studies of P/Cs.

#### 3.6.1. A Corpus Defined in Corpus Linguistic Terms

The word 'corpus' comes from the Latin 'body', and the plural is corpora. Examples of corpora in the literature are collections of short stories stored in electronic form and even the whole World Wide Web. However, there is a lack of consensus on the necessary and sufficient

conditions for a text to be a corpus. Therefore, to discuss the fundamental principles of corpus linguistics, it is essential to start establishing certain limits around what can and cannot be considered a 'corpus-based' study of language.

A number of different definitions cover the term corpus and emphasise different aspects of this type of resource. McEnery and Wilson (1996: 87) define a corpus as 'a body of text which is carefully sampled to be maximally representative of a language or language variety'. In their definition, they emphasise representativeness, which might be difficult to evaluate depending on what the corpus is used for. Bowker and Pearson (2002: 9) offer another definition of a corpus: 'a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria'. By using more selection criteria, Bowker and Pearson allow for greater flexibility than McEnery and Wilson, even when the corpus is still intended to be 'used as a representative sample of a particular language or subset of that language' (Bowker and Pearson, 2002: 9), because these selection criteria more accurately reflect the fact that corpus representativeness is always dependent on the purpose of using the corpus and on the specific linguistic features under study. Representativeness cannot be ensured at some particular design stage. For example, a corpus may accurately represent the distribution of a common feature, such as pronouns in one subset of a certain language, but may not accurately represent a rarer feature, such as the use of reported speech, in the same subset. Corpora are generally intended to be long-term resources and to be used for a variety of studies, and representativeness hence cannot be ensured at the design stage (Saldanha, 2009).

Apart from selection criteria, Bowker and Pearson's definition contains three additional aspects to differentiate a corpus from other collections of texts: the authenticity of the data, the means of storage, and the size. Authentic data is data that occurs naturally and has not been elicited especially for the purpose of linguistic analysis. The means of storage is the main aspect which differentiates current corpus linguistics from a longer-established tradition of manually

analysing collections of texts for the purposes of extracting data (Saldanha, 2009). Regarding size, Bowker and Pearson suggest that a corpus contains 'a greater number of texts than you would be able to easily collect and read in printed form' (2002: 10).

Leech's (1992: 106) definition of a corpus is more flexible: 'a helluva lot of text, stored on a computer'. In his definition, he did not offer criteria to distinguish a corpus from other collections of texts, as it seems he had no need for such a distinction; however, his emphasis was on size and medium. Likewise, Kilgarriff and Grefenstette's (2003: 334) definition of a corpus is also flexible: 'A corpus is a collection of texts when considered as an object of language or literary study'. They did not add any constraints regarding what can be considered a corpus, because they felt that the focus on linguistics study can be underestimated in corpus linguistics.

A more stable definition of a corpus is offered by Sinclair (2004: 19): '[a] corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research'. He justifies the use of the word 'pieces' instead of 'text', since sampling methods were still used in building some corpora rather than gathering complete texts or transcripts of complete speech events. He focuses on selecting the same external criteria as in Bowker and Pearson's definition. Moreover, the primary purpose of corpora is stressed so that they are not confused with other collections of language. In conclusion, what makes a 'good corpus' is sometimes size (which must be large enough, and Kilgarriff and Grefenstette even talk about the whole World Wide Web as a corpus), which can outweigh the benefits of careful selection criteria (Saldanha, 2009).

### 3.6.2. The Object of Study in Corpus Linguistics and Language Studies

In this section, I briefly define corpus linguistics and sociolinguistics, and I explain how they relate to each other. This is followed by some of the advantages and disadvantages of using

corpus evidence in sociolinguistic studies prior to coming to an informed conclusion that most disadvantages can be solved as technology advances.

Based on the definitions of a corpus in Section 3.6.1, corpus linguistics is an approach that is gaining increasing prominence in the field. It has become one of the most widespread methods of linguistic investigation in recent years. As Joseph (2004: 382) states, 'we seem to be witnessing as well a shift in the way some linguists find and utilize data – many papers now use corpora as their primary data, and many use internet data'. Corpus linguistics can be described as the study of language based on text corpora, where corpora are important sources of systematic collections of naturally occurring data (of both written and spoken language) for a number of areas within the wide scope of linguistics. According to Müller and Waibel (2000), corpus linguistics is 'the study of language by means of naturally occurring language samples; analyses are usually carried out with specialised software programmes on a computer. Corpus linguistics is thus a method to obtain and analyse data'. They also define it as 'a method to obtain and analyse data quantitatively and qualitatively rather than a theory of language or even a separate branch of linguistics on a par with e.g. sociolinguistics or applied linguistics'.

Corpus linguistics has been demonstrated to have the potential to yield fundamental and often surprising new insights about language (Nesselhauf, 2011). The corpus approach is employed not only in describing language features but also in testing hypotheses formulated in many different linguistic frameworks. In short, corpus linguistics serves to answer two fundamental research questions: 1) What particular patterns are associated with lexical or grammatical features? and 2) How do these patterns differ within varieties and registers? (Bennett, 2010). Corpora were originally used in pedagogy and lexicology, whereas nowadays, computers and specialised software are used in different fields of language studies (Birkner, 2015), as will be explained later. Language variation in GPA is examined in this study from a sociolinguistic point of view, since the study of linguistic variation in contact languages can make a valuable

contribution to the field of (socio)-linguistic variation and change. Sociolinguist researchers were traditionally not interested in using corpora in their investigations (Baker, 2010:1) until 1996, when McEnery and Wilson demonstrated a first possible relation between sociolinguistics and corpora. They showed the value of supplementing the qualitative analysis of language with quantitative data (McEnery and Wilson, 2003). Later, Hunston (2002) and others, such as Beeching (2006), made further contributions. For example, in 2006, McEnery et al. also indicated that along with the speed of information processing, specialised software can classify and select words to assess their frequencies between major classes, for example male and female usage.

I review some of the benefits and limitations that have been found in the literature regarding the use of corpora evidence in sociolinguistic research, focusing on contact languages. First, however, I define sociolinguistics as a branch of linguistics and discuss language variation and change as an interesting subject for sociolinguists.

Wolfram and Fasold (1974) define sociolinguistics as a study of language in its social context. The major concern in sociolinguistics is to identify the relationship between the identity of people and the way in which they use language. The goal of sociolinguistics is to understand how social variables, such as age, sex, and social class influence people's use of language. Variation and language changes are major interests in sociolinguistics, and variation (i.e. the different ways in which people say the same thing, influenced either by regional or social variables) is a characteristic of language. In studying variation, sociolinguists often focus on analysing natural conversations. Furthermore, the size of the data varies; it can comprise a short conversation or several years of recorded interviews (Brinker, 2015). Since sociolinguistics depends essentially on authentic data, the use of corpora is useful for sociolinguistic studies, as in the case of my current study. However, there are some limitations to consider.

From one perspective, most researchers agree that the rapidness of processing a vast amount of data in a reasonably short period of time is one of the major advantages of using corpus evidence in sociolinguistics research when compared to older methods. Brinker (2015) states that the powerful software in computers enables sociolinguist researchers to have immediate access to accurate and authentic evidence contained in written and spoken corpora at a high processing speed. Computers can also process data more consistently and accurately (Barnbrook, 1996:11) compared with using paper-based corpora manipulated by human beings. McEnery et al.'s (2006: 6) article further emphasises the importance of using a computer and some specialised software to show the frequency of certain words, for instance in male and female usage. Corpora can be useful for both written and spoken data. In written corpora, such as the International Corpus of English (ICE), varieties of a language can be compared. This kind of corpus allows sociolinguist researchers to conduct comparative studies of English around the world (Brinker, 2015). Likewise, using spoken evidence from corpora has many advantages in sociolinguistic research.

Here, 'spoken language' means any language whose productions was originally in oral form, although the data of spoken language can include recordings of scripted speech (Wynne, 2005). Since the source of the data in my current study is a spoken language, I focus mainly on the use of recordings of naturally occurring spoken language. Compared to written language data, spoken language data is much more difficult to work with, and handling spoken language is a complex task that requires careful planning.

On the one hand, the data of a written language is typically composed of orthographic words, which can easily be stored in electronic text files (Wynne, 2005). However, representing the original features of a written language corpus text is still problematic, requiring consideration of aspects such as layout, accompanying diagrams, and font size; however, as Thompson (2005) states, the main problem is how to describe 'what can be seen' in written language.

In the case of spoken language data on the other hand, the main problem is how to represent the corpus in orthographic transcription or other symbolic means:

> Not only the transcription but also the process of data capture itself is problematic: an audio recording of a speech event is only an incomplete view of what occurred, not only because of possible technical deficiencies, but also because visual and tactile features are lost. (Thompson, 2005 in Querol-Julián and Fortanet-Gómez, 2014: 81)

To overcome this problem, Thompson suggests that it is possible for an analyst to also use a video recording to present the speech event; however, this video recording cannot capture the faces and body language of the participants themselves in the event. He adds that the corpus developer should bear in mind the complexities of working with spoken language data, such as during the task of compiling and designing a spoken language corpus. It is clear that building a spoken corpus highly depends on the usage purposes, which vary for each corpus. In my study, I am interested in the patterning of language: the morphosyntactic features of GPA.

For the study of morphosyntactic features, there is little need for highly sophisticated transcription – the quantity and speed of transcription work are the main concerns – whereas the study of segmental phonology requires less data and a high degree of accuracy and detail in the phonetic transcription of recordings, with links, where possible, to the sound files. However, Thompson (2005) comments that the possibilities for linking the transcript with audio/video files are limited at present, but should develop rapidly in the coming years. Furthermore, McEnery (2012) states that it is possible to link back transcriptions of spoken materials systematically to the original recording through a process called time alignment. This process, done through a computer, allows one to search a spoken corpus easily and to hear the section of the recording that matches a particular search result. Examples of corpora which use time the alignment process include, but are not limited to, the COLT corpus of London teenage speech, the International Corpus of English British component (ICE-GB), and the Origins of New Zealand English (ONZE) corpus. Moreover, McEnery (2012) suggests that this process is much more useful in phonemically transcribed material than in orthographically transcribed

material because the former kind of spoken corpora ‘often normalise the form of the words in the text to standard spellings, meaning that orthographically transcribed material is rarely a reliable source of evidence for research into variation in pronunciation’.

Today, a few disadvantages still remain. Throughout Lindquist (2009: 10), cited by Brinker (2015) that ‘the amount of evidence that is available is limited but it does not mean that it actually does not exist’. The next section will highlight some of these disadvantages which still remain today and which are overcome.

### 3.6.3. Disadvantages of Using Corpora in Sociolinguistic Studies

Researchers argue that there are some limitations in using spoken corpora, such as the limited amount of available evidence, errors made by speakers, a lack of metadata, and the absence of contact with the informants (Brinker, 2015). From the 1950s onwards, the use of corpora in linguistic analysis has been strongly criticised in the literature, notably by Noam Chomsky. The following discussion focuses on Chomsky’s criticisms and determines which are true versus untrue today.

Chomsky’s initial criticism was presented in the late 1950s when the corpus methodology became marginalised, if not totally banded, because of the alleged skewedness of corpora. He claimed that

> Any natural corpus will be skewed. Some sentences will not occur because they are obvious, others because they are false, still others because they are impolite. The corpus, if natural, will be so wildly skewed that the description would be no more than a mere list. (Chomsky, 1962: 159)

At the time, Chomsky’s criticism was true when linguists of that era, such as Jesperson (1909, 1949) and Fries (1952), used shoeboxes filled with paper slips to store their data instead using computers, and the size of the shoebox corpora was small and not representative. They also used paper slips as well as human hands and eyes to study grammar, which refers to paper-based corpora. It was impossible to collect large bodies of language data by using paper-based corpora. Thus, Chomsky’s criticism at that time was true, and those corpora could not be

anything other than skewed. However, now, with the significant development of technology, especially with powerful computers that offer massive storage and high processing power, the creation of massive corpora has become feasible.

Second, McEnery et al. (2006: 3-4) reported that Chomsky believed that the use of corpora is not valid for research purposes: 'corpus linguistics doesn't mean anything' (cited in Andor, 2004: 97) from a language perspective, as he claimed that corpus linguistics is a waste of time, because it is capable of focusing only on external phenomena of language, whereas the main concern is to know how language works (I-language) not how language is observed (E-language), as in the corpus-based approach (i.e. 'I' means 'internal' and 'E' means 'external'); thus, it is not scientific (Richards and Pilcher, 2016). With regard to Chomsky's second criticism, a number of linguists, for example McEnery and Hardie (2012: 26) have argued:

> Chomsky's view is at best somewhat naïve and at worst deliberately misleading…just as observation of the universe through astronomy can help to prove the hypotheses of physicists such as Einstein, so observation of language through corpora can help linguists to understand language.

Furthermore, McEnery and Wilson (2001) indicated that in a remarkably short period of time, Chomsky changed the direction of linguistics away from empiricism and towards rationalism. It has been argued that 'corpus linguistics is concerned with language use in real contexts [and] therefore, it is often contrasted with Chomskyan linguistics, which emphasizes language competence and often involves made-up examples as the basis of its exploration of language' (Adolphs and Lin, 2013: 597; cf. McEnery and Wilson, 2001).

However, this was not the only criticism that Chomsky had of the early corpus linguistics approach. McEnery and Wilson (2001) argued that Chomsky considered the language of corpus approach as finite and able to be collected and enumerated, while natural language is infinite, and a corpus thus cannot describe a natural language entirely. In response, McEnery and Wilson (2001) stated that all these claims and assumptions are twisted and express partial views

of linguistic methodology, some contain mis-truths and exaggeration (see Chapter 2 in McEnery and Wilson, 2001 for a detailed examination). They concluded that

> Although nowadays modern computer technology allows us to collect much larger corpora than those that Chomsky was thinking about, his criticisms still must be taken seriously. This does not mean that we should abandon corpus linguistics, but instead try to establish ways in which a much less biased and representative corpus may be constructed.

Furthermore, researchers argue that language is in continuous change, and collecting all the evidence needed is therefore impossible, but nowadays there are corpora known as monitor corpora, which are frequently updated and supplemented with new data (Brinker, 2015), such as the Bank of English (BOE). However, some linguists, such as Leech (1991:10), did not consider a monitor corpus to be a real corpus but rather an 'ongoing archive'.

A second disadvantage of using corpora in sociolinguistics studies is related to the errors that speakers might make when producing evidence for these studies. Lindquist (2009:10), cited in Brinker (2015), reported that a 'corpus does not filter mistakes or errors present in speech'. To avoid errors such as grammar mistakes, Brinker (2015) suggested that native speakers may assist the sociolinguist in making decisions regarding what is correct and what is wrong. Thus, it is the researcher's responsibility to select the appropriate evidence carefully.

There are three limitations in using corpora in sociolinguistic studies which should be dealt with carefully (McEnery and Wilson, 2001: 116). One of these limitations is 'the lack of sociolinguistic metadata encoded in currently available corpora'. McGlashan (2013) defines metadata as 'data about that data' – it offers information about, for example, the author, publication date, and title of a written text, while in spoken texts the information is about the speaker (Baker et al., 2006 cited by Brinker, 2015: 115). Knowledge of such information is important to interpret the evidence as accurately as possible.

Moreover, Biber, Conrad, and Reppen (1998) claim that to have a wide view of an interaction, it is not enough to know only who addresses the utterances; the exact category of all participants

involved in the conversation must also be known. Based on Biber et al.'s claim, Flowerdew (2012) stated that using all of this information in sociolinguistic variables would be a demanding and time-consuming task. In addition, McEnery and Hardie (2011: 61) indicated that there are other kinds of ethical issues, such as metadata, that affect respondents. They wrote:

> Not only the privacy of what is said in their conversation, but also the privacy of their personal information….they may provide information about themselves to the corpus complier which is useful in the generation of demographic metadata but is not to be incorporated into the metadata itself.

Another limitation is that sociolinguistic studies are mainly concerned with spoken language, and most corpora are comprised of material which comes in written form. Brinker (2015) provides an example: Access to spoken data in the first generation of spoken corpora, developed in 1960, is not possible because they contain only written words. Moreover, the developed corpora of the second generation in 1990, such as the British National Corpus (BNC), contain 100 million words or more, yet the underlying spoken data is not widely accessible. Therefore, Andersen (2010: 555), cited in Brinker (2015), draws the conclusion that this limitation can decrease the success of using corpus evidence for studies focused on spoken data. In my current study, conducted on non-standard varieties of the spoken GA language, I encountered the obstacle that most corpora contain written forms of Standard Arabic or other standard varieties which are not suitable for my variation study. Regarding transcription of spoken corpora, there is no agreement among transcribers in relation to spoken data interpretation, especially at the phonetic or prosodic level (ibid). Furthermore, the quality of sound recordings is usually poor because some researchers record with old equipment and others have background noise.

A further limitation in using corpora in sociolinguistics is the use of informants. The researcher has the accessibility to know informants' personal information, but their identities are still kept anonymous, and the researcher must not have contact with them (ibid). However, Andersen

(2010: 556) claims that the best corpus evidence will not be complete without contact with the informants. Moreover, researchers will obtain a complete picture of the situation if they collect evidence through face-to-face interaction. For instance, recording informants makes them feel uncomfortable, and it affects the way they speak, especially if they talk about their personal experience. In such situations, the researcher may fail to transcribe their emotions properly on paper. Alongside this, Widdowson (2000) argues that one of the weaknesses of researchers in their analysis involves depending exclusively on textual corpora. For example, a researcher does not have to access the speakers and writers' convictions and self-perceptions, such as attitudes, since 'Users are not always aware of how they use the language, and corpora are unable to provide researchers with a full description of what the speakers mean from a personal perspective' (ibid). For example, 'dear' can have different meanings; it can express gratitude, affection, or sarcasm. Corpora only represent lexical and grammatical features of language within a 'surface-structure' level (ibid in Friginal and Hardy, 2013: 62).

I now discuss other problems in using corpora in a number of interesting areas in sociolinguistic research, as identified by Brinker (ibid), cited in Friginal and Hardy (2013: 64). Corpora are not suitable tools for determining and analysing the following areas accurately. The first area pertains to the sociophonetic characteristics of speech, such as vowel and consonant sounds (segmental features), intonation, rhythm, and stress (suprasegmental). The second area is diglossia, first defined by Ferguson (1959) as the H (high) variety and L (low) variety to be two divergent forms of the same language. Later on, the definitions of diglossia were reviewed and amended by Fishman (1971) and subsequently by many other sociolinguists over the years. It has become a term used to classify communication situations in societies, where exchanges are made in two distinct codes which are either two language varieties or two languages. The third area encompasses pronunciation and accent, which are difficult to analyse, particularly in

studies of attitude and prejudice (ibid). Finally, another inconvenience is that annotation might affect corpus evidence indirectly.

Corpus annotation is the interpretative linguistic information added to a corpus (Leech, 2004), and two methods are employed to annotate a corpus – either manually or automatically. Hunston (2002: 91) argues that the results produced by computer software are not as accurate as those produced by humans. Sinclair (1991), however, does not agree with Hunston's views on annotating corpora by humans. He states that annotation made by a human might result in inconsistencies. Brinker (ibid) reports that annotation might cause some changes to the corpora and that the interpretation of the evidence might be affected by such changes. In addition, Barnbrook (1996: 82) claims that the researcher will be confused while looking for some evidence because this evidence is sometimes omitted, and occurrences which are not required are shown; therefore, the researcher ends up performing additional work.

Every point of view has another angle, and every angle has its merit. McEnery and Wilson (1996: 89-90) stated that a spoken corpus can be used in contact language studies because it provides a broad sample of speech with a wide range of selected variables, such as speaker gender, age, class, and genre (e.g. news readings, poetry, legal proceedings, etc.). The greatest advantage of this spoken corpus is that generalisations can be made about spoken language, as the corpus is as wide and as representative as possible and allows for variation within spoken language to be studied (ibid). The study of linguistic variation in contact languages can make a valuable contribution to the field of sociolinguistic variation and change. Unlike Baker (2010: 1) who stated that until 1996, sociolinguist researchers were not interested in using corpora in their investigations.

Yet, within the field of P/C research – with the exception of a small number of studies, some of which I review in Section 3.18 above – linguistic variability is hardly ever investigated. Thus, corpus linguistics will afford me the opportunity to explore the variations among GPA.

Moreover, a spoken corpus provides naturalistic samples of a speech to a greater extent than elicited speech; therefore, the findings will reflect language as it is spoken in 'real life' (ibid). The main purpose of my research is to study morphosyntactic structure of GPA, and with this transcribed and annotated corpus, it would be easier to carry out large-scale quantitative analyses than with fresh raw data (ibid).

To sum up the discussion about both the advantages and the disadvantages of using corpus evidence in sociolinguistic studies, I conclude that the advantages considerably outweigh the limitations for several reasons. First, the development of computer technology and the use of specialised software have revived corpus-based approach studies. Such software enables researchers to have instant access to vast amounts of authentic data which is more natural to analyse, process, and manipulate rapidly at minimal cost, for example searching, selecting, sorting, and formatting. Furthermore, computers can avoid human bias (i.e. annotation bias) in an analysis, thus making the result more reliable. Second, studying language variation requires accurate data, which nowadays can be found in a number of specialised corpora, for example the Michigan Corpus for Academic Spoken English (MCASE). Corpus data can highlight differences that intuition alone cannot perceive, for example in the use of 'totally', 'absolutely', and 'completely'. However, while using corpora has numerous advantages, there are still limitations that can be improved and overcome. Although a spoken language is the most common way in which people interact with one another in everyday life, corpora of natural speech are still relatively rare and relatively costly (Newman, 2008: 27). Unfortunately, only a small percentage of corpora are based on spoken language texts, while most contain written texts. A third disadvantage, in the words of Chomsky, is that 'a corpus does not even tell you what is possible' (Aarts, 2000: 6). Hilpert (2014: 20) adds that 'Chomsky would have a point if the quote were altered to "A single, isolated example never tells you what is impossiple. In

fact, it does not even tell you what is possible". So, if that is the case how can we determine what is possible and impossible?' grammatically or syntactically.

Moreover, researchers agree that there are still problems in transcribing spoken language and that there are disagreements among transcribers (Brinker, ibid). Sometimes, the sociolinguist wants to research some aspects of the spoken laguage which cannot be retrieved by written transcription (ibid). One possible solution sugsted by Brinker (ibid) is that it would be more suitable to use corpora as a complemenary tool among other methods. In fact, there are not many mistakes in a corpus, and words andsentences are mostly correct. Regarding transcription, modern corpora comprise a grat deal of detailed contextual transcription. Finally, Friginal and Hardy (2013:65), cited in Brinker (ibid), state that 'annotation is also increasingly improving and supported with digital recordings'. Most disadvantages will likely be solved as technology advances. Next, Section 3.7 provides the conclusions of this chapter.

### 3.7. Conclusion to Chapter 3

This chapter mentioned that the emergence of P/C languages is still a matter of ongoing discussion and debate. The two most opposing views are found in the universalist and the substratist linguistics theories (see Holm, 2000; Ramat, 2009; Siegel, 2008a, 2010). Regarding the lack of extensive historical documentation on the emergence of creole languages, it is difficult to choose between the opposing views. Interestingly, the supporters of both views can often use the same evidence. For instance, Bickerton (1984) stated that one of the universal features of adult SLA is that P/C languages tend to be analytic rather than synthetic languages. Opposing this, Holm (2000) reported that one of the most common features of many African substrate languages is analytic morphosyntax. It is noticeable that the construction of universalist and substratist linguistics theories are based on structurally similar languages (i.e. either European superstrate languages or African substrate languages), where such theories are supposed to apply to all contact languages, regardless of their superstrate or substrate. The

purpose of the current study is to investigate the gender and language variation in an Arabic lexified pidgin (i.e. GPA) resulting from morphosyntactic differences in the length of stay in the Gulf and language variation in the speakers' L1s, which I compare to the potential effect of the universal parameters in adult SLA. Thus, I aim to find new evidence for some of the above theories. The next chapter provides a morphosyntactic description of selected linguistic features of GPA and GA, the superstrate language of the pidgin under investigation.

# Chapter 4: Description of Gulf Arabic and Gulf Pidgin Arabic

This chapter is devoted to a syntactic description of the following five selected morphosyntactic features of the contact language under investigation in this study, namely, GPA, and the lexifier, GA: definiteness and indefiniteness, coordination, copular verbs, pronouns, and agreement in the VP and in the NP and ADJP. These five morphosyntactic features are adequate to test the proposed typological features that might be found in all P/C languages worldwide (irrespective of their input languages). As discussed in Chapter 3, P/C languages are generally believed to have a reduced, if not absent, inflectional system and reduced derivation (reduction of agreement markers in verb and noun agreement), and a small inventory of function words (copulas, definite and indefinite articles, and pronouns). These generalisations reported in the literature have been proposed after careful examination of tens, or sometimes even hundreds, of P/C languages (e.g. Bakker, 1995: 2003; Roberts and Bresnan, 2008; Bakker, Daval-Markussen, and Muysken, 1994; Parkvall, 2011; Sebba, 1997; Siegel, 2004). Thus, this chapter also defines them in GA, stating strategies used to form them in sentences and clarifying contexts in which they might be absent. I discuss each morphosyntactic feature in turn in the following subsections, as they are considered in my quantitative investigation (see Section 6.1). For all their variety, the Arabic dialects share a number of features that collectively distinguish them from CA. Some of these features are easily spotted and have long been well known (e.g. the lack of case and mood endings); others require more careful study (Blanc, 1970: 42). All examples below are provided by me, a native speaker of GA.

### 4.1. Description of the Selected Features of Gulf Arabic

#### 4.1.1. Definiteness and Indefiniteness in Gulf Arabic

A noun in Arabic – any noun that is derived from the verb – can be either indefinite نَكِرَة (*nakirah*) or definite مَعْرِفَة (*ma'rifah*). The indefinite النَّكِرَة (*an-nakirah*) is almost always changeable to the definite form following two morphological rules[1]. In contrast, not every definite noun is changeable to indefinite; certain nouns exist only in the definite form.

- **Definiteness:** A definite noun in Arabic is a noun that refers to a specific entity. There are three types of definite nouns. First, there are definite nouns in the Arabic language to refer to something specific only. These include pronouns, such as 'I' (*ana* انا) , 'he' *(hu* هو), and 'she' *(hi* هي); demonstrative pronouns, such as 'this.M' (*haða* هذا) and 'this.F' (*haði* هذه) versus 'that.M' (*haðak* هذاك) and 'that.F' (*haðik* هذيك ); relative pronouns, such as 'who' (*alði* الذي and *allati* التي). Second, there are proper nouns – nouns that are used to refer to a specific person or place – such as Ahmed (احمد), Dubai (دبي), and so forth. Third, there are definite nouns that are derived from the basic form (i.e. changing the indefinite noun into a definite noun). The prefix *al-* is the marker for definiteness in GA.
- **Indefiniteness:** An indefinite noun in Arabic is a noun that refers to a non-specific entity. In GA, the absence of the definiteness marker *al-* from the noun or adjective typically marks indefiniteness (Almoaily, 2012). While there are no overt markers for indefiniteness in GA, it can be expressed indirectly.

To summarise, GA definiteness and indefiniteness have been discussed in this section. Definiteness in GA only has the overt marker (the prefix *al-*), but actually there are other ways

[1] There are two ways to change an indefinite noun in Arabic to a definite noun: 1) by adding the Arabic definite article الـ (*al*) to the beginning of the indefinite noun and 2) by adding the indefinite noun to a definite noun; that is, forming the إِضَافَة (*idafah*, or addition) structure called a *construct state* (see Hassan, 1987, and Schulz, 2004).

to express definiteness in the language, such as *idafah* (‘addition’) or pronominal clitics (Almoaily, 2012). As for GA indefiniteness, while there are no overt markers, it can be expressed indirectly.

### 4.1.2. Coordinating Conjunctions in Gulf Arabic

A coordinating conjunction, *haraf ataf* (حرف عطف), is a particle which connects two words, phrases, or clauses together. According to Feghali (2004), the most common conjunction markers used in GA are the prefixed particles *wa* (وا), *laakin* (لكن), and *aw* (او), which usually translate as ‘and’, ‘but’, and ‘or’, respectively. There are also less common coordinating conjunction markers, according to Almoaily (2012), such as ‘then’*f* (*a* فا), ‘or’ (*willa* ولا), ‘either.. or’ (*ya ... ya....* يا ...يا...) , ‘but’ (*bass* بس), and ‘not.. nor.’ (*la... wala* …لا....ولا). The following is a brief discussion of the most commonly used of these GA conjunction markers, *wa* (‘or’).

**(i) *wa* و (‘or’)**: This is an additive conjunctive marker that links nouns, phrases, clauses, and paragraphs (Feghali, 2004). As can be seen in Example (1) below[2], *wa* is a prefix attached to the linked element at the beginning of a sentence or a paragraph. In multiple coordination, as in (2), *wa* is attached to every linked element, while in listing, it is replaced by a comma in English.

(1) akala al-walad w = šarab
ate the boy and= drank
‘the boy ate and drank’

[2] For every example throughout this thesis, there are three lines: the first is a transliteration in Roman script; in the second line, the sentence/utterance is glossed by using all the Leipzig Glossing Rules Conventions (refer to page x for a full list of conventions used in this study); and in the third line, I provide an English translation of the extract.

(2) šif-t Nada w =Amal w =Sara

Saw-1SG Nada and =Amal w =Sara

'I saw Nada, Amal, and Sara'

This section provided a discussion of GA conjunction markers. It concludes that most of the GA conjunction markers are free morphemes. The next section discusses copular verbs in GA.

### 4.1.3. Copula in Gulf Arabic

The word copula originates from the Latin noun for a 'link or tie' that connects two different things (Tahir, 2009). In English, it refers to the verb *be*, and verbs which function as a copula are variously referred to as copulative, equative, intensive, or support verbs, such as *seem*, *become*, *appear*, *feel*, and *look* (Quirk, Greenbaum, Leech, and Svartvik, 1985). Some grammarians assign the term non-copulative to languages which do not have a copula, such as Arabic, Russian, and Hungarian. Other grammarians state that such languages have a copula but that it is expressed in different ways depending on the tense, case, and context. For example, Holes (1990) states that the GA copula is only explicit in past and future sentences, whereas it is implicit in the present tense. In this regard, the term 'zero copula' refers to the linguistic phenomenon in which the presence of the copula is covert rather than expressed overtly. In Arabic, a sentence has three elements: *Al-musnad*[3], *Al-musnad Ilayhi*, and *'isnad*[4]. In English, a copula is called the link between the subject and the predicate. There is no copula in Arabic corresponding to *be* in English. Therefore, some grammarians consider Arabic to be a non-copulative language. However, the copula does exist in Arabic; it is just not expressed syntactically in the same way as it is in English (Alotaibi, 2018). Rather, it is expressed

[3] According to Sībawayh and al-Mubarrad, the definitions of *al-musnad* and *al-musnad 'i/avhi* are both indispensable parts of the sentence; it is not possible to form a sentence without them, and neither of them can exist without the other.

[4] According to Wright and Caspari (1896), the term *'isnad* denotes the relation between the subject and the predicate as 'the assignment of a predicate to a subject'.

according to the tense, case, and context. It is consequently expressed by different phonetic, linguistic, grammatical, and morphological phenomena. For example, the copulas ‘be.PST’ (*kaan* كان), ‘will be’ (*B-itikuun* بيكون), ‘became’ (*sˁaar* صار), and ‘will become’ (*B-yi-sˁiir* بيصير) agree with the subject in number and gender. Only past and future tense sentences display an overt verbal copula which agrees with the subject. To illustrate, Examples (3), (4), (5), and (6) display the presence of the copulas *kaan* and *sˁaar* in the past and future tenses, while Example (7) demonstrates copula absence in the present tense:

(3) kann Ahmed khayf ams

COP.PST-SG.M Ahmed afraid -SG.M yesterday

‘Ahmed was afraid yesterday’

(4) sˁaar el-bayt gideem

COP.PST.SG.M DEF-house old.M

‘The house became old’

(5) B-itikuuni hina bukra?

FUT- COP.2SG.F here tomorrow?

‘Are you going to be here tomorrow?’

(6) B-yi-sˁiir el-bayt gideem baʕd ʕašir sanaw-at

FUT-SG.M-COP DEF-house old.M after ten.M years-PL.F

‘The house became old.’

(7) Ø alħiin ana mašgu:l

Ø now I busy

‘I am busy now’

In Sentences (5) and (6), the prefix *b-* indicates the future tense, where the rest of the word in Sentence (5) indicates the second-person feminine, while in Sentence (4) it indicates the third-person masculine.

Arabic conjugation involves gender, number, and person. Table 4.1 shows conjugation of the copula in the imperfective or future tense. The next section discusses subject, object, and possessive pronouns in GA.

| **Person and Gender** | **Singular** | **Dual** | **Plural** |
|---|---|---|---|
| **1F and M** | *b-kūn* (will be) | - | *b-nakūn* (will be) |
| **2 M** | *b-ikūn* (will be) | *b-ikūnun* (will be) | *b-ikūnūn* (will be) |
| **2 F** | *b-itakūnī* (will be) | *b-itkūnān* (will be) | *b-itkunan* (will be) |
| **3M** | *b-iakūn* (will be) | *b-iakūnun* (will be) | *b-iakūnūn* (will be) |
| **3F** | *b-itkūn* (will be) | *b-itkūnān* (will be) | *b-itkūnān* (will be) |

Table 4.1: The conjugation of GA imperfective copula.

### 4.1.4. Pronouns in Gulf Arabic

#### *4.1.4.1. Personal Pronouns in Gulf Arabic*

This section briefly describes and exemplifies GA subject, object, and possessive pronouns. They all inflect for number, person, and gender. The following three subsections discuss the various inflections (gender, number, and person) for GA personal pronouns.

**A) Subject Pronouns**

Subject pronouns, *Damaa'ir al-faa3il* (ضمائر الفاعل), are used in Arabic to substitute nouns and to refer to them. They are considered to be nouns and stand-alone (free morpheme), as in Sentence (8):

(8) El-talab ġayab, Hu mariiðˤ

DEF-student absent PST.S.M, he-SGM ill

'The student is absent; he is ill.'

Table 4.2 lists the forms of the GA subject pronoun:

| Pronoun | | Transliteration | Examples |
|---|---|---|---|
| **1st** | SG | *ana* | *ana hina* (‘I am here’) |
| | PL | *Ħinna (inna)* | *ħinna hina* (‘We are here’) |
| **2nd** | SG.M | *int (inta/ anta)* | *int hina* (‘you-SGM are here’) |
| | SG.F | *Inti (anti)* | *inti hina* (‘you-SGF are here’) |
| | PL | *Intum (into)* | *intum hina* (‘you-PL are here’) |
| **3rd** | SG.M | *hu* | *hu hina* (‘he-SGM is here’) |
| | SG.F | *hi* | *hi hina* (‘she-SGF is here’) |
| | PL.M | *hum* | *hum hina* (‘they-PLM are here’) |
| | PL.F | *hn* | *hin hina* (‘they-F are here’) |

Table 4.2: GA subject pronouns (adapted from Feghali, 2004).

It is worth mentioning that in GA, the subject pronoun is frequently dropped (optional). The subject is apparent from the verb conjugation, so using the subject pronoun is not necessary in such cases except for emphasis – refer to Example (8) above. However, in equational (verbless) sentences such as Example (8) above, the subject pronoun is needed.

**B) Object and Possessive Pronouns**

Possessive pronouns, *Damaa ʕ ir al- mulkiah* (ضمائر الملكية), and object pronouns, *Damaa’ir al-maf ʕuul bihi* (ضمائر المفعول به) do not appear as separate words; they are always bound morphemes marked as suffixes and attached to the end of nouns, verbs, and particles (Qafisheh, 1977). These pronouns indicate possession when they are attached to nouns, and they function as the object when they are attached to verbs, active participles, or prepositions (Almoaily, 2012). Table 4.3 lists the GA object and possessive pronouns.

| Pronoun | | Transliteration | | Examples |
|---|---|---|---|---|
| **1st** | SG | *-i* | OBJ | *gal-l-i* (‘said-to-me’) |
| | | | POSS | *kitab-i* (‘book-my’) |
| | PL | *- na* | OBJ | *gal-li-na* (‘said-to-us’) |
| | | | POSS | *kitaab-na* (‘book-our’) |
| **2nd** | SG.M | *- ik* | OBJ | *gal-l-ik* (‘said-to-you-SGM’) |
| | | | POSS | *kitab-ik* (‘book-your-SGM’) |
| | SG.F | *-ič (-its -iš -is)* | OBJ | *gal-l-its* (‘said-to-you-SGF’) |
| | | | POSS | *kitab-ič* (‘book-your-SGF’) |
| | PL | *-kum* | OBJ | *gal-li-kum* (‘said-to-you-PL’) |
| | | | POSS | *kitab-kum* (‘book-your-PL’) |
| **3rd** | SG.M | *- ah* | OBJ | *gal-l-ah* (‘said-to-him’) |
| | | | POSS | *kitab-ah* (‘book-his’) |
| | SG.F | *- ha* | OBJ | *gal-l-ha* (‘said-to-her’) |
| | | | POSS | *kitab-ha* (‘book-her’) |
| | PL.M | *-hum* | OBJ | *gal-l-hum* (‘said-to-them-M’) |
| | | | POSS | *kitab-hum* (‘book-their-M’) |
| | PL.F | *- hin* | OBJ | *gal-li-hin* (‘said-to-them-F’) |
| | | | POSS | *kitab-hin* (‘book-their-F’) |

Table 4.3: Object and possessive pronouns in GA (adapted from Feghali, 2004).

It is evident from the discussion above that GA personal objects and possessive pronouns have affixes attached to the noun or to the verb, while subject pronouns are free. Both are used to mark for gender, number, and person. The following two sections explore agreement in the VP and the NP.

### 4.1.5. Agreement in Gulf Arabic

#### *4.1.5.1. Verbal Agreement in Gulf Arabic*

In traditional Arabic grammar, there are two basic types of sentences, based on what the sentence’s first word is. A nominal sentence, الجملة الاسمية (*al-jumla l-ismiyya*), is one in which the sentence’s first word is a noun (9), and a verbal sentence, الجملة الفعلية (*al-jumla l-fiʔliyya*), is one in which the sentence’s first word is a verb (10).

(9) al-walad mariiðˤ

DEF-boy ill

‘The boy is ill’

(10) jlis-at al- xadem-ah fi al-bayt

stayed-3SG.F DEF-house maid- SG.F at DEF-home

‘The house maid stayed at home’

(11) ʕamam -*i* ṭalʕ -*aw* badri

uncle-PL.M went out-3PL.M early

‘My uncles went out early’

Although the equational sentence in (9) contains a verb in English, it does not have one in Arabic[5]. However, it literally translates as ‘The boy ill’; the ‘is’ is understood. The verbal system in GA agrees with its subject in both gender and number. Although the declarative sentences in (10) and (11) have two basic word orders – VSO (10) and SVO (11) – the verb obligatorily agrees with an NP subject in all Ø- features.

According to Almoaily (2012), the GA verb root and tense determine the way in which the subject-verb agreement is applied in gender, number, and person. This section briefly discusses how the GA verb inflects for number, gender, and person. I provide examples (all examples below are provided by me, a native speaker of GA) of the different structures of the GA verb and conjugation tables for each kind. First, there are two moods/tenses in Arabic:

- **The perfect/past**, *al-maadi* (الماضي), is used to indicate actions that have been completed.

This conjugation involves adding suffixes to the ‘base’ form of the verb.

[5] Arabic does not use a present-tense form of ‘to be’; instead, it uses a verbless sentence consisting only of a noun and an adjective.

- **The imperfect/present**, *al-mudaariʕ* (المضارع), is used to indicate actions that have not yet been completed. This conjugation requires the addition of prefixes and, in some cases, suffixes as well.

Some considerations are worth mentioning here. There are no infinitive forms of verbs in Arabic. Instead, the masculine third-person perfect form is typically used, for example *darasa* (درس) 'to study' (actual meaning: 'he studied').

Furthermore, according to Feghali (2004), the verb in GA falls into two main classes: 1) sound, *Sahiih* (الصحيح), and 2) weak, *muʕ tall* (المعتل). The subsections below explain these two classes and provide examples of each.

**1)** Sound verbs / Strong verbs, *al-fiʕl as-sahiih* (الفعل الصحيح): These are the verbs whose root is composed of three consonants, where the last two consonants are not alike (e.g. k-t-b كتب 'to write', and d-r-s درس 'to study'). There are two types of sound verbs: regular and irregular sound verbs.

- Regular sound verbs, *al-fiʕl as-sahiih as-saalem* (الفعل الصحيح السالم): These are the first types of sound verbs that do not have a *w* (و) or *y* (ي) as one of the three root letters. There are two moods for regular sound verbs: perfect mood and imperfect mood. Table 4.4 shows how the various forms of the strong verb *k-t-b* 'to write' agree with the subject in number, gender, and person[6] .The transliterations in the third column represent two of the common ways in which to pronounce the GA verb.

[6] Note that other, less common, dialectal possibilities are not mentioned (e.g. *kitbti* and *ktibti* 'write-2SGF').

| Subject | Arabic derivation | Transliteration | Meaning |
|---|---|---|---|
| **1SG** | كتبت | *katab-t* | ‘write-1SG' |
| **1PL** | كتبنا | *ktab-na/ katab-na* | ‘write-1PL' |
| **2SGM** | كتبت | *ktab-t* | ‘write-2SGM' |
| **2SGF** | كتبتي | *katabt-i/ktabt-i* | ‘write-2SGF' |
| **2PL** | كتبتو | *ktab-tu/ katab-tu* | ‘write-2PL' |
| **3SGM** | كتب | *katab/ktab* | ‘write-3SGM' |
| **3SGF** | كتبت | *ktibt-at/ ktab-at* | ‘write-3SGF' |
| **3PLM** | كتبو | *ktib-aw/ ktab-aw* | ‘write-3PLM' |
| **3PLF**[7] | كبتن | *ktab-an/ ktib-an* | ‘write-3PLF' |

Table 4.4: Various agreement forms of the GA verb stem *k-t-b* in the past tense.

Table 4.4 indicates that any GA strong verbs can be conjugated by replacing the three consonants of the stem and retaining the rest of the consonants and phonemes analogous to the *k-t-b* stem (Almoaily, 2012); for instance, *d-r-b* ‘to hit’ can be done with 1SG subject *drab-t*, 1PL subject *drab-na/ darab-na*, and so on.

Almoaily (2012: 70) mentioned that in the present tense, the stem of a strong verb will take another form than those stated for the k-t-b stem in Table 4.4 above. He also stated that ‘the procedure which creates these forms is still the same: prefixation, suffixation, and/or infixation’. See Almoaily (2012) for an exemplification all of the various agreement forms of the verb stem (j-l-s) in the present tense. He also indicates that the future marker /b/ can be attached to the stem of the strong verb (i.e. present tense forms) as a prefix. Thus, the future form for *ktb* is *b-aktab* (‘I will write’), the future form for *naktab* is *b-n-iktib* ‘we will write’, and so on.

The second type of GA sound verb is the irregular sound verb. These verbs of other verbal stems may take different conjugations, as highlighted in the next subsections.

- Irregular sound verbs: This category of GA verbs can be further subcategorised into two subtypes, namely, geminate/doubled verbs and hamzated verbs.

[7] For many GA speakers, the 3PL.F form is used interchangeably with this form in a random manner.

**(i)** Geminate/doubled verbs, *al-fiʕ l al-muDaʕaf* (الفعل المضعف): These are verbs in which the second and third consonants of the tri-consonantal stem are the same (a doubled consonant), for example دقّ *daqq* PST.3SG.M, يدقّ *yadiqq* PRT.3SGL.M (to knock), and ردّ *radd* , يّرد *yardd* (to reply). Table 4.5 shows how the verb r-d-d ('to reply') agrees in person, gender, and number with its subject in the past tense.

| Subject | Arabic derivation | Transliteration | Meaning |
|---|---|---|---|
| **1SG/2SG.M** | ردّيت | *radd-eet* | 'reply-1SG'/'reply-2SG.M' |
| **1PL** | ردّينا | *radd-eena* | 'reply-1PL' |
| **2SGF** | ردّيتي | *rdd-eeti* | 'reply-2SG.F' |
| **2SG.M** | ردّيتو | *radd-eetu* | 'reply-2SG.M' |
| **2PL.F**[8] | رديتن | *rdd-eetin* | 'reply-2PL.F' |
| **3SG.M** | رد | *radd* | 'reply-3SG.M' |
| **3SG.F** | ردّت | *radd-at* | 'reply-3SG.F' |
| **3PL.F** | ردن | *radd-an* | 'reply-3PL.F' |
| **3PL.M** | ردّو | *radd-u* | 'reply-3PL.M' |

Table 4.5: Agreement between doubled GA verb stem (r-d-d) and its subjects in the past tense.

For the present tense, the stem of a GA doubled verb has different forms than the stem *r-d-d*, as illustrated in Table 4.5. The procedure which creates these forms is still the same: prefixation, suffixation, and/or infixation, for example *a-ridd* ('PRS.1SG-reply') and *ni-ridd* ('PRS.1.PL-reply') – these verb forms are presented in Almoaily (2012). He shows how the verb *ʕ-d-d* ('to count') agrees in person, gender, and number with its subject in the present tense.

**(ii)** Hamzated verbs, *al-fiʕl al-mahmuuz* (الفعل المهموز): These are verbs where the glottal stop[9] 'ʕ(ء) is one of its stem consonants, for example اكل *akal* - يأكل *ya'kul* (to eat), سأل *sa'al* - يسأل *yas'al* (to ask). Table 4.6 shows how the verb a-k-l ('to eat') agrees in person, gender, and number with its subject in the past tense.

[8] For many GA speakers, the 2PLM is used interchangeably with this form in a random manner.

[9] Note that in many cases, the glottal stop in GA is converted to the short vowel /a/.

| **Subject** | **Arabic derivation** | **Transliteration** | **Meaning** |
|---|---|---|---|
| **1SG** | اكلت | *akal-t* | ‘eat-1SG' |
| **1PL** | اكلنا | *akal-na* | ‘eat-1PL' |
| **2SGF** | اكلتي | *akal-ti* | ‘eat-2SGM' |
| **2SG.M** | اكلتو | *akal-tu* | ‘eat-2SGF' |
| **2PL.F** | أكلتن | *akal-tin* | ‘eat-2PL' |
| **3SG.M** | اكل | *akal* | ‘eat-3SGM' |
| **3SG.F** | كلت | *kal-t* | ‘eat-3SGF' |
| **3PL.F** | كلن | *kal-an* | ‘eat-3PLM' |
| **3PL.M** | كلو | *kal-u* | ‘eat-3PLF' |

Table 4.6: Agreement between hamzated verbs and their subjects in the past tense.

In the present tense, the GA doubled verb has different forms to the ones illustrated in Table 4.6. These verb forms are shown in Table 4.7.

| Subject | Arabic derivation | Transliteration | Meaning |
|---|---|---|---|
| 1SG | اكل | *akal* | ‘PRS-1SG-eat' |
| 1PL | ناكل | *na-akal* | ‘PRS-1PL-eat' |
| 2SG.M | تاكل | *t-akal* | ‘PRS-2SGM-eat' |
| 3SG.F | | | ‘PRS-3SG.F-eat' |
| 2SG.F | تاكلين | *t-akl-iin* | ‘PRS-2SGF-eat' |
| 3SG.M | ياكل | *y-akl* | ‘PRS-3SG.M-eat' |
| 2PL | تاكلون | *t-akl-uun* | ‘PRS-2PL-eat' |
| 3PL.M | نياكلو | *y-akl-uun* | ‘PRS-3PL.M-eat' |
| 3PL.F | يأكلن | *y-akl-an* | ‘PRS-3PL.F-eat' |

Table 4.7: Various agreement forms of the verb stem (a-k-l) in the present tense.

**2)** Weak verbs, *al-fiʕl al-muʕtall* (الفعل المعتل): In this category of GA verbs, the root contains at least one vowel such as *w* (و) or *y* (ي) as one or more components of the root. These verbs take different forms to those shown in Table 4.4. Weak verbs can be further subcategorised into three subtypes, namely, assimilated verbs, defective verbs, and hollow verbs. Just as for strong verbs, all weak verbs inflect for tense, person, number, voice, and gender.

**i) Assimilated verbs**, *al-fiʕl al-mithaal* (الفعل المثال): The roots of these verbs begin with the semivowel *w* (و) or *y* (ي), but usually *w* (و); in the imperfect and in other situations, the *w* (و) often disappears (e.g. *wasal* وصل- يصل *yasil* 'to arrive').

**ii) Hollow verbs**, *al-fiʕl al-ajwaf* (الفعل الأجوف): These are verbs whose second component of the root is either the vowel *a* (ا) و or *y* (ي); in the perfect, the *w* (و) or *y* (ي) is replaced by an *alif* (ا). Hollow verbs are verbs whose root is composed of the vowel /a/ (e.g. 'sold' باع *baaʕ* - 'to sell' يبيع *yabiiʕ*).

**iii) Defective verbs**, *al-fiʕl al-naaqis* (الفعل الناقص): The final root of a GA defective verb ends with the semivowel *w* (و) or *y* (ي); for example, 'forgot' (نسى *nasa*) – 'to forget' (ينسى *yansa*) and 'appeared, seemed' (بدا *bada*) – 'to appear, seem' (يبدو *yabdu*) or ends with a short /a/ (e.g. قرأ *g-r-a* 'to read'. The final root radical is either و or ي and is doubled (i.e. the middle and the last consonants in the stem are one doubled consonant), where the second and third radicals of the root are the same.

This section shed light on subject-verb agreement in the GA verb. It is clear from the examples in Tables 4.4 to 4.7 above that the GA verb is relatively complex in that the subject and the verb root determine how the verb inflects for tense, number, gender, person, and voice. The next section discusses agreement in the GA NP and ADJP.

### *4.1.5.2. Agreement in the Noun Phrase and in the Adjective Phrase in Gulf Arabic*

Adjectives are words that modify nouns. In other words, an adjective has to be preceded by a noun, and this makes up a noun-adjective structure (Kremers, 2003). This section discusses the properties of adjectival agreement in NPs, adjectival or nominal agreement in numeric phrases, and adjectival or nominal agreement in demonstrative phrases in GA. Since my study is compared with that of Almoaily (2012), I investigate the same GA features that Almoaily considered in his study.

**1-Agreement Between the Noun and the Adjective(s) in Gulf Arabic**

Arabic has a particular construction in which an adjective has a subject internal to the ADJP (Kremers, 2003). In GA, adjectives agree with the head noun in gender, number, and definiteness (cf. Feghali, 2004; Smart, 1990; Almoaily, 2008, 2012; AlQahtani, 2016; Brustad, 2000; Holes, 1990). Agreement in number in the VP inflects for number as singular versus plural, whereas in the NP and in the ADJP, it also inflects for dual. However, GA has a dual marker only for some nouns, as opposed to Standard Arabic for which Examples (12–14) below provide an illustration of noun-adjective agreement forms in GA for the noun *mudarris* 'teacher' and the adjective *jidiid* 'new'.

(12) mudarris -ah jidiid-ah
teacher-SGF.INDEF new-SGF.INDEF
'A new female teacher'

(13) il-mudarris-een il-judad
DEF-teacher.PL.M DEF-new.PL.M
'The new male teachers'

(14) mudarris-tain jidid-tain
Teacher-DUF.INDEF new-DUF.INDEF
'Two new female teachers'

Adding more descriptive words (adjectives) to the noun is applicable because adjectives follow nouns in Arabic. Thus, they should agree with the head noun in gender, number, and definiteness (AlQahtani, 2016), as illustrated in Example (15):

(15) al-madrasa aS-aġira wa al-bay ðˤa

DEF-school.SG.F DEF-small.SG.F and=DEF- white.SG.F

'The small and white school'

When combining two head nouns regardless of their genders, the adjective is always in the masculine form. Agreement in number and in definiteness is still in effect, as shown in Example (16):

(16) il-mudarris-een w=il-mudarrris-at il-judad

DEF-teacher.PL.M and=DEF-teacher-PL.F DEF-new.PL.M

'The new male and female teachers'

The next subsection demonstrates the slight differences between noun/adjective agreements with number and the noun-adjective agreement discussed here. The following subsection thus describes the agreement between the noun/adjective and the number that Almoaily (2012) refers to.

**2- Agreement Between the Noun/Adjective and the Number in Gulf Arabic**

In English, words are either singular or plural, yet in Arabic, words can be singular *mufrad* (مفرد), dual *muthana* (مثنى), or plural *jamʕ* (جمع). The same applies to numbers, as numbers are words. The numbers one and two are adjectives, and behave as such, following the noun and agreeing in number and gender. They are rarely used except in counting, as the singular or dual form of the noun makes the number of items sufficiently clear. Used for emphasis or contrast, 'one' (واحد *waħid*) and 'two' (ثنين *Θnēn*) follow the noun they qualify and agree in gender and number (singular for *waħid*, dual for *Θnēn*). In this respect, they behave like any ordinary adjective. Examples (17–20) explain how to say 'one' and 'two' in GA:

(17) šarē-t el-kitab-ēn eθ:-intēn

Buy.1SG.PST DEF-book-DU DEF-two

‘I bought the two books’

(18) šarēt el-kitab-ēn

Buy.1SG.PST DEF-book-DU

‘I bought the two books’

Or

(19) garē-t maqal-ah *waħi-ah*

Read-1SG.PST article-SG.F one-SG.F

‘I read one article’

Or

(20) garē-t maqal-ah

Read-1SG.PST article-SG.F

‘I read one article’

In Sentence (18), the two (*eθ:-intēn* ثنين) and the one (*waħid* واحد) are optional. The dual or singular form of the noun makes the number of items sufficiently clear, such that both (17) and (18) have the same meaning, as do (19) and (20). Note that the cardinal numbers are the ones used for counting (one, two, three, four, five, etc.), and there is a masculine and a feminine variant for each number. Table 4.8 lists the GA feminine and masculine cardinal numbers from one to eleven.

| No. | Masculine | | Feminine | |
|---|---|---|---|---|
| **1** | واحد | *waħid* | وحده | *waħdah* |
| **2** | ثنين | *Θnēn* | ثنتين | *θintain* |
| **3** | ثلاث | *Θalaθ* | ثلاثة | *θalaθah* |
| **4** | أربع | *arbaʕ* | اربعة | *arbaʕah* |
| **5** | خمس | *χams* | خمسة | *χamsah* |
| **6** | ست | *Sitt* | ستة | *sittah* |
| **7** | سبع | *sabiʕ* | سبعة | *sabʕah* |
| **8** | ثمان | *θiman* | ثمنية | *θamanyah* |
| **9** | تسع | *tisiʕ* | تسعة | *tisʕah* |
| **10** | عشر | *ʕašir* | عشرة | *ʕašrah* |
| **11** | حدعش | *ħdaʕaš* | حدعش | *ħdaʕaš* |

Table 4.8: GA cardinal numbers (adapted from Almoaily, 2012).

The numbers from three to ten are also adjectives, but they are distinctive in some ways. They precede the counted noun, which is in the plural form. The number and the counted noun form 'addition' (إضافة *idafah* 'constructs[1]). The number appears in the *opposite gender* to the noun, which is known as 'reverse agreement', 'polarized' gender distinction, or 'chiastic agreement': the unit of number must appear in reverse gender agreement with the noun that is referred to. In other words, the number is feminine whenever the singular of the counted noun is masculine (cf. Bolozky and Haydar, 1986; Feghali, 2004; Shaalan, Monem, and Rafea, 2006). This means that the numbers from three to ten end in ة (*ta marbu ṭa*) before a masculine noun such as *kitab* (book), and with a feminine noun such as *ṭaliba* (female student), the number must not be marked by the feminine. However, Badawi's (1996) result shows that just half of his subjects were able to produce the reverse form for the masculine noun, while less than one third were able to produce the feminine form. He concluded that Colloquial Egyptian Arabic has no reverse agreement rule, and most of his subjects used the colloquial form in the feminine *ta marbu ṭa.* As a GA native speaker, I agree with Almoaily (2012) and Badawi (1996) that most people ignore *fuṣha* (Standard Arabic) rules and go with their native colloquial feelings.

[1] Known by Arab grammarians as idafah, or 'addition' (see Hassan, 1987, and Schulz, 2004).

Furthermore, numbers that are larger than ten always come in one form – a masculine form (Almoaily, 2012). For example,

(21) ʔišriin mudarris w= ʔišriin mudarris-ah
twinty teacher-PL.M and=twenty teacher-PL.F
'Twenty male teachers and twenty female teachers'

This subsection discussed agreement between numbers and the nouns or adjectives to which they refer in GA. The examples from (17) to (21) clearly depicted the agreement between the numeral and the quantified noun in number and gender. This agreement seems rather complicated, since it occurs in some cases and is absent in others. The following subsection discusses agreement between the noun or adjective and the demonstrative.

**3- Agreement Between the Noun/Adjective and the Demonstrative in Gulf Arabic**

Cheng and Corver (2013: 327) define the phenomenon of demonstrative agreement as a 'phenomenon whereby a demonstrative element agrees in number and gender with the noun of which it is a determiner'. In GA, demonstrative pronouns, *asma' alešarah* (أسماء الأشارة), are used in the same way as English demonstratives. However, Arabic has many more demonstratives than English due to gender differentiation – masculine, *muḏker* (المُذَكَّر), and feminine, *mu'nṯ* (المُؤنَّث). Demonstrative pronouns in GA inflect for number, gender, and proximity (Qafisheh, 1977; Holes, 1990). GA demonstrative pronouns for proximate objects are 'this.M', or *haða* (هذا), and 'this.F', or *haði* (هذه), while for distanced objects, they are 'that.M', or *haðak* (هذاك), and 'that.F', or *haðik* (هذيك). To illustrate how GA demonstrative pronouns inflect for gender and proximity, I examine the examples below:

(22) haða l- rajjal
this.MSG.PROX DEF-man.MSG
'This man'

(23) haðak l- rajjal

that-MSG. DIST DEF-man.MSG

‘That man’

(24) haði i- siyarh

this-FSG.PROX DEF-car.FSG

‘This car’

(25) haðik i- siyarh

‘that-FSG.DIST DEF-car.FSG

‘That car’

GA demonstrative plural forms inflect only for proximity, but not for gender (Hole, 1990). Accordingly, the plural demonstrative for proximate objects, *haðoli* (هذولي), and the plural demonstrative for distant objects, *haðolik* (هذوليك), are both used with feminine and masculine objects. Consider Examples (26) and (29) below:

(26) haðoli el- mudarris-iin

this.PL.PROX DEF-teacher.PL.M

‘These teachers’

(27) haðoli el- mudarris-at

this.PL.PROX DEF-teacher.PL.F

‘These female teachers’

(28) haðolik el- mudarris-iin

this.PL.DIST DEF-teacher-PL.M

‘Those teachers’

(29) haðolik el- mudarris-at

this.PL.DIST DEF-teacher-PL.F

‘Those female teachers’

In summary, while the verb root determines the verbal inflection in the GA verb, which leaves us with numerous verbal conjugations in GA, the agreement in the GA NP and ADJP also involves a large number of other conjugations. In the following sections, I describe GPA with respect to the morphosyntactic features that have been investigated above and which have been selected by Almoaily (2012). Simplification in the GPA morphosyntax is expected in comparison to the GA morphosyntactic system. The linguistic simplification in the GPA morphosyntax is discussed in Section 3.3.2.1 in Chapter 3. Thus, this section shows linguistic simplification in the GPA morphosyntax.

## 4.2. Gulf Pidgin Arabic

This section discusses the five selected morphosyntactic features listed in Section 4.1 and in Almoaily (ibid), as the results of my study are compared with his in order to investigate gender variation in GPA. This section provides a descriptive account of GPA, where all the examples below are from my fieldwork data. The code number of the informant is placed immediately after each example.

### 4.2.1. Definiteness and Indefiniteness in Gulf Pidgin Arabic

Definiteness is expressed inflectionally in GA (Roberts and Bresnan, 2008); the prefix *al-* is the marker for definiteness in GA, whereas there is no overt marker for indefiniteness in the language, although indefiniteness can be expressed indirectly (see Section 4.1.1). The GA definiteness marker is normally dropped in GPA, as Smart (1990) and Almoaily (2008, 2012) mentioned. The notion of definiteness and indefiniteness is not inherent inflection; it is fully

contextualised in GPA (Almoaily, 2012). However, Examples (30) and (31) from my data indicate that the GA definiteness marker *al-* in GPA is sometimes dropped, but retained in other cases (refer to Chapter 6 for the number of instances for each informant).

(30) awal ana šughul fi-al-*ʕiyadh* (M1)
first I work in-DEF-clinic
'I used to work in the clinic'

(31) bʕdyeen yi-ji Ø mustašfa
Then 1SG.F.PRS.come Ø hospital
'After that I came to work in the hospital'

In GA, the sentence in (30) would be as follows:

b*ʕd* *ki*ða ana jy-et a-štghl fi-al*musta*šfa
after such I come-1SG.PST 1SG.PRS-work in-DEF-hospital

It is clear from the discussion in this section that definiteness and indefiniteness in GPA are contextualised. However, the GA definiteness marker *al-* is used by a number of GPA speakers.

### 4.2.2. Coordination in Gulf Pidgin Arabic

In GPA, Smart (1990), as reported by Almoaily (2012), referred to the coordination system as asyndetic[2] linkage. Most of the GPA speakers in my study tended to drop coordination markers, while a large number of the speakers used some of the GA coordination conjunction markers (e.g. the GA markers *wa* 'and', *laakin* 'but', and *willa* 'or' were the most commonly used by the GPA informants).

[2] When coordination can occur without the presence of a coordinator, this is known as asyndetic coordination (Gast and Diessel, 2012). For example, 'quickly, resolutely, he strode into the bank.'

(32) Khlas ba'den *wa* khalas ya'nee arbaah sana (T2)
Finished later CONJ finish mean four year
'Later I will finish and I will complete four years'

(33) Wahad sanah Ø aθnen shahar (P2)
One year Ø two month
'I have been here for one year **and** two months'

The next subsection introduces a rather uncommon feature of pidgin languages. This feature is the use of a copula.

### 4.2.3. Copula in Gulf Pidgin Arabic

As discussed in Section 4.1.3 above, the copula is always absent in GA in the present context (Bahloul, 1993; Ryding, 2005, 2014; Benmamoun, 2008; Aoun, Benmamoun, and Choueiri, 2010). However, it is used for past and future tenses (Almoaily, 2012). Although a zero copula is possible for a number of extended pidgins (Velupillai, 2015), a non-zero copula exists in GPA, namely, *fi*, which is also used in the present tense (Almoaily, 2012). Potsdam and Alanazi (2014) and Avram (2012, 2017) suggest that this GPA copula *fi* derives from the Arabic morpheme *fi(h)* found in existential sentences which means 'there'. To exemplify the word *fih* in GA as an existential particle, examine Sentence (34):

(34) fii(h) awrag fi al-maktabah (GA)
there paper PL.F in DEF-library
'There are papers in the library'

In GPA, the word *fi(h)* might be used as a copula, as in (36), or as an existential particle, as in (35). In this project, I only investigate the use of *fi(h)* as a copula in GPA as a focus point in Almoaily's study.

(35) hena fi alheen tam tisʕah sana (M11)

here there now complete nine year

'I have now been here for nine years'

(36) hena fi muhandes (B1)

here Cop. engineer

'[My son] is engineer in Saudi Arabia'

Example (37) illustrates copula dropping in the past tense in GPA:

(37) ams ana Ø mariið (Su9)

Yesterday I Ø sick

'I was sick yesterday'

It is noticeable through the examples above that the use of the copula is optional in GPA. However, most participants in this study show a tendency to drop the copula, both in the present and in the past tenses.

### 4.2.4. Personal Pronouns in Gulf Pidgin Arabic

As discussed in Section 4.1.4, the GA pronominal system inflects for number, person, and gender in both independent (subject pronoun), and object and possessive pronouns. Smart (1990) and Almoaily (2012) both categorised personal pronouns in GPA into five pronouns: *ana* 'I', *inta/anta* 'you', *hu* (*wa*) 'he/they', *hiy:a* 'she', and *nihn* 'we'. Smart (1990) stated that the number of personal pronouns in GPA has reduced to only five, instead of the 18 GA pronouns illustrated in Tables 4.2 and 4.3 above. To understand the reason behind this reduction, examine the examples below excerpted from my data:

(38) banat huwa šughul katheer (B1)

Girls he work too much

‘All girls work too much’

Sentence (38) shows that there is no agreement in number; the third-person singular masculine subject pronoun *huwa* is used instead of the third-person plural feminine subject pronoun *–hin*. The GA third-person plural masculine pronoun *hum* (‘they.M’) and the third-person plural feminine pronoun *hin* (‘they.F’) are both replaced with the third-person singular masculine subject pronoun *huwa* (‘he.M’). Moreover, GPA speakers do not distinguish between masculine and feminine third-person plural pronouns in their speech.

Another noticeable reason is that GPA speakers use the second-person singular pronoun *inta*, regardless of gender, whereas in GA, the second-person singular pronoun inflects for gender (i.e. *int* ‘you.SGM’ and *inti* ‘you.SGF’). Examine Sentence (39):

(39) Zawj inta fi mawjowd? (S7)
Husband you Cop. alive?
‘Is your husband alive?’

The pronoun *ana* is used as a free morpheme and replaces all of the pronominal suffixes that are used as possessive pronouns and object pronouns (see Næss, 2008; Al-Azraqi, 2010).

(40) ay fi kalam hag ana saw-I (T12)
anyone Cop. speak of 1SG do-1SG
‘If anyone tells me to do that, I will do it’

(41) omer baby ana sitta (T9)
old baby 1SG six
‘My baby is six years old’

This subsection concludes that GPA speakers tend to use full pronominal forms instead of pronominal suffixes, as can be seen in Examples (39) and (40). Furthermore, as discussed in

Section 3.1.2, GPA seems to correspond to the global structural tendency of pidgin languages to lose inflections and use free pronouns rather than clitic ones (Almoaily, 2012; Avram, 2014).

### 4.2.5. Agreement in Gulf Pidgin Arabic

This section provides a description of subject-verb agreement in addition to agreement in the NP and in the ADJP in GPA. One of the predominant characteristics of pidgin languages encountered in the literature is a reduction in the agreement system (Bakir, 2010), and GPA is no exception. Section 3.3 in Chapter 3 illustrates that agreement markers are rare in pidgin languages. The reduced agreement system in GPA is discussed next.

#### *4.2.5.1. Subject-Verb Agreement in Gulf Pidgin Arabic*

In GPA, the subject-verb agreement is observed more strictly than in GA (Almoaily, 2012; Bakir, 2010; Næss, 2008; Potsdam and Alanazi, 2014; Al-Ageel, 2015). My fieldwork data demonstrates that the most obvious feature of the verbal system in this linguistic variety is the generalisation of using specific verbal forms. Bakir (2010: 205) states that '[no] changes are applied to this form to indicate difference in tense, mood, aspect, voice, or agreement in gender, number, or person with the subject'. In my study, the participants expressed the GA verb in different ways. They either use the third-person singular masculine imperfect or the perfect form of the GA verb, regardless of the gender, number, or person of the subject, as in (42) below:

(42) ana šughul fi mustašfa (T2)
I work. 3SG.PST in hospital
'I work in the hospital'

In GA, the 1SG form of the verb is used with the 1SG pronoun *ana*. Thus, the GA form of Sentence (42) would be

(43) ana a-štughel fi al- mustašfa

I PRS.1SG-work in DEF-hospital

'I work in the hospital'

Moreover, GPA speakers use the imperative form of the GA verb instead of the GA present or past verb (44):

(44) waraq hagi yanee khali (T2)

Papers mine means leave-SG.M.IMP

'I gave them my papers before I left'

In Example (44) above, T2 reports that she gave the Philippine government her papers before she left the country. In GA, the form of the verb used in this case would be the 3PL.M. past form, as demonstrated below:

(45) aʕṭit-hum awrIg-i gabal ma-safer

I gave-3PL.M.PST my papers before I left

'I gave them my papers before leaving '

In addition, GPA speakers replace verbs with nouns (46):

(46) nass kalam Arabi (Su 6)

people speech Arabic

'People speak Arabic'

In (46), the informant has replaced the GA inflected verb *tt-kalam* ('PRS1PL.M-say') with the noun *kalam* ('speech'). Hence, when translated to GA, (45) would be

(47) al-nas hena tt-akalm Arabi

DEF-People here PRS.2PL.M- speak Arabic

Apart from the above strategies that the participants use to express the GA verb, I also noticed verb deletion in my data. The verb is dropped in some situations when the meaning can be

inferred from the context (Almoaily, 2012). Sentence (48) illustrates that although there is no verb, the meaning is still clear from its context:

(48) bass Ø English wa=Tagalog (T2)
only Ø English and=Tagalog
'I speak only English and Tagalog'

In GA, the verb is *atkalam* ('speak-1SG PRS'). Thus, (48) in GA would be

(49) at-kalam bass engalizi wa=tagalog
at-kalam bass engalizi wa=tagalog

The discussion of verbal forms in GPA above generally shows the lack of subject-verb agreement in GPA and that it is heavily dependent on the context. This typical pidgin feature leads us to the conclusion that the GPA morphosyntactic system has a reduced, if not absent, inflectional system and agreement markers. The verb in GPA does not agree with the noun in person, number, or gender, as in (43). Instead, one form of the verb, the third-person singular masculine, tends to be used with all subjects, as in (44) and (45). The next subsection discusses agreement in the GPA NP and ADJP.

#### *4.2.5.2. Agreement in the Noun Phrase and in the Adjective Phrase in Gulf Pidgin Arabic*

This section illustrates agreement in the NP and in the ADJP in GPA. As discussed in Section 4.1.5.2 above, the adjective in GA agrees with the noun in definiteness, gender, and number. In addition, the demonstrative pronouns in GA, which inflect for proximity, agree with the noun in number and gender. Unlike the noun in GA, the noun in GPA has some inconsistency in its agreement with the numeral in number, gender, and definiteness (Almoaily, 2012 and Badawi, 1996). This section demonstrates agreement in the NP and in the ADJP in GPA. The following subsections illustrate agreement between noun-adjective and agreement between numerals and quantified nouns in GP.

**I- Noun-Adjective Agreement in Gulf Pidgin Arabic**

In GPA, the noun-adjective does not agree with definiteness, gender, and number (Almoaily, 2008, 2012; Smart, 1990; Roberts and Bresnan, 2008). Examine the example below from my fieldwork data:

(50) yabɣi ʕa'lah kabeer (M5)
want family.SG.F big. SG.M
'I want to make a big family'

In this excerpt, M5 uses the masculine form of the adjective with the feminine noun *family*. In GA, the adjective form *kabeer* in (50) would be *kabeer-ah* ('big-SG.F').

**II- Agreement Between the Noun or Adjective and the Numeral in Gulf Pidgin Arabic**

As shown in Section 4.1.5.2, in GA the noun or adjective following a number can be singular, dual, or plural; feminine or masculine; and definite or indefinite. In contrast, GPA speakers always use the indefinite, singular, masculine form, regardless of the number (specifically for numbers from one to ten). In the following excerpt, the singular form of the adjective *kamel* ('complete') is used with the feminine form of the number *sanah* ('one year'):

(51) yaʕni mumken sanah kamel (Su 2)
means probably year.SG.F complete.SG.M
'It means I probably have one complete year'

Sentence (51) could be translated to GA as follows:

(52) ana le tagriban sanah kaml-ah
I have approximately year.SG.F complete-SG.F

**III- Demonstratives in Gulf Pidgin Arabic**

The use of demonstrative pronouns in GPA is limited and inconsistent with the use of such pronouns in GA (Al-Ageel, 2015; Al-Azraqi, 2010; Almoaily, 2012; Johnstone, 1967; Naess, 2008; Smart, 1990). Demonstratives in GPA do not agree with nouns or with adjectives in number and gender, nor do they inflect for proximity (Almoaily, 2012; Al-Azraqi, 2010; Smart, 1999). My fieldwork data suggests that the existence of a single demonstrative[3] *haða* ('this') is used with all GPA nouns and adjectives regardless of their gender, number, or proximity. To investigate the use of GA demonstrative pronouns, during interviews[4] I asked my informants to use the demonstrative with objects of different genders, numbers, and proximity. Indeed, most of the informants used the demonstrative *hatha* (meaning 'this', the singular masculine proximal demonstrative) with all objects, irrespective of their gender, number, or proximity. Next, Section 4.3 provides the conclusions of this chapter.

**4.3. Conclusion to Chapter 4**

This chapter delivered a descriptive account of the five morphosyntactic features of GA, which is the superstrate language of GPA, and GPA, which is the pidgin under investigation. All the examples are thus from my fieldwork data, which was collected especially to test the first hypothesis that there are differences between standard GA and GPA. A qualitative comparison was conducted in this chapter to determine whether differences exist between standard GA and GPA.

It has been shown that inflection is heavily used in GA morphosyntax. The following three of the five morphosyntactic features investigated in this study are consistently expressed as bound morphemes: the definiteness marker *al*, the agreement markers, and the object and possessive

[3] Singular masculine proximate demonstrative.
[4] My interviews contained a direct elicitation exercise for demonstratives (refer to Appendix A).

pronouns. At the same time, inflections in the other two features – coordination and copular verbs – still exist and are expressed by free morphemes.

As regards the description of the GPA system, the use of inflections was less frequent compared to GA through adopting a reduced agreement system (as demonstrated in Section 4.2.5.1). Furthermore, the discussion in Section 4.2.4 indicates that GPA is more analytic: The GA object and possessive pronouns are bound morphemes, whereas they are free morphemes in GPA. Next, Chapter 5 delivers a definition of each substrate language of GPA and a descriptive account of the relevant morphosyntactic features, which are limited to the same morphosyntactic features discussed in this chapter.

# Chapter 5: Substrate Languages of Gulf Pidgin Arabic

All the varieties reported from the Gulf region have emerged in a multilingual environment where several substrate languages are involved, from both South Asia (Urdu, Bengali, Pashto, Punjabi, Sinhala, Malayalam, and Tamil), and South East Asia (mainly Tagalog and Indonesian). Since this thesis aims to discover language variation in males and females' GPA speech resulting from different morphosyntactic structures of the substrate languages and from length of exposure to GA, I use Almoaily's (2012) corpus as a reference for male GPA speech and compare it with the corpus of the female GPA speech which I built in SA. Therefore, this chapter aims to describe the six main substrate languages of GPA. It starts with a review of the substrate languages and discusses the methodology of determining the languages with the largest number of female speakers in the Gulf region. Then, I review the morphosyntactic features relevant to this project (i.e. agreement, pronouns, definiteness/indefiniteness, coordination, and copula). This chapter concludes with a summary of the main results of the comparison between GA and GPA, along with a discussion of how the similarities and differences between the substrate languages may possibly play a role in the variation encountered between the speakers of GPA.

## 5.1. Substrate Languages of Gulf Pidgin Arabic

To determine the countries from which the majority of temporary immigrant workers originate, Almoaily (2008) adopted a pilot study which revealed that the three largest language male groups are Bengali, Malayalam, and Punjabi. In addition, several substrate languages are involved, among which are Tagalog, Pashtu, Sinhala, and Indonesian. It seems impossible to test all the substrate languages of GPA as potential sources of linguistic variation in female GPA. Therefore, I follow Almoaily's methodology to determine whether these three substrate

language groups in male GPA are the same largest substrate language groups in female GPA (please refer to Section 6.6.3.1 for more details on the problem in determining the most common substrate languages of GPA). Regarding the migration profiles in SA (2013), the migrant stock by origin reveals that the top five countries or areas of origin are India, Pakistan, Bangladesh, Egypt, and the Philippines, as indicated in Table 5.1.

| **Top five countries or areas of origin** | **Total** |
|---|---|
| India | 1 761 857 |
| Pakistan | 1 319 607 |
| Bangladesh | 1 309 004 |
| Egypt | 1 298 388 |
| Philippines | 1 028 802 |
| **Total** | **6 717 658** |

Table 5.1: Migrant stock by origin (2013).

Based on these numbers, India, Pakistan, and Bangladesh are the largest substrate language groups in SA. However, it is worth noting that the immigration policies for FWs in SA state that all FWs should work under a 2-year extendable permit, which means that the total number of speakers for every substrate language in SA is constantly changing (refer to Section 2.5). I faced this problem, as did Almoaily in his study. Therefore, the approach in determining the largest number of language groups was to ask the GPA female speakers themselves. According to the migration profiles in SA (2013), the total number of male expatriates is generally higher than the total number of female workers, as illustrated in Figure 5.1, although the international migrant stock only offers statistics for the number of immigrants without any reference to the countries they come from or the languages they speak.

| | 1990 | | | 2000 | | | 2013 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | Total | Male | Female | Total | Male | Female | Total |
| 0-4 | 239 537 | 240 924 | 480 461 | 269 804 | 258 549 | 528 353 | 339 913 | 273 842 | 613 755 |
| 5-9 | 216 424 | 220 600 | 437 024 | 242 922 | 237 664 | 480 586 | 327 889 | 262 937 | 590 826 |
| 10-14 | 152 082 | 157 520 | 309 602 | 194 588 | 193 595 | 388 183 | 299 193 | 239 562 | 538 755 |
| 15-19 | 104 604 | 88 460 | 193 064 | 139 656 | 117 293 | 256 949 | 288 161 | 216 089 | 504 250 |
| 20-24 | 244 347 | 133 874 | 378 221 | 198 993 | 112 869 | 311 862 | 453 069 | 262 558 | 715 627 |
| 25-29 | 556 309 | 217 136 | 773 445 | 430 609 | 165 076 | 595 685 | 851 555 | 354 955 | 1 206 510 |
| 30-34 | 645 660 | 240 181 | 885 841 | 597 965 | 205 331 | 803 296 | 1 031 747 | 335 880 | 1 367 627 |
| 35-39 | 494 005 | 170 680 | 664 685 | 547 607 | 186 562 | 734 169 | 963 239 | 263 170 | 1 226 409 |
| 40-44 | 317 556 | 89 814 | 407 370 | 392 111 | 117 516 | 509 627 | 755 978 | 164 679 | 920 657 |
| 45-49 | 172 113 | 45 643 | 217 756 | 240 732 | 63 221 | 303 953 | 516 770 | 97 158 | 613 928 |
| 50-54 | 96 595 | 26 705 | 123 300 | 134 342 | 35 448 | 169 790 | 311 360 | 59 403 | 370 763 |
| 55-59 | 41 105 | 13 968 | 55 073 | 70 145 | 21 243 | 91 388 | 157 756 | 36 189 | 193 945 |
| 60-64 | 22 967 | 11 754 | 34 721 | 33 035 | 14 849 | 47 884 | 85 441 | 29 421 | 114 862 |
| 65+ | 21 569 | 16 313 | 37 882 | 25 220 | 16 443 | 41 663 | 55 199 | 27 320 | 82 519 |
| Total | 3 324 873 | 1 673 572 | 4 998 445 | 3 517 729 | 1 745 659 | 5 263 388 | 6 437 270 | 2 623 163 | 9 060 433 |

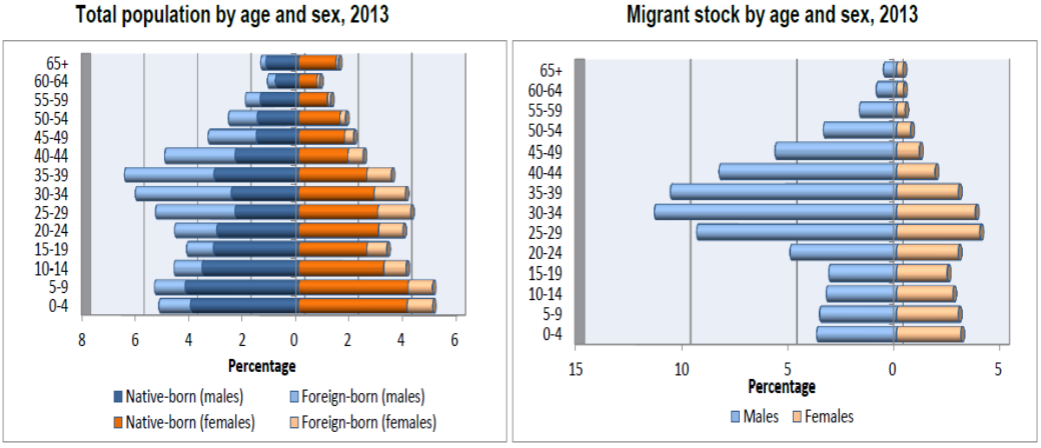



Figure 5.1: International migrant stock by age and sex in Saudi Arabia, adopted from migration profiles in Saudi Arabia (2013)[1].

Figure 5.1 above illustrates that the number of male migrants grew significantly more than the number of female migrants. For example, in 2013, 75% of the total number of migrants aged 30–34 were male, while just 25% were female. This was fuelled in part by the strong demand for male migrant workers in the oil-producing countries of Western Asia, such as the Gulf countries. According to the report of the International Migration Report 2015 (refer to Section 2.8), women comprise slightly less than half of all international migrants worldwide. Female migrants outnumber male migrants in Europe and Northern America, while in Africa and Asia, particularly Western Asia, migrants are predominantly men. The proportion of female migrants fell from 49% in 2000 to 48% in 2015. The reason for this decline lies primarily in the high

[1] Retrieved on 17 May 2017 from https://esa.un.org/miggmgprofiles/indicators/files/SaudiArabia.pdf

concentration of males among recent migrant inflows. According to International Migrant Stock (2019)[2], SA hosted the second largest number of migrants worldwide (around 13 million). In 2015, SA had the highest number of international female migrants workers with 32% of total migrints population as reported by the International Migration Report 2015 (see Table 2.1).

The next section describes the methodology which I followed to determine the largest substrate language groups of GPA.

### 5.2. Determining the Most Common, Largest Language Groups of Gulf Pidgin Arabic

Based on the general population census in SA (2010), the largest six nationality groups with GPA speakers in SA are from India, Pakistan, the Philippines, Bangladesh, Sri Lanka, and Indonesia. This is illustrated in Figure 5.2.

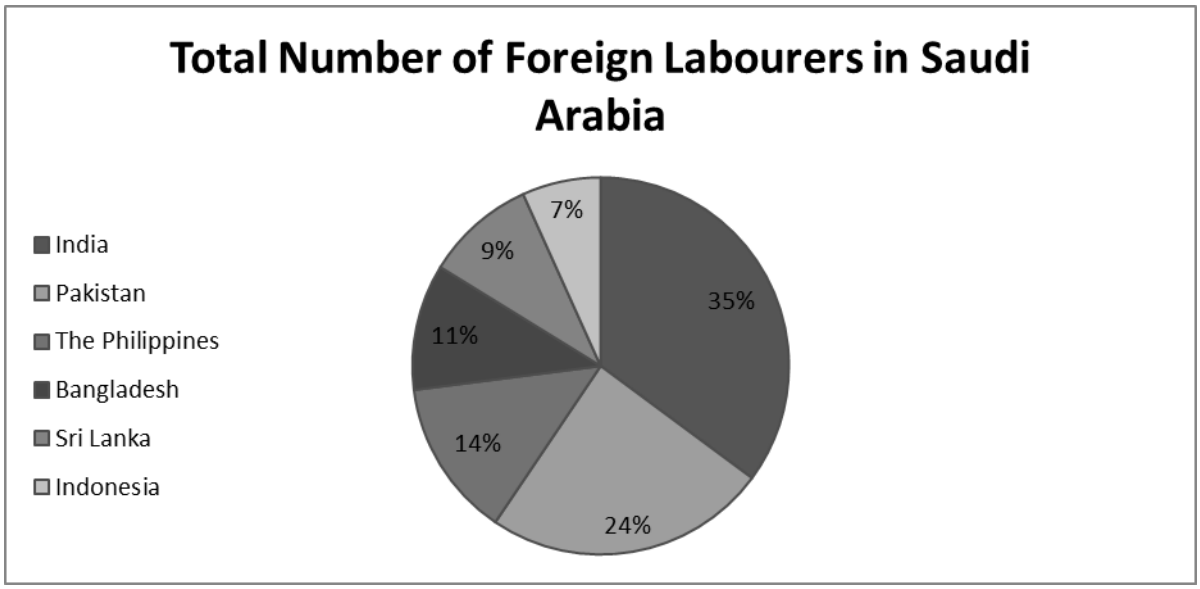



Figure 5.2: The estimated number of foreign labourers in Saudi Arabia based on their nationalities in 2010.

[2] Retrieved on 10 April 2019 from https://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationStock2019_TenKeyFindings.pdf

Thus, I conducted a pilot study with some female expatriates (among whom the majority of temporary immigrant workers come from India, Pakistan, and Bangladesh) in the form of short oral interviews. I met them in their workplaces (e.g. beauty salons, homes, hospitals, etc.) in SA to describe the three aforementioned languages regarding the morphosyntactic features that are the focus of this study. I asked them the following questions:

- From which city/state do you think the majority of people from your country come to SA?
- What language do they speak as a mother tongue?
- Do you speak that language?

Based on the findings of this pilot study and as a native – since childhood, I have always been in contact with people from the large migrant labour force – it could be supposed that the most common, largest language groups which GPA female speakers in SA speak as an L1 are Malayalam, Bengali, Tagalog, Sinhala, Punjabi, and Sunda (also refer to Figure 5.2). The next section describes the largest three language groups in Almoaily's (2012) study (Malayalam, Bengali, and Punjabi). In addition to these three languages, I describe the other three largest language groups that GPA female speakers in SA speak as an L1 (Sinhala, Tagalog, and Sunda). Based on gender variation analysis, I assess the potential influence of the length of stay on informants' choices as regards the studied morphosyntactic features by comparing GPA males and females' linguistic production.

### 5.3. Description of the Six Substrate Languages of Gulf Pidgin Arabic

This section provides a short introduction to each language under investigation. It starts with a brief sociolinguistic, typological, and geographic account of each language, and it ends with a summary table containing each morphosyntactic feature relevant to this project (i.e. agreement, pronouns, definiteness/indefiniteness, coordination, and copula).

### 5.3.1. Bengali

The Bengali language, Bengali Bangla, is a member of the Indo-Aryan language group from the Indo-Iranian branch of the Indo- European language family (Dasgupta, 2003; Lewis, 2009). It is spoken as an L1 or L2 by more than 210 million people in South Asian countries such as Bangladesh, India, and Nepal, among others (see Map 3 in Appendix D). There are two standard styles in Bengali: Sadhubhasa (elegant or genteel speech) and Chaltibhasa or Cholit Bangla (current or colloquial speech). The former is an old form that is used rarely, and only in formal settings, and it became standardised as the literary language; Sadhubhasa was not the language of daily communication. Chaltibhasa is the dominant literary language and has become a standard colloquial form of speech among the educated (Encyclopaedia Britannica, 2009). The pronouns and verb forms of the Sadhubhasa are contracted in Chaltibhasa, and there is also a marked difference in vocabulary. Bengali follows subject–object–verb (SOV) word order with a split-ergative case and agreement system (Almoaily, 2012). It is worth noting that the Bengali language does not have gender-specific pronouns, but it does have gender-specific terms, such as *abhineta* (actor) and *abhinetri* (actress), and *Baba*/*pita* (father) and *Maa*/*mata* (mother). Verbs do not change based on the gender of the person, as they do in Hindi. A few of the adjectives, however, do change based on gender, for example খুশি মহিলা *aahladita* (delighted female) and প্রসন্ন পুরুষ *aahladito* (delighted male). Examples of nouns with gendered differences are দূর্ভাগ্যজনক/দুর্ভাগ্য নারী *bhagyoheena* (unfortunate/unlucky female) and দুর্ভাগা/দুর্ভাগ্য পুরুষ; *bhagyoheen* (unfortunate/unlucky male), পুরুষ ছাত্র *chhatro* (male student) and মহিলা ছাত্র *chhatri* (female student), and বৃদ্ধ মানুষ *buro* (old man) and বৃদ্ধা *buri* (old woman). For the purposes of the current study, I am interested in the Modern Standard form of Bengali, Chaltibhasa, because any possible effect on the Bengali informants' production of GPA will

likely be due to Chaltibhasa rather than Sadhubhasa. Below is a brief description of the morphological features in Bengali (based on Ray, 1966 and Nasrin and Wurff, 2015).

#### *5.3.1.1. Agreement in Bengali*

**1. Verbal Agreement in Bengali**

Verbs in Bengali do not inflect for number and gender despite the large number of affixes. The roots of verbs are mono- or disyllabic, and the suffix *-a* is added to nouns to form verbs. Moreover, markers for aspect/tense follow the stem and precede the personal inflections. The four tenses (continuous, future, past, and perfect), apart from the present tense which is formed by attaching the personal endings directly to the verb stem (the verbal noun *minus -a*), are formed by adding one tense/aspect marker to the stem: (***c***)*ch* for the present continuous, *b* for the future, *l* for the simple past, and *t* for the past habitual-past conditional. Furthermore, three tenses are formed by a combination of two or more tense–aspect markers: the past continuous is marked by (***c***)*ch* + *l*, the present perfect by *e* (a completive marker) + *ch*, and the past perfect by ***e*** + *ch* + *l*. For example, the conjugation[3] of *jan-* ('know') is presented in Figure 5.3.

| | *present* | *pres. cont.* | *future* | *simple past* | *past habitual/cond.* |
|---|---|---|---|---|---|
| 1st | jani | janchi | janbo | janlam | jantam |
| 2nd inferior | janis | janchis | janbi | janli | janti |
| 2nd familiar | jano | jancho | janbe | janle | jante |
| 3rd familiar | jane | janche | janbe | janlo | janto |
| 2nd/3nd polite | janen | janchen | janben | janlen | janten |

| | *past continuous* | *pres. perfect* | *past perfect* |
|---|---|---|---|
| 1st | janchilam | jenechi | jenechilam |
| 2nd inferior | janchili | jenechis | jenechili |
| 2nd familiar | janchile | jenecho | jenechile |
| 3rd familiar | janchilo | jeneche | jenechilo |
| 2nd/3nd polite | janchilen | jenechen | jenechilen |

Figure 5.3: The conjugation of *jan-* (from Guntman and Avanzati, 2013).

[3] Black font: simple stem, orange: progressive marker, green: future marker, brown: past marker, red: habitual-counterfactual marker, light blue: completive marker, blue: personal markers.

## 2. Nominal Agreement in Bengali

The word order in Bengali is a fairly strict SOV structure, due to its relatively poor case system (Guntman and Avanzati, 2013). Nouns in Bengali inflect for number, size, and proximity, but not for gender or animacy, while attributive adjectives are not inflected for case, gender, or number. Appropriate suffixes are added to nouns to mark for plurality, definiteness, and case. However, suffixes attached to nouns are also attached to adjectives only to distinguish between them on a semantic level (Ray, 1966).

- **Gender**: Like other Eastern Indo-Aryan languages, Bengali has no grammatical gender but distinguishes sex.
- **Number**: *Singular* and *plural* exist. The plural forms are found in animate and definite inanimate nouns, as well as pronouns. In the case of pluralisation, only countable inanimate nouns may be pluralised, while uncountable nouns are not usually marked for the plural. The two plural suffixes are described by Thompson (2012) as follows:
    - *-(e)ra* is for animate nouns only, definite or indefinite
    - *-gulo/-guli/-gula* is for animates and definite inanimates
    - *chele* ('boy') → *chelera* ('boys/the boys') → *chelegulo* ('boys/the boys')
    - *juto* ('shoe') → *jutogulo* ('the shoes')

### *5.3.1.2. Definiteness in Bengali*

According to Ray (1966), definiteness in Bengali is expressed indirectly by using the suffix *-ṭa/-ṭi* (e.g. *cheleṭi* 'the boy') discussed above. This suffix is also employed when a noun is preceded by a possessive adjective or a demonstrative[4].

[4] There are 10 forms of demonstratives in Bengali that inflect for number (e.g. *-ta* 'that one', *-to* 'those two', *-tin* 'those four', and *-car* 'those four') and size (e.g. *-ti* 'that small'), but not for gender (Ray, 1966).

### *5.3.1.3. Pronouns in Bengali*

Bengali is a subject pro-drop language; therefore, according to Ferguson (1991), subject pronouns are optional. Bengali pronouns (see Figure 5.4) inflect for person, politeness, and number, but not for gender. Personal, demonstrative, interrogative, relative, and indefinite pronouns are free morphemes as detailed below:

- Personal pronouns are genderless but distinguish between human and non-human entities and have several degrees of status (familiar and polite forms): three in the second person and two in the third person.
- Demonstrative pronouns distinguish three deictic degrees (e.g. proximate *ei/e*, distal *oi/o*, and invisible *śei/śe*).
- The interrogative pronouns are *ke* (‘who?’) and *ki* (‘what?’). Other interrogatives are *kothae* (‘where?’), *kæmon* (‘how?’), *kon* (‘which?’), *kæno* (‘why?’), and *kɔto* (‘how much? / how many?’).
- The relative pronoun is *ja*.

| | singular | plural |
|---|---|---|
| 1 | ami | amra |
| 2 inferior | tui | tora |
| 2 familiar | tumi | tomra |
| 2 polite | apni | apnara |
| 3 familiar | śe | tara |
| 3 polite | tini | tãra |

Figure 5.4: Bengali pronouns (from Guntman and Avanzati, 2013).

### *5.3.1.4. Coordinating Conjunctions in Bengali*

Conjunctions in Bengali normally lack an overt coordinator, as in (1). However, in some cases, coordinator is expressed by having some overt linking devices such as conjunctions: *ba* (‘or’) and *ar* (‘and’), as in (2). For example,

(1) Aaoka Aatmaja Aadhila

‘Aaoka, Aatmaja, and Aadhila or ‘Aaoka Aatmaja, or Aadhila’

(2) Aaoka Aatmaja ***ba*** Aadhila

Aaoka Aatmaja or Aadhila

‘Aaoka, Aatmaja, or Aadhila’

Similarly to English, the conjunction marker is placed before the final coordinated element.

#### *5.3.1.5. Copula in Bengali*

According to Ray (1966), the copula in Bengali is not usually expressed in the affirmative present tense. However, it can be overt in some cases (Finch, 2001). He adds that the copula may appear as an overt verb (*achh*) in some cases, and it may be covert (true copula) in others depending on its subsequent element. Furthermore, Finch noticed that ‘the distribution of the copula in Bengali obeys the following generalisation: the overt copula indicates the stage level[5] reading; the zero copula favours the individual level reading, but allows for the stage level reading too’ (Tatevosov, 2005:186). For stage-level predicates, Bangla uses the *-ach* copula, while for individual-level predicates, the common zero copula is favoured (Ghosh, 2009).

### 5.3.2. Punjabi

An alternative name for Punjabi is ‘Panjabi’, which originally meant ‘belonging to the [land of the] five rivers’, in reference to the prominent, wealthy people who lived beside the Indus river. Punjabi is classified as a member of the Indo-Aryan subgroup of the Indo-European family of languages. Punjabi is among the 10 largest languages of the world, spoken in the UK, Canada,

[5] Carlson and Pelletier (1995) and Diesing (1992) classify predicates into stage-level and individual-level forms in standard literature of semantics. The predicate which refers to a temporary state of the individual is known as a stage-level predicate, and the predicate that refers to some permanent property of the individuals is called an individual-level predicate (Ghosh, 2009).

Malaysia, SA, Australia, the UAE, the US, Singapore, and Kenya (Rodriguez, 2010). According to the Ethnologue 2017[6] estimate, approximately 99 million people worldwide speak Punjabi. Shackle (1970) suggested that Punjabi is believed to have been influenced by many languages, including English, Urdu, Persian, and Arabic, pointing out that there are three languages, namely, English, Punjabi, and Urdu, that are concurrently spoken in Lahore, the capital of the province of Punjab. These three languages have different statuses: English and Urdu are the languages used in formal settings, while Urdu is widely used in public speech. Moreover, English is used amongst highly educated people in the state of Lahore. In fact, Punjabi has the lowest rank of the three; it is typically the language of uneducated people and is used by lower-class and lower-middle-class family members. (Almoaily, 2012). Punjabi has many different dialects, including Punjabi Proper, Majhi, Doab, Bhatyiana, Powadhi, Malwa, Bathi, Western Panjabi, and Eastern Panjabi (Rodriguez, 2010). Th language is written in both the Shahmukhi and the Gurmukhi scripts, the former mainly by Muslims and the latter mainly by Sikhs and Hindus. One of the linguistic features of the Punjabi language is its SOV word order (Wikipedia, 2011)[7]. It is also a gendered language, although its pronouns do not distinguish gender. Punjabi learners of English often treat prepositions as nouns, reanalysing English relational terms as names of locations on the pattern of the mother tongue and producing forms such as 'Put the down chair' (Perdue, 1993: 246). Some other examples of this type of error are 'in upstairs', 'I live with enjoy', and 'It's belong to me'.

When speaking to people older than oneself, the style is mostly indirect. However, when speaking to people of one's own age or younger, the style is usually direct.

- Direct eye contact is usually the norm between members of the same gender and age.

---

[6] Retrieved on 05 February 2018 from https://www.ethnologue.com

[7] Punjabi Grammar. (2011). In Wikipedia. Retrieved on February 05 2018 from http://en.wikipedia.org/wiki/Punjabi_grammar

- Indirect eye contact is usually the norm when speaking to elders and members of the opposite gender.
- An arm's length of personal space is common when speaking to members of the opposite gender. This distance tends to be shorter with members of the same gender.

#### *5.3.2.1. Agreement in Punjabi*

**1. Verbal Agreement in Punjabi**

In Punjabi, most finite verbs are a combination of aspect (imperfective and perfective) and tense, using the auxiliary 'to be' plus a non-finite form of the lexical verb, and they produce 12 aspectual tenses (Guntman and Avanzati, 2013). Apart from these compound tenses, Punjabi has two non-tense forms (unspecified habitual-contrafactual and unspecified perfective) and two non-aspectual forms (contingent future and definite future). For example, the list of Punjabi finite forms of the verb *āuṇā* ('to come') in the first-person singular are presented in Figure 5.5.

| | *masculine* | *feminine* | |
|---|---|---|---|
| *present habitual* | āundā hã̃ | āundī hã̃ | I come |
| *present continuous* | ā riˈā hã̃ | ā raˈī hã̃ | I am coming |
| *past habitual* | āundā sã̃ | āundī sã̃ | I used to come |
| *past continuous* | ā riˈā sã̃ | ā raˈī sã̃ | I was coming |
| *presumptive habitual* | āundā hovã̃ gā | āundī hovã̃ gī | I probably come |
| *subjunctive habitual* | āundā hovã̃ | āundī hovã̃ | (perhaps) I come |
| *contrafactual habitual* | āundā hundā | āundī hundī | had I come (regularly) |
| *present perfective* | āīā hã̃ | āī hã̃ | I have come |
| *past perfective* | āīā sã̃ | āī sã̃ | I had come |
| *presumptive perfective* | āiā hovã̃ gā | āī hovã̃ gī | I must have come |
| *subjunctive perfective* | āiā hovã̃ | āī hovã̃ | (if) I have come |
| *contrafactual perfective* | āiā hundā | āī hundī | (if) I had come |
| *unspecified habitual-contraf.* | āundā | āundī | I would come; had I come |
| *unspecified perfective* | ãiã̃ | ãĩ̃ | I came |
| *contingent future* | āvã̃ | āvã̃ | (that) I may come |
| *definite future* | āvã̃ gā | āvã̃ gī | I shall come |

Figure 5.5: Punjabi finite forms of the verb *āuṇā*, meaning 'to come' (from Guntman and Avanzati, 2013).

## 2. Nominal Agreement in Punjabi

According to Tolstaja (1981), Punjabi nouns and some adjectives are inflected for case (direct, oblique, vocative), gender (masculine, feminine), and number (singular, plural). There are two declension types for masculine nouns – one ends in *ā,* and the other is in another letter – while feminine nouns have just one declension type. Attributive adjectives precede nouns, and once they are used predicatively, they are positioned before the verb. Some masculine adjectives ending in *ā* and feminine adjectives ending in *ī* are inflected, while others are not, as depicted in Figure 5.6.

| | masc. | fem. |
|---|---|---|
| dir. sg. | -ā | -ī |
| obl. sg. | -e | -ī |
| dir. pl. | -e | -iã̄ |
| obl. pl. | -e, -iã̄ | -iã̄ |

Figure 5.6: Punjabi inflected adjectives (from Guntman and Avanzati, 2013).

Furthermore, inflected adjectives agree with their noun in case, gender, and number. For example, the adjective *cangaa* ('good') may take the following forms: *cangaa*-SGM, *cange*-PLM, *cangii*-SGF, and *cangiaa*-PLF.

#### *5.3.2.2. Definiteness in Punjabi*

There are no markers for definiteness/indefiniteness in Punjabi: '[T]he concept of definiteness and indefiniteness is expressed indirectly by means of pronouns and the numeral *ikk* "one"' (Bhatia, 2013: 99). Definiteness in Punjabi is similar to Bengali; it can be expressed indirectly by using demonstratives such as ***é*** /ih/ (in close objects, i.e. this/these) and ***ó*** /uh/ (in distant objects, i.e. that/those) and can retain the same form for both the numbers.

#### *5.3.2.3. Pronouns in Punjabi*

In Punjabi, there are six types of pronouns, and they are all free morphemes. For the purpose of this study, I briefly discuss the most important pronouns that have distinct marker contrasts in the pronominal system of this stock language (see Malik, 1995; Bhatia 2013; Dasgupta, 2007).

**(i) Subject (Personal) Pronouns:** They inflect for person and number and are used to refer to the first, second, and third person in sentences: ਮੈਂ *maiṃ* ('I') is used for the singular first person, and ਅਸੀਂ *asīṃ* ('we') is used for the plural; ਤੂੰ *tūṃ* ('you') is used for the second singular person, and ਤੁਸੀਂ *tusīṃ* ('you') is used for the plural. For the third person, demonstrative pronouns are ਉਹ ***ó*** /uh/ ('that') and ਇਹ ***é*** /ih/ ('this'), and they retain the same form for both

the numbers. All personal pronouns can equally be used for both genders, except the genitive case forms. According to Brown and Ogilvie (2009), subject pronouns in Punjabi are optional in sentences. Thus, Punjabi is a pro-drop language.

**(ii) Object pronouns**: They take the same form for the first and second person of the subject pronouns above. However, the only distinguishable form is the third-person object pronouns: ***us*** for the singular third person and ***is*** for the plural.

**(iii) Possessive pronouns**: They take a different set of possessive pronouns which inflect for person, number, and gender. Examples include *mer-aa* ('my-M'), *mer-e* ('our-M'), *mer-ii* ('my-F'), *mer-iaa* ('our-F'), *teraa* ('your.MSG'), *terii* ('your.MPL'), and *uhdaa* ('his').

### *5.3.2.4. Coordinating Conjunctions in Punjabi*

The following key conjunction markers, among others, are used in Punjabi: ਅਤੇ *atē* ('and'), ਜਾਂ *jāṃ* ('or'), ਕਿ *ki* ('that'). They can be classified into two categories[8] based on their uses, as outlined below:

- **Coordinate conjunctions:** ਅਤੇ *atē/tē* ('and') and ਜਾਂ *jāṃ* ('or') act as coordinate conjunctions in the following examples:

(3) muṇḍā atē kuḍaī āē sī.

'The boy and girl came'

(4) pañchī gīt gā rahē han tē ṭhaṇḍhī havā vag rahī hai.

'The birds are singing and a cool breeze is blowing'

- **Subordinate conjunctions:** ਜੇ *jē* ('if'), ਉਂਞ *uññ* ('by'), and ਤਾਂ *tāṃ* ('then') are examples of subordinate conjunctions.

[8] Retrieved on 10 December 2016 from http://punjabi.aglsoft.com/punjabi/learngrammar/?show=conjunction

#### *5.3.2.5. Copula in Punjabi*

A copula is used in positive sentences and dropped in negative ones (Bhatia, 2013). The copula agrees with the subject in number, just as it does in English. For example, the copula *ai* is used when the subject is singular, while the copula *ne* is used when the subject is plural:

(5) muṇḍā sōhṇā hai (*ai*)
Boy-M.SG handsome-M.SG is
'The boy is handsome'

(6) kuṛīāṃ sōhṇīāṃ han (*ne*)
Girls-F.PL beautiful- F.PL are
'The girls are beautiful'

### 5.3.3. Malayalam

Malayalam is an official language spoken by the native people of southwestern India. It belongs to the Dravidian family of languages spoken in the Indian state of Kerala[9] (see Map 2 in Appendix D). Malayalam is believed to be the mother tongue of more than 38 million people, as claimed by the Official Language (Legislative) Commission (2015). The Malayalam language is alternatively called Alealum, Malayalani, Malayali, Malean, Maliyad, and Mallealle (Ethnologue, 2012). Caldwell (1875) notes that Malayalam languages branched from Classical Tamil[10] and over time gained a large amount of Sanskrit vocabulary and lost the personal terminations of verbs. According to Subramoniam (1997), dialects of Malayalam are distinguishable at regional and social levels. On the regional level, dialects of Malayalam can

[9] A state in South India on the Malabar Coast. Malayalam is the most widely spoken language and is also the official language of the state.
[10] Tamil is a Dravidian language predominantly spoken by the Tamil people of India and Sri Lanka. Tamil is an official language of two countries: Sri Lanka and Singapore ("Kodagu-Kerala association is ancient". The Hindu. Chennai, India. 26 November 2008 and "Virajpet Kannada Sahitya Sammelan on January 19". The Hindu. Chennai, India. 9 December 2008.)

be divided into 13 dialect areas. Such divergence among the dialects of Malayalam can be seen in almost all aspects of language, including phonetics, phonology, vocabulary, and grammar. The difference between any two given dialects of Malayalam can be counted in terms of the presence or absence of specific units at each level of the language. According to Pillai (1965), Malayalam has a noticeably high number of Sanskrit loanwords, but they are rarely used. Loanwords and influences also come from Hebrew into Judeo-Malayalam, English, and Portuguese, and into Mappilah, Syriac, and Greek in the Christian dialects, while Arabic and Persian elements predominate in the Muslim dialects. As regards word order, SOV is predominant in Malayalam, as in other Dravidian languages, although a rare object–subject–verb word order occurs in interrogative clauses when the subject is the interrogative word (Jayaseelan, 2001). As in English, both adjectives and possessive adjectives precede the nouns they modify.

Malayalam has six grammatical cases, according to Asher and Kumari (1997), or seven of them, as reported by Aparnna, Raja, and Soman (2010). Unlike other Dravidian languages, in Malayalam verbs are conjugated only for TMA, but not for person, gender, or number (Asher and Kumari, 1997). Moreover, tense is expressed through suffixes attached to the final verb.

#### *5.3.3.1. Agreement in Malayalam*

1. **Verbal Agreement in Malayalam**

Malayalam is unusual among Dravidian languages in that it has no agreement between the subject and the finite verb in person and number, but conjugates only for TMA (Asher and Kumari, 1997). Malayalam is an agglutinative language adding suffixes to nominal and verbal stems to mark grammatical categories. Furthermore, tense is expressed via suffixes (tense markers) attached to the root or derived stem. According to Rajaraja and Roy (1999), to form the present, the suffix *-unnu* is added to the stem, while the suffixes *-um* and occasionally ***-ū*** are marked to form the future tense, and the past tense markers are ***-i*** or ***-u***. For example, the

verb *po* (‘go’) is: *poy**i*** (past), *pokunnu* (present), and *pokum* (future). Further examples of past conjugations are *pāṭ-**i*** (‘sang’), *iru-nnu* (‘sat’), *kara-ñña* (‘wept’), *etu-ttu* (‘took’), and *kaṅ-ṭu* (‘saw’) based on the number of consonants preceding the tense markers. For the non-finite forms (infinitive and participles), the infinitive is formed by adding ***-(k)****kuka* or *-yuka* to the stem. for instance *varuka* (‘to come’) and *paṛayuka* (‘to speak’).

## 2. Nominal Agreement in Malayalam

As mentioned earlier, Malayalam is an agglutinative language which requires adding suffixes to nominal and verbal stems to mark grammatical categories. Nouns in Malayalam are marked for case, number, and natural gender, and there are few true adjectives in the language; attributive adjectives come before the noun and are invariable (Guntman and Avanzati, 2013). Gender (masculine, feminine, and neuter) in Malayalam is not marked on the noun, but it is indicated on third-person and demonstrative pronouns or in the form of a noun expressing natural gender. Regarding number (singular and plural), the Malayalam masculine and feminine nouns ending in *a* take the plural suffix *-mār*. Only feminine nouns ending in *i* take *-mār* or *-kaḷ*, while other feminine nouns take *-kaḷ*. Moreover, masculine or feminine nouns ending in *ā* or *ṛ* take *-kkaḷ*; only masculine nouns ending in *an* change it to *-ar*. However, most neuter nouns take *-kaḷ*, *-kkaḷ*, or *-ṅṅaḷ*, except if they are preceded by a quantifier when unmarked for plural, as shown in Figure 5.7.

| | | |
|---|---|---|
| addhyāpaka-n | addhyāpakanmār | teacher (masc.) |
| addhyāpika | addhyāpikamār | teacher (fem.) |
| sahōdari | sahōdarimār | sister |
| strī | strīkaḷ | woman |
| pitṛ | pitṛkkaḷ | ancestor (masc.) |
| manuṣyan | manuṣyar | man |
| āna | ānakaḷ | elephant (neuter) |
| pū | pūkkaḷ | flower (neuter) |
| maram | maraṅṅaḷ | tree (neuter) |

Figure 5.7: Suffixes added to Malayalam nouns to mark number (from Guntman and Avanzati, 2013).

#### *5.3.3.2. Definiteness in Malayalam*

In Malayalam, there are no definiteness markers; they could only express indirectly by using the demonstratives *ii* ('this') and *aa* ('that'). As regards indefiniteness, Malayalam uses markers which take the following forms, according to Nair (2012):

- The indefinite marker *oru* ('a') is only used with countable singular nouns. For example,

(7) oru kuttiye kayarraattatŭ

'One student enters'

- Malayalam quantifiers, such as *cila* ('few'), *pala* ('several'), and *kuracce* ('little'), are used with plurals and mass nouns to express indefiniteness. For example,

(8) cila kuttikal

'Some children'

(9) pala aalukal

'Many people'

#### *5.3.3.3. Pronouns in Malayalam*

Pronouns in Malayalam are always free morphemes which inflect for person, case, number, gender, politeness, and proximity (Nair, 2012). Subject pronouns for the third person are identical to the remote demonstrative pronouns and distinguish gender and number as follows:

| | |
|---|---|
| 1SG: ñān | 1PL.incl.: nām/nammaḷ |
| 2SG: nī | 1PL.excl.: ñaṅṅaḷ |
| 3SG.M: avan | 2PL.: niṅṅaḷ |
| 3SG.F: avaḷ | 3PL.M -F: avar(kaḷ) |
| 3SG.NT: atu | 3PL.NT: ava/atukaḷ |
| 1SG: ñān | 1PL.incl.: nām/nammaḷ |

Malayalam demonstrative pronouns are marked for proximation ('this') and distance ('that') as well as for gender and number. On the one hand, proximate demonstrative pronouns are formed by replacing the initial ***a*** of third-person pronouns by ***i***, for example *ivan* ('this.M'), *ivaḷ* ('this.F'), *itu* ('this.NT'), *ivar* ('these.M,F'), and *iva* ('these.NT'). Distance demonstrative pronouns, on the other hand, are identical to third-person pronouns.

#### *5.3.3.4. Coordination in Malayalam*

Nair (2012) indicated that suffixes such as the coordinating suffix–*um* or the disjunctive suffix ***-oo*** is used to connect two or more sentences into a single coordinate sentence. For instance,

(10) avar vaayikkukayum elutukayum ceytu
They read-INF-CONJ write-INF-CONJ do-PAST
'They read and wrote'

(11) avar vaayikkukayoo elutukayoo ceytu
they read-INF-DISJ write-INF-DISJ do-PAST
'They read or wrote'

#### *5.3.3.5. Copula in Malayalam*

In Malayalam, there are two copulas, *uNTE* and *aaNE,* which are used interchangeably in a large number of contexts. The copula *aaŋtə* is also widely used – more than the former copula and especially with nominal complements (Asher, 1968; Variar, 1979; Asher and Kumari, 1997). Mohanan and Mohanan (1999) stated that '[o]ne strong intuition is that uNTE and aaNE

correspond to the English verbs “have” and “be”; another is that they should be viewed as the “existential” and “equative” copulas respectively’. In some cases, the use of the copula is optional if, intuitively, the tense of the verb is clear from the context; if not, then the use of the copula is obligatory (Jayaseelan, 2010, cited in Almoaily, 2012). The following examples show the interchangeability of the copulas:

(12) aanakkE pani uNTE / illa
elephant-D fever-N BE-PR BE:NEG:PR
‘The elephant has / doesn’t have a fever’

(13) puucca tooTTattil uNTE / illa.
cat-N garden-L BE-PR BE:NEG:PR
‘The cat is/isn’t in the garden’

### 5.3.4. Sinhala[11]

Sinhalese belongs to the Indo-Aryan branch of Indo-European languages (DeVotta, 2004). Apart from Tamil, Sinhalese is one of the official languages of Sri Lanka[12] (Dharmadāsa, 1992); see Map 2 in Appendix D. It is the native language of the Sinhalese people, who, with a population of more than 20.3 million, make up the largest ethnic group in Sri Lanka, according to the Census of Population and Housing, 2012[13], and it is an L2 spoken by other ethnic groups in Sri Lanka, totalling approximately 4 million. According to Geiger (1995), Sinhalese has some different features that distinguish it from other Indo-Aryan languages: (1)

[11] Alternative names: Sinhala, Singhala, Singhalese, and Cingalese. Retrieved on 12 February 2018 from http://www.languagesgulper.com/eng/Sinhalese.html
[12] It is an island country in South Asia, located southeast of India.
[13] Census of Population and Housing, (2012). Department of Census and Statistics Ministry of Finance and Planning. Retrieved on 29 February 2018 from http://www.statistics.gov.lk/PopHouSat/CPH2011/Pages/Activities/Reports/CPH_2012_5Per_Rpt.pdf

the influence of the substratum, the parent stock of the Vedda language[14] – Sinhalese has its own words that are only found in that language or shared between Sinhalese and Vedda; (2) influences from neighbouring languages – apart from Tamil loanwords, several phonetic and grammatical features present in neighbouring Dravidian languages; and (3) foreign influence – modern Sinhalese contains some foreign loanwords, namely, from Portuguese, Dutch, and English, due to centuries of colonial rule.

Despite the differences between the dialects of Sinhalese, native speakers find all dialects to be mutually intelligible and unnoticeable, as reported by Geiger (1995). Due to regional variation, Sinhalese has a distinctive diglossia involving literary Sinhalese (e.g. literary texts, speeches, TV and radio news broadcasts, etc.), and spoken Sinhalese, which is the language of communication in everyday life (Gair, 1968). Both varieties differ from each other in many aspects; they have important morphological, syntactical, and lexical differences. Since the informants in the current study speak informally, for the sake of simplicity I talk about 'colloquial Sinhala', which is the spoken, and more simplified, version of Sinhala.

#### *5.3.4.1. Agreement in Sinhalese*

1. **Verbal Agreement in Sinhalese**

In written Sinhalese, the verb agrees with the subject in person, gender, and number, while in spoken Sinhalese, the verb inflects for tense (past and non-past, present, and future) but not for number, person, or gender. In other words, there is no subject-verb agreement (Gunasekara, 1891). According to Geiger (1968), Sinhalese has three basic tenses: present, past, and future. These basic tenses are combined to form compound tenses, such as present continuous, past continuous, present perfect, or past perfect, as Sinhalese has many compound verbs (e.g. noun-verb, verb-verb).

[14] It is an endangered language which was used by the indigenous Vedda people of Sri Lanka (Hammarström, Forkel, and Haspelmath, 2017).

In Sinhalese, the simple present generally may have a future or a progressive sense, and it is the same as in Colloquial Sinhalese where the simple present is the same as the infinitive (e.g. *enavā* 'come[s]/coming/will come'). Unlike the simple past, it is marked by a change in the verbal base, for example the verb *āvā* ('came'). The stem of the verb 'to come' changes from *e* to *āv* to *æv* to express different tenses. To mark the future, the suffix *–nnam* is used, and it is only marked for the verb in the first person (e.g. *ennam* 'I/we will come').

Moreover, Sinhalese has an SOV word order, which is common to most left-branching languages. Sinhalese is a head-final language, which means that the modifiers of the noun and the verb precede their heads. This is not applied for the numerals because they follow the head noun (Geiger, 1995). As a left-branching language, there are no prepositions, only postpositions. Unlike a preposition, a postposition comes after its complement. For instance, the NP 'under the book' translates in Sinhalese to පොත යට (*poṭə yaṭə*), literally 'book under'. The structure of a sentence is commonly a nominal rather than verbal one.

## 2. Nominal Agreement in Sinhalese

Nouns in Sinhala inflect for definiteness, number, case, and animacy. As all Indo-Aryan languages, adjectives head NPs and are indeclinable.

- **Gender**: Sinhalese is a gendered language; it has animate masculine and feminine and inanimate neuter. The observable distinction between masculine and feminine nouns is only found in written Sinhalese and in some spoken dialects. The masculine noun ends in *-a* or ***-ā***, and the feminine noun ends in *-ə* or *I*, while the neuter noun ends in *-ə* or *–ē* (Gair and Karunatilaka, 1974).
- **Number**: Sinhalese has singular and plural forms for masculine, feminine, and neuter nouns (Gair and Karunatilaka, 1974). Singular masculine nouns end in *-a/-ā* and mark their plurals in *-o* or *-u*, with possible gemination of the final consonant (e.g. *daruva* 'child' → *daruo* 'children'). Most feminine nouns end in ***ə*** and mark their plurals in ***-o*** or ***–u*** (e.g. *denə* 'cow'

→ *dennu* ‘cows’). Singular neuter nouns end in ***ə*** or ***ē***, and their plurals in -*val* – a nasal or involvement apocope[15] (e.g. *pārə* ‘way’ → *pārəval* ‘ways’). Regarding the plural of kinship nouns, pronouns and proper nouns, they are marked by the suffix –*lā* (e.g. *ammə* ‘mother’ → *ammələ̄* ‘mothers’).

- **Case**: Sinhalese has several cases: nominative, accusative, genitive, and dative, as indicated in Table 5.3. Some of them are less common, such as the instrumental. The exact number of these cases depends on the exact definition of cases one wishes to use.

| | 'man' (m) | | 'book' (n) | |
|---|---|---|---|---|
| | *singular* | *plural* | *singular* | *plural* |
| nom. | minihā | minissu | potə | pot |
| acc. | minihāvə | minissunvə | potə | pot |
| ins. | minihāgeṅ | minissungeṅ | potəṅ | potvəliṅ |
| gen. | minihāge | minissunge | potē | potvələ |
| dat. | minihāṭə | minissunṭə | potəṭə | poṭvələṭə |

Figure 5.3: Sinhalese cases (source: The Language Gulper)[16].

### *5.3.4.2. Definiteness in Sinhalese*

Suffixes are used to mark Sinhalese indefinite nouns. They are used only in the singular (-*ak/ek*). In the plural, (in)definiteness does not receive special marking; definiteness will be marked by the absence of such marking (Geiger, 1995). Moreover, masculine nouns usually take -*ek*, while feminine and inanimate ones take –*ak*, and they come before the case markers (e.g. *minihā* ‘the man’ and *minihek* ‘a man’).

### *5.3.4.3. Pronouns in Sinhalese*

Personal pronouns are used mostly for the first-person singular and plural (e.g. *mama-1SG*, *api-1PL*), although they are also used with the second-person form (e.g. *chē)* for polite address.

[15] Apocope in phonology is the loss (elision) of one or more sounds from the end of a word, especially the loss of an unstressed vowel (Wheeler, 2007).
[16] The Language Gulper http://www.languagesgulper.com/eng/Sinhalese.html
Accessed February 2018.

Dresser (2011) suggests that pronouns are not preferable for addressing someone; instead, friendship terms or titles are preferred, such as *mahattayā* (‘sir’), *nōnā* (‘madam’), *eyā* (‘he’), *ē gollo* (‘they’).

Demonstrative pronouns in Sinhala are not inflected for definiteness or number. There are four different ways in which to distinguish the demonstrative for proximate objects: this close object to ‘me’ (*mē*), that close object to ‘you’ (*oya*), that close object to a visible third person (*ara*), and that close object to an invisible third person (*ē*).

| **humans** | | **animals** | | **inanimates** | |
|---|---|---|---|---|---|
| *sing.* | *plural* | *sing.* | *plural* | *sing.* | *plural* |
| meyā | meyāla | mū | muṅ | mēkə | mēva |
| oyā | oyāla | ōka | okuṅ | ōkə | ōva |
| arəya | arəyəla | arū | aruṅ | arəkə | arəva |
| eyā | eyāla | ū | uṅ | ēkə | ēva |

Figure 5.9: Sinhalese demonstrative adjectives (source: The Language Gulper).

#### *5.3.4.4. Coordinating Conjunctions in Sinhalese*

According to MacDougall and De Abrew (1979), Sinhala conjunctions are particles which follow all nouns that need to be linked. Particles can take different shapes depending on the nouns they need to follow and whether they end with consonants or vowels. In general, there are almost no conjunctions equivalent to the English ‘that’ or ‘whether’, and there is no single word for ‘and’.

#### *5.3.4.5. Copula in Sinhalese*

As the focus in this study is on Colloquial Sinhala, Paolillo (2000) pointed out that Sinhala has no lexical copula in informal speech. For example,

(14) rajjuruwoo diga rñwulak ñti minihek

king.NOM long beard.IND.NOM having man.IND.NOM

‘The king was a long-bearded man’

(15) මම පොහොසත්

mamə poːsaʈ̪

literally ‘I rich’

‘I am rich’ in Sinhalese

### 5.3.5. Tagalog

Tagalog (Filipino) is derived from *Taga-ilog*, which literally means ‘from the river’. It is an Austronesian language family which belongs to the Malayo-Polynesian branch. A quarter of the population (21.5 million people) in the Philippines speaks Tagalog as an L1, and it is spoken as an L2 by a great majority of Filipinos. Filipino is the national language of the Philippines. It is a standard register of Tagalog and a native language which is spoken and written, in Manila, the capital city of the Philippines (Gripaldo, 2005).

Tagalog has no grammatical gender, no grammatical cases, and no conjugation based on person or number (Llamzon, 1966). Regarding the headedness of Tagalog’s morphological structure, it is left-headed, as prefixation, infixation, circumfixation, and reduplication are the majority of the morphological processes found in Tagalog, while suffixation plays a minor role in its morphological processes (Salt, 1992). Tagalog generally tends to have a great deal of derivation, but no inflection. An extensive derivational system can be found in verbs, nouns, adjectives, numbers, gerunds, and compounds (Salt, 1992).

Sections 5.3.5.1–5.3.5.5 provide a brief description of the five morphosyntactic features of Tagalog relevant to this project. The language has some infixes or prefixes (depending on whether the word begins with a consonant or a vowel) to assign a tense to the root verb – past, present, or future (Voskuil, 1996).

### *5.3.5.1. Agreement in Tagalog*

#### 1. Verbal Agreement in Tagalog

The following is a typical sentence pattern sequence in Tagalog: action (verb) – doer (subject) – receiver (object) and modifier. However, Tagalog uses affixes and markers to show word order in a sentence, so in this sense, the language does not have a strict 'word order' rule. The focus generally comes last, but not when it is a pronoun – pronouns tend to move after the predicate. In the Tagalog system, the marker on the noun indicates which word is in focus, and affixes in the verb indicate what the focus is doing. There is no subject-verb agreement, only a single form of verbs for all pronouns in each tense. In fact, Tagalog verbs are inflected only for aspect, as the language does not have a tense system (Schacter and Otanes, 1972/1983). All verbs in Tagalog undergo a derivation process. Voskuil (1996) indicates that Tagalog has a rather complex verbal system, which is based on the use of affixes (infixes or prefixes) depending on whether the word begins with a consonant or a vowel, and these affixes are used to assign a tense to the root verb (past, present, or future). For example, the root word *Kain* (eat) becomes the past tense/infinitive *Kumain*, the present tense *Kumákain*, and the future tense *Kakain*.

#### 2. Nominal Agreement in Tagalog

Nouns in Tagalog do not inflect for number, size, and proximity, and only some borrowed nouns from Spanish are marked for gender. According to Schachter and Otanes (1983), there is no agreement between the adjective and the NP; instead, the linker *na*[17] is used to link or hook up the modifier (adjectives) and the noun it is describing, for instance *matapang na lalaki* ('brave man') / *matapang na mag lalaki* ('brave men') and *matalinong bata* ('smart child') /

[17] The linker *na* has three forms: 1) *-ng* is used when a word ends with a vowel; 2) *-g* is used when a word ends with the letter *n*; and 3) *na* is used when a word ends with a consonant. The first two (*-ng* and *-g*) are attached directly onto the end of the word, while the third form *na* is written as a separate word (Henry and Zerwekh, 2002).

*matalinong mag bata* (‘smart children’). As for plurality in Tagalog, the plural marker *mga* is always placed directly before the noun to indicate pluralisation, for example *bato*- SG (‘stone’) and *mga bato*- PL (‘stones’). Schachter and Otanes (1983) also indicate that numbers, when used as modifiers of nouns, behave differently from adjectives in that they can only come before the noun being modified, and the linker *na* is still necessary.

#### *5.3.5.2. Definiteness in Tagalog*

Tagalog has markers to indicate definiteness. According to Henry and Zerwekh (2002), *ang* is a noun definitiser; it functions like the definite article *the*, while *ng* functions like *a* or *an*, in English. For instance,

(16) Gusto ko *ang*- DEF mangga.
‘I like the mango’.

(17) Gusto ko *ng*- INDF mangga.
‘I like mangoes’

#### *5.3.5.3. Pronouns in Tagalog*

According to Henry and Zerwekh (2002), Tagalog has free personal pronouns, which are utilised to replace NPs that refer to a person(s). These pronouns are used among young Tagalog-speakers in the Philippines to refer to both animate and inanimate nouns. There is no gender distinction in the use of third-person pronouns (*siya* ‘he/she’) as in Sentence (18) below:

(18) *Si* (Jose) went to the park
Then *siya* (he) listened to music
Then *siya* (he) climbed a tree

Schachter and Otanes (1983) state that Tagalog markers such as *ng*, *ni* and *nina* are also used to indicate possession. The possessed or owned thing comes before *ng*, and the possessor or owner comes after *ng*. Below are some examples:

(19) kotse *ng* babae

'Car of the woman'

(20) *kaibigan ni Juan*

'A friend of Juan'

(21) *ang bahay nina Tina*

'The house of Tina and family'

Other markers to indicate possession are *kay* and *kina*, which come before the noun possessed. For instance,

(22) *kay* Mariang kapatid

'Maria's brother'

(23) *kina Litang* hardin

'Lita's family's garden'

#### *5.3.5.4. Coordination in Tagalog*

Schachter and Otanes (1983) report that Tagalog uses several free conjunction markers, and each one has possible functions (e.g. *at* 'and', *at saka* 'and… too', *ngunit* 'but', and ***o*** 'or'). Below are some examples:

(24) Tumawag si Mary *at* si John

'Mary and John called'

(25) itim *o* puti'

'Black or white'

#### *5.3.5.5. Copula in Tagalog*

According to Schachter and Otanes (1983) and Coon (2013), Tagalog does not have a copula. However, Richards (2009) argues that Tagalog does have a copula but that it is often null. He adds that the null copula is possible only in contexts where aspect might be omitted, and it is overt only when it is used to support overt aspectual morphology.

### 5.3.6. Sunda / Basa Sunda

The word 'Sunda' originates from Sanskrit and literally means either 'light' or 'water'. Sundanese is a Western Austronesian language spoken by more than 39 million native speakers in the region of western Java, and it is the third most-spoken language in Indonesia, after Bahasa Indonesia and Javanese (Sukmawan, 2017). In West Java, particularly in Sundanese society, Sundanese is used in daily live interaction. However, most Sundanese people, similarly to other Indonesians, are bilingual; they can speak both Sundanese – their mother tongue – and the Indonesian national language.

Sundanese has several regional and sociolinguistic varieties. Levels of speech vary depending on the relative social status of the speakers. They use the same grammar and syntax, but differ in vocabulary. Sundanese nouns and verbs have no inflections; their morphology is based on affixing and reduplication. The normal word order in Sundanese is SVO (Müller-Gotama, 2001). Sections 5.3.6.1–5.3.6.5 highlight the morphological features of Sundanese under investigation.

### *5.3.6.1. Agreement in Sunda*

#### 1. Verbal Agreement in Sunda

According to Hardjadibrata (1981:1985), in Sundanese, the verb does not inflect for number, gender, or tense. Furthermore, verbs are based on social status (e.g. close friends, mothers, old people, etc.). They can be transitive or intransitive; unaffixed root words can express intransitive verbs, while transitive verbs can be expressed through the active or passive voice. Suffixes *-i* or *–ake* are used in most transitive verbs. For example, the suffix ***-i*** is attached to the verb to establish the relationship between the verb and its direct object, and it can also attach to a noun to convert it into a verb. The suffix *–ake* is attached to the verb and makes it transitive with causative meaning and also marks it as the indirect object. Sundanese verbs can inflect for aspect/tense by using adverbial markers that precede the verb. For instance,

*tau* → for remote past
*wis*→ for past (‘already’)
*bakal/arep*→ for future (‘will’)
*tansah*→ for habitual action (‘always’)
*lagi*→ for an ongoing action (‘in the process of doing’)

Negation is also expressed by adding *aja* (‘don’t’), *durung* (‘not yet’), or *ora* (‘not’), and *ya* is an imperative marker.

#### 2. Nominal Agreement in Sunda

Nouns and adjectives are not inflected for gender, case, or number (Hardjadibrata, 1985). Sundanese uses reduplication to express plurality, and partial reduplication to derive a noun from an adjective, for example *peteng* (dark) > *pepeteng* (darkness). The prefix *ka-* can be combined with the suffix *–an* in abstract nouns to form a concrete noun or to form an adjective, for example *rosa* (strong) > *karosaan* (strength). Moreover, another combination of a prefix *pa-* with a suffix *-an* is attached to verbs or nouns to form a noun denoting place. If the suffix *–an* is added by itself to certain verbs, then it produces nouns that are the result of their action, such as *nandur* (to plant) and *tanduran* (a crop), or that indicate the instrument by which an

action is carried out. Adjectives can also be formed by *-an* or other affixes, for example *jamur* (fungus) and *jamuren* (mouldy).

#### *5.3.6.2. Definiteness in Sunda*

Definiteness may be indicated by adding the suffix *-(n)é* to a noun. Indefiniteness may be indicated by reduplication of the noun (Hardjadibrata, 1985).

#### *5.3.6.3. Pronouns in Sunda*

According to Hardjadibrata (1985), possessive pronouns are free or bound morphemes used to express possessive and object pronouns. For example,

*Pun baba* → 'my father'
*-na→Babana→* 'his/her father'

#### *5.3.6.4. Coordinating Conjunction in Sunda*

In Sundanese, conjunction is normally expressed using syndetic linkage, such as *je* ('and'), *atawa* ('or'), and *tapi* ('but').

#### *5.3.6.5. Copula in Sunda*

Sundanese does not have copulative verbs. However, as stated by Wetzer (1996: 359) that:

> Both Hardjadibrata (1985) and Robin (1968) mentioned the existence of the verb *jadi* "be, become" without giving further information about its actual use as a copula (with predicate nouns). Since both grammars explicitly state that nominal predicates are generally expressed without an overt copula, Sundanese is classified as a type-B language.

In Table 5.2, I attempt to provide a concise and easily accessible cross-linguistic comparison of GPA, GA, and the major substrate languages of GPA.

| Language / Feature | GPA | GA | Bengali | Punjabi | Malayalam | Sinhala | Tagalog | Sunda |
|---|---|---|---|---|---|---|---|---|
| Definiteness | ø Markers | *al*-(prefix) | Expressed indirectly (DEMs e.g. *–ta*) | ø Markers (expressed indirectly) | ø Markers (expressed indirectly) | ø Markers | Using *ang* for DEF | *-na* for DEF |
| Indefiniteness | ø Markers | ø Markers (Expressed indirectly) | prefix *ek*-('some') | ø Markers (expressed indirectly) | *oru* (only with SG count Ns) | ø Markers (expressed indirectly)<br>*-ek* for animates and<br>*-ak* for inanimate | *ang* for INDEF | No indefinite |
| Coordinating Conjunction | ø Markers | Markers used | Optional CONJ markers | *ate* ('and') *jar* ('or') | Markers are used | No conjunctions but only non-finite clauses | Markers are used<br>*at* ('and')<br>*o* ('or') | *Je*' and *' atwa* ('or')<br>*tapi* ('but') |
| Copula | *fi* (optional) | PRS: ø PST: *kan*.AGR *Sar*.AGR | ø, except short-term ADJs *achh* | In positive sentences only | *aakuka* (optional, but highly frequent) | No copula, except use two existential verbs only | ø, except optional in SVO sentences | No copula in predicate nominal |
| Verbal Agreement | Ø Default is GA 3SGM | P, N, G, and case | P, but not N or G | N, G, and P | no S-V AGR | AGR P, N, G, and tense AGR only in literary | No S, V AGR | AGR in P, N Inflected for P, A |
| Adj + NP Agreement | Ø Default is GA SGM | N-ADJ AGR in DEF, N, and G | N-ADJ AGR exists, DEMs inflect for N and size | N and G (only in adjectives ending with *-aa*) | P, G, and N (predicative adjectives only) | N-ADJ AGR in N, G, P and DEF | No G AGR | No AGR |
| Possessive Pronouns | Same as SBJ PROs | Suffix (inflect for P, N, and G) | Inflect for P, politeness, and N, but not for G | Inflect for P, N, and G | Inflect for P, case, N, G, politeness, and proximity | Inflected for P, N and not for G | Inflect for P, N, and politeness (*ho*/*po*), but not G | Inflected for N and P, not for G |
| Object Pronouns | Same as SBJ PROs | Suffix (inflect for P, N, and G) | Inflect for P, politeness, and N, but not for G | Same as SBJ PROs (Except 3rd P) | Inflect for P, case, N, G, politeness, and proximity | Inflected for P, N and not for G | Inflect for P, N, and politeness (*ho*/*po*), but not G | Inflected only for P |

Table 5.2: A cross-linguistic comparison of the morphosyntax of GPA, Gulf Arabic, and the six substrate languages (GPA = Gulf Arabic Pidgin, GA = Gulf Arabic, DEF = definite, , DEM = demonstrative, INDEF = indefinite, CONJ = conjunction, PRS = present, PST = past, VA = verbal agreement, ADJ AGR = adjective agreement, SBJ PRO = subject pronoun, P = person, G = gender, N= number, SG = singular, N = noun, N-ADJ AGR = noun adjective agreement).

The description of the substrate languages in Table 5.2 highlights some differences in the morphosyntactic systems of these six substrate languages. For example, Tagalog has no subject-verb agreement, whereas in Punjabi, the verb agrees with the subject in person, number, and gender. Thus, language variation patterns were expected in GPA based on the morphosyntactic differences.

The end of this chapter contains a detailed morphosyntactic description and analysis of the substrate languages or L1 mother tongues of the female speakers recorded in the corpus. The analysis of morphosyntactic features in six L1s, as summarised in Table 5.2, is a significant novel research, and it is one of the research contributions of this study. Almoaily (2012) provides an accessible, cross-linguistic comparison of GPA, GA, and the three substrate languages of GPA (Bengali, Punjabi, and Malayalam), while I provide a linguistic comparison of the remaining three substrate languages of GPA (Sinhala, Tagalog, and Sunda). Next, Section 5.4 provides the conclusions of this chapter.

### 5.4. Conclusion to Chapter 5

The first purpose of this chapter was to define the six main substrate languages of GPA – Bengali, Malayalam, Punjabi, Sinhala, Tagalog, and Sunda – and then describe the relevant morphosyntactic features in these languages (i.e. agreement, pronouns, definiteness/indefiniteness, coordination, and copula). The second purpose of the chapter was to conduct a comparison between GA, GPA, and the described languages and discuss how the similarities and differences between the substrate languages may possibly play a role in the variation encountered between the female speakers of GPA. The next chapter presents in detail the methodology and data used in my study.

# Chapter 6: Research Methodology

This chapter concentrates on the methods employed to solve the problems identified in this study. The research problem in a nutshell is the analysis of gender variation among GPA speakers in the Gulf region, with special reference to the influence of their L1 and length of stay in the Gulf countries on their production of GPA. This chapter delivers a detailed background of this current work, with a particular focus on the research purpose and the structure of the study's corpus. A significant part of this chapter is devoted to an overview of the procedures and outcomes of the pilot study. This is followed by details about the process of creating the corpus (i.e. sampling strategies, the rationale behind their selection, instruments of data collection, the process of conducting and transcribing the interviews, and the procedures followed in counting and labelling the tokens). Section 6.6.9 presents a summary and discussion of some potential limitations in the data used for the pilot study and preparation of datasets for analysis in the main study. Finally, the chapter ends with the summary and conclusions of this chapter, providing the justifications for the choice of approaches applied for the purposes of this thesis.

## 6.1. Background Description of the Current Study

The review of literature in Section 3.1.1.4 highlighted the grey area surrounding studies on GPA in several countries of the Gulf, such as the UAE (Smart, 1990), Kuwait (Wiswall, 2002), SA (Almoaily, 2008; Alshammari, 2018), Oman (Naess, 2008), and Qatar (Bakir, 2010). Most previous studies on GPA have concentrated on male speech, and conclusions have been drawn without quantitative variationist analysis investigating GPA variability. Despite the increasing

labour market demand for women migrants in the Gulf – a demand which is often more stable than that for men – studies are yet to investigate the gender variation in Gulf countries (see Section 2.10). The review of literature further demonstrated that no quantitative variationist study has been done so far on GPA speakers who are all women. The present study aims to provide a quantitative analysis of language variation in GPA based on gender differences, different morphosyntactic systems of the substrate languages of female GPA, and the number of years they spent in the Gulf according to five morphosyntactic phenomena: definiteness and indefiniteness, coordination, copular verbs, pronouns, and agreement in the VP and in the NP and ADJP. These morphosyntactic features of the substrate languages of GPA, which is the pidgin under investigation in this project, and its lexifier, GA, have been detailed in Chapter 4 and 5. Based on these differences (summarised in Table 5.2, Chapter 5), a set of hypotheses were formulated and broadly identified for carrying out the study, as described in Section 6.2 below.

Regarding gender variation based on the duration of the speakers' stay in the Gulf, Almoaily (2012) reported that male GPA speakers tend to shift to GA in only one feature (i.e. conjunction markers), after spending 10 years or more in the Gulf countries, while Bakir (2010) reported that long-staying speakers of GPA tend to shift to GA. Thus, in this study, I compare my results of female GPA speakers with the data from Alomoaily's (2012) male GPA speakers. This allows me to investigate the question of whether female GPA speakers shift towards GA more than male GPA speakers after spending more than 10 years in the Gulf.

Furthermore, Versteegh (2014) emphasised one of the main differences between the language acquisition of children and that of pidgin speakers. He claimed that a child tends to shift towards the TL, while pidgin speakers tend to fossilise at a certain stage of the process of language acquisition. According to his claim, I investigate the impact of the length of stay in the Gulf on female GPA speakers by comparing the data of newly settled female GPA speakers

with that of speakers who have stayed longer in the Gulf. This is done by comparing the proportional use of GA tokens among those with a long stay in the Gulf with those of their newly settled counterparts.

As for the influence of a substrate L1, the role of language transfer (i.e. cross-linguistic interference) is not only restricted to full-fledged languages, as is widely discussed in the literature of SLA, but also carried out in P/C languages. As a result, numerous intersections of the phenomena have been studied in the fields of contact and transfer. For example, Thomason and Kaufman (1988: 37) have discussed studies that combine both language transfer and language contact, distinguishing between borrowing and substratum interference (i.e. transfer). Siegel (1999, 2003) outlined various transfer constraints and substrate influences of Central-Eastern Oceanic languages on Melanesian Pidgin.

### 6.2. Hypotheses

This analysis is based on informants' use of the variants of five selected morphosyntactic phenomena. I kept the same five morphosyntactic features that Almoaily adopted in his study for two reasons. First, to investigate gender variation in GPA speakers, I must compare my female corpus with Almoaily's male corpus. Second, these five grammatical features of atypical contact languages are believed to be adequate to test the proposed typological features of P/Cs (Almoaily, 2012). These typological features are as follows: (1) reduced inflection – reduction of agreement markers in verb and noun and adjective agreement – and (2) reduced inventory of function words – (3) copulas, (4) definite and indefinite articles, and (5) pronouns. These may be found in all P/C languages worldwide (irrespective of their input languages) – see Almoaily (2012); Bakker (1995, 2003); Roberts and Bresnan (2008); Bakker, Daval-Markussen, Muysken, and Parkvall (2011); Sebba (1997); and Siegel (2004).

The analysis is based on female informants' use of the variants of these five selected morphosyntactic features (note that the contrasts marked in the substrate languages are summarised in Table 5.2). Next, I briefly summarise these phenomena.

1. **Object or Possessive Pronouns**

The object and possessive pronouns in GA are always suffixes (bound morphemes) attached to the noun or the verb (Qafisheh, 1977). In GPA, they are used as independent pronouns to replace the bound object and possessive pronouns that are bound in GA (Smart, 1990), for example, in GPA, *bēt ana* ('house I') instead of GA's *bēt-i* ('house-my'). However, in the speech of GPA, I have noticed that the GA object and possessive pronouns may occur as Almaoily (2013) mentioned. The object and possessive pronouns in GPA can take three variants: free pronouns, bound pronouns, and the dropping of the pronoun.

2. **Presence or Absence of the Arabic Definiteness Marker *al***

GPA speakers usually drop the GA definiteness marker *al-* (الـ). However, it has been noticed that some GPA speakers use it infrequently, as also documented by Almoaily (ibid). For example, a GPA speaker may say *aish maudu?* ('what is matter?') or sometimes use the GA definiteness marker *aish almaudu?* ('what is the matter?').

3. **Presence or Absence of Arabic Conjunction Markers**

According to Feghali (2004), GA has a range of coordinating conjunction markers, such as *willa* and *aw* ('or') and *wa* ('and'), which are normally dropped in GPA (Almoaily; ibid). For example, in *θalaθah Ø arbaah* ('three four') and, less frequently, when using the conjunction markers, as in *θalaθah aw arbaah* ('three or four'), both of these sentences can be translated into English as 'three or four'.

### 4. Presence or Absence of the GPA Copula *fi*

The copula in GA is used overtly only in the past and future tenses, whereas it is covert in the present tense (Holes, 1990). In other words, the GA copula is not used in the present tense and is optional in GPA. For example, the English sentence 'I am a teacher' can be translated into GPA as either *ana fi mudarris* or *ana Ø mudarris.*

### 5. Presence or Absence of Agreement in the Verb Phrase and the Adjective Phrase

#### *1) Agreement in the Verb Phrase*

In GA, the verb agrees with the subject in gender, number, and person, whereas in GPA, it does not agree with the noun in person, number, or gender. Instead, the third-person singular masculine form of the verb is usually used with all subjects (Almoaily, ibid). For example, the GA sentence *hi t-drus fi il-midrisah* ('she 3SGF.PRES-study at -the-school') in GPA is *hiya y-drus fi madrasah* ('she 3SGM-study at school'). Almoaily (2013) stated that it is unusual to find examples in GPA where the verb agrees with the noun in person, gender, and number. Therefore, a set of variants are addressed here: (i) use of the GA agreement markers with the agreeing noun; (ii) use of GA agreement markers, but where the marker does not agree with the noun; and (iii) use of verb-fewer utterances.

#### *2) Agreement in the Adjective Phrase*

The GA adjective agrees with the noun in number and gender (Almoaily, ibid), as in *il- madras-ah gidi:m-ah* ('the-school-SGF old-SGF'), whereas the singular masculine form is always used with GPA adjectives irrespective of the head noun (e.g. *madras-ah gadiim* 'school- SGF old.SGM'). In some cases, the GPA adjective may agree with the noun in number and/or gender.

A list of hypotheses is formulated below to determine the potential effect that the three factors – (i) the speakers' gender, (ii) their L1 and (iii) the number of years they have spent in the Gulf – have on variability in GPA (based on the five morphosyntactic phenomena). Differences in

the substrate languages are expected to have an effect on the choice of available GPA variants, since the six languages tested in this study (i.e. Malayalam, Punjabi, Bengali, Sinhala, Tagalog, and Sunda) are relatively dissimilar typologically. Tagalog and Sunda are closely related, as they belong to the Austronesian languages. However, Punjabi, Bengali, and Sinhala are dissimilar to the languages above, as they are related to a different family, the Indo-Aryan languages (see Chapter 5). In the current study, I investigate whether the participating informants use morphosyntactic features similar to those found in their L1s when they speak GPA. If the findings reveal that the sampled informants produced universal features of contact languages which cannot be traced to their L1s, and furthermore that divergent properties of their L1s did not have a significant effect on their production of GPA, then this study would support universalist theories of genesis.

**Hypothesis 1**: **There are differences between standard GA and GPA.** This hypothesis is tested through a qualitative investigation, which involves describing the selected aspects of the morphosyntax of GPA and its lexifier, GA. These features are agreement in the VP and in the NP and ADJP, definiteness and indefiniteness, pronouns, coordination, and copular verbs. I will discuss each in turn in the following subsections. The female speakers recorded in the corpus are described, evaluated, and compared qualitatively.

**Hypothesis 2: There is a difference between the GPA spoken by speakers with different L1s**. This hypothesis is tested by conducting a quantitative analysis based on informants' use of the variants of the five selected morphosyntactic phenomena.

**Hypothesis 3: Length of stay in the Gulf produces accommodation to standard GA**. This hypothesis is tested by conducting a comparative quantitative analysis of the data from newly settled GPA speakers with that of speakers who have stayed longer in the Gulf. The findings will answer the question of whether GPA speakers shift to GA after spending some time in the Gulf.

**Hypothesis 4**: **Length of stay in the Gulf produces more accommodation to standard GA in women than men**. This hypothesis is tested by conducting a comparative quantitative analysis of my female data with Almoaily's (2012) male data.

### 6.3. Data and Methodology

This section explains the various research methods that were employed to gather data in this study. It also provides valuable evidence in favour of one or two of the competing theories of genesis of P/Cs. It is divided into two main parts, the first of which is the pilot study phase (Section 6.6), which is further split into three subdivisions. First, Sections 6.4 and 6.5 describe the purpose of a pilot study, illustrate how the pilot study was conducted, and explain how the resulting corpus has been structured. The subsequent subsections present details about the procedures of creating the corpus (i.e. sampling, preparing for, conducting, and transcribing the interviews), as well as the procedures followed in counting and labelling the tokens (Sections 6.6 to 6.6.8). Lastly, Section 6.6.9 highlights the main findings of this pilot study and concludes with a discussion around the potential limitations of the data used for this research and the ways in which these might be resolved or minimised.

The second main part is the main data collection phase (Section 6.7), which is subdivided into two First, Sections 6.7.1, 6.7.2, and 6.7.3 discuss the number of changes during the process of building the corpus in the pilot study (i.e. increasing the sample size and sampling process). Then, Sections 6.7.4, 6.7.5, and 6.7.6 describe the process of analysing the main corpus data (i.e. transcription of the interviews, annotation and counting of the tokens, and quantification of tokens).

### 6.4. Research Design

Dörnyei (2007) stated that, generally, there are two approaches to research methodology: qualitative and quantitative approaches. Although they have particular differences in many aspects (i.e. purpose, strength, and limitation), they also correspond to each other (Neuman,

2007). In fact, both are used to answer research questions, but they look for answers differently (Goertz & Mahoney, 2012). It is currently common to employ both qualitative and quantitative methods in the same research. This so-called mixed method approach can be effective and useful to simultaneously issue in-depth descriptions and take a general conclusion (Sinaga, 2014).

This study is best described as an integrated research design applied to combined/integrated qualitative and quantitative phases of analysis within a single study (Mayring, 2001). Since the present study is a corpus-based study, as it relies on a designed corpus of sociolinguistic interviews with female GPA speakers, the qualitative material was collected and transformed into categorical data for further quantitative analysis aimed at deriving generalisable results (Srnka and Koeszegi, 2007). However, the integrated design constitutes a relatively new approach that is still developing (Creswell, 2003), and integrated research design studies suggest that the generalisation model represents a research approach that successfully accomplishes two goals. First, it provides significant insights into the research problem and thus responds to the many calls for discovery-oriented research. Second, it assures scientific rigor and allows for the derivation of generalisable results from qualitative data (Mayring, 2001).

The current study combines both approaches of qualitative and quantitative research to analyse form selection in GPA. The combination of the two analyses (qualitative and quantitative) occurs when the interpretive analysis is integrated with descriptive statistics. One of the main reasons I adopted an integrated research design is because it can maximise the potential of my analysis, as I seek to develop coherent outcomes that achieve the goals of my research, and this can also strengthen the findings and increase the validity of the results. Another reason I adopted an integrated research design is to fully understand the phenomenon at hand, namely, GPA, ‘when either the quantitative or qualitative approach by itself is inadequate to best

understand a research problem or the strengths of both quantitative and qualitative research can provide the best understanding' (Creswell, 2009: 18).

Therefore, this design is an attempt to uncover the linguistic behaviours surrounding GPA, from both social and linguistic situations which lead to the emergence of this contact language. The results gathered from the two research approaches are complementary to one another and therefore offer a fuller picture.

This study was conducted in two phases. First, a pilot phase took place over 7 weeks in June and July, 2017. The second phase ensued over a 3-month period from February to April 2018; this is referred to as the main data collection phase. Sections 6.5–6.7 outline the objectives, data collection instruments, participants, and procedures for the pilot and for the main data collection phase of the study.

### 6.5. "Do not take the risk. Pilot test first"[1]

A pilot study is considered to be a fundamental preliminary phase in the research process, especially in a quantitative study of linguistic variation: 'Acquaintance with previous work and perhaps a pilot study should help to narrow the focus of the project' (Feagin, 2013: 23). A pilot study is a "small- scale test of the methods and procedures to be used on a larger scale …" (Porta, 2008), with the main purpose being to examine the feasibility of an approach that is proposed to be carried out in a larger scale study. In other words, in a pilot study, the entire process is carried out, but with fewer participants than would be used for an extensive study.

In addition to achieving all of the usual objectives, such as improving data-collecting routines and checking the appropriateness of standard measures, the pilot study plays an essential role in providing additional knowledge that can be used to improve and modify the design of a larger hypothesis testing study (Leon, Davis, & Kraemer, 2011). It can also be the pre-testing

[1] De Vaus (1993: 54).

or ‘trying out’ of a particular research instrument (Baker, 1994: 182-3). The advantage of conducting a pilot study is that it greatly reduces the number of treatment errors because unforeseen problems revealed in the pilot may be overcome in redesigning the study; these may include the main research project failing, research protocols not being followed, or proposed methods being inappropriate or too complicated. Moreover, Feagin (2013: 21) indicated that ‘a small-scale pilot project along the general lines of the main research will indicate more precisely what might be feasible goals and procedures’.

## 6.6. The Pilot Phase

The primary aim of this pilot phase was to test the reliability of linguistic annotation, as recommended by Granger (1997), whatever the type of corpus used, for example with learner corpora.

### 6.6.1. Location

The pilot study was conducted in Riyadh, which is located in the Central Province of SA. As the country’s capital, Riyadh hosts numerous government ministries and public service headquarters, making the public sector the city’s largest employer, and it is dominated by FWs. Riyadh is an important financial, business, and manufacturing centre (refer to Section 2.7 for more details regarding the main setting of the research study, Riyadh City).

During the 7-week study period between June and July, 2017, I started my field trip by seeking out places where I was likely to meet female GPA speakers whose L1s are Tagalog, Punjabi, Sinhala, Malayalam, Sunda, and Bengali. These places included family homes where GPA speakers live and work as house maids, and work places of GPA speakers, such as hospitals and hairdressing salons.

### 6.6.2. The Corpus

The corpus comprises the speech of informants participating in the interviews, which I conducted in SA, Riyadh. The database consisted of interviews with 12 female GPA-speaking

informants from six linguistic backgrounds, Malayalam, Bengali, Punjabi, Sinhala, Tagalog, and Sunda. Half of the data was produced by informants who had spent five years or less in the Gulf, while the other half had spent 10 years[2] or more in the area at the time of interview (for example, one female Malayalam who had stayed for a long term was balanced by one female Malayalam who had stayed for a short time).

This study focuses on examining the structure of GPA rather than its lexicon; thus, counting of the lexical features has been excluded. Additionally, vocabulary studies are more related to developed languages. For example, Malmasi et al. (2016) identified a set of four regional Arabic dialects (Egyptian, Gulf, Levantine, North African) and MSA, which are all native languages, unlike GPA, which has no native.

General principles for the quantification of variability above the level of phonology[3] are still subject to debate (Macaulay, 2002). A number of researchers have devised several approaches for the quantification of tokens. On the one hand, some quantify them by the number of words. For example, Precht (2008) quantified gender similarities and differences per 1,000 words in American English conversations, and Cheshire, Kerswill, and Williams (2005) calculated variation in discourse per 1,000 words. On the other hand, some researchers prefer to quantify tokens per minute or hour of speech in a sociolinguistic interview, as was done by Rickford and McNair-Knox (1994), who examined the effect of the interviewer's race by calculating the tokens of African American syntactic features per hour of speech. In my case, I preferred to calculate the tokens per number of words, as Almoaily (2013) suggests, irrespective of the length of the turn or the number of words produced in a minute of speech. The reason was that

[2] I was unable to conduct interviews for two short-stay informants, one Bengali speaker and one Sunda speaker, as it was difficult to find them due to time limitations.

[3] Combith, Barlow, and Sanchez (2019) examined the single-word productions of 275 children with phonological disorder from the Learnability Project (Gierut, 2015b) to confirm the relationship between phonemic inventory (a measure of phonological knowledge) and consonant accuracy – a quantitative, relational measure that directly compares a child's phonological productions to the target (i.e. adult-like) form.

the informants of the study have been exposed to GPA over a period ranging from eight months to 25 years, and newly arrived speakers are expected to speak more slowly than those who have spent more than 10 years in the Gulf.

I took 1,000-word samples from each informant, which are sufficient to analyse the data, as Almoaily (ibid) suggested. These 1,000-word samples allowed me to conduct the comparison between the following: (1) the relative number of linguistic features produced by female speakers of different language groups (e.g. tokens produced by the Tagalog female respondents vs. tokens produced by Sinhalese female respondents), (2) female informants with different lengths of stay (e.g. among the newly settled Malayalam-speaking female informants vs. long-staying Malayalam female speakers), and (3) gender differences based on lengths of stay (e.g. newly settled Bengali-speaking female informants vs. newly settled male Bengali speakers, and long-staying Bengali-speaking female informants vs. long-staying male Bengali speakers). Overall, I created a balanced database containing 10,000 words: 2,000 words per substrate language, 1,000 of which were produced by newly settled informants, and the remaining 1,000 words from long-term residents. The purpose of analysing an equal number of words from every language group is to have a fair means of comparison for the GPA speakers' data, regardless of the length of their turns or the words they utter per minute or per hour of speech. All the participants were equally distributed between the two criteria, except for two language groups which did not adhere to this distribution – two short stay informants (one female Bengali speaker and one female Sunda speaker) – because it was a little difficult to find and conduct interviews with them due to time limitations. The data distribution is illustrated in Figure 6.1, where the 'new' and 'old' tags do not refer to the actual age of the participants, but to the number of years spent in the Gulf. Moreover, the informants are all females, to enable me to compare them with Almoaily's male data and investigate gender variation.

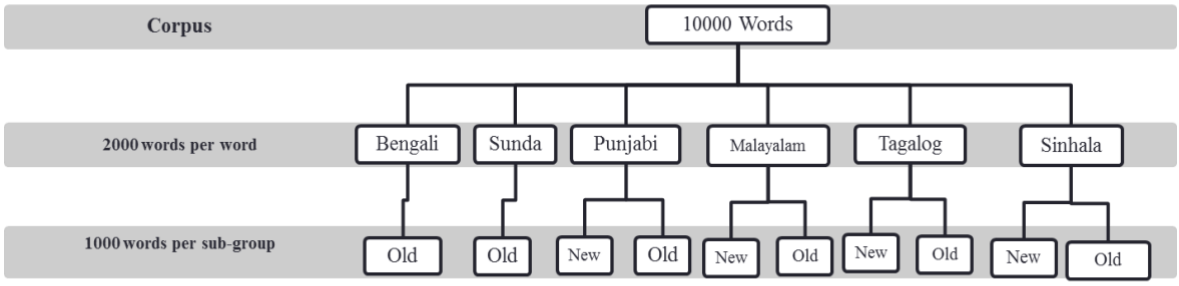



Figure 6.1: Distribution of the data.

### 6.6.3. Building the Corpus

This section discusses each of the four stages in building the corpus. They are (1) preparation (strategies in deciding on sampling methods, interview procedures, and informants participating in this study), (2) data collection by conducting the interviews, (3) the process of transcribing the interviews, and (4) extraction of the required amount of data.

This study relies on the use of a suitable corpus, and since GPA is only a spoken variety of the Arabic language, no such corpus was previously available in electronic form. Therefore, I had to design and build my own corpus. A number of issues must be taken into consideration when building a corpus, and they are discussed below.

Feagin (2013) pointed out that the ultimate goal of sociolinguistic research is to solve questions of linguistic importance, such as investigations of language variation and change or analyses of the roles that language or particular linguistic features play in the construction of individual or group identities. These types of studies cannot be accomplished unless the researcher first enters a community and conducts sociolinguistic fieldwork to collect data which will provide the basis for any linguistic questions. The dominant problem and the major challenge in collecting sociolinguistic data is that when people know they are being listened to, they tend to pay more attention to their speech than normal. This is what Labov describes as the *observer's*

*paradox*[4]: 'our goal is to observe the way people use language when they are not being observed' (1972a: 61). Observing speech within a natural context in sociolinguistics fieldwork – whether tape-recorded interviews, participant observations, or street conversation – must be geared to work around this problem.

Recording of speech within a natural context is one form of sociolinguistic fieldwork that captures the way people actually talk in casual settings. On the one hand, this kind of data collection affords the researcher the best possible representation of the natural linguistic world by creating a comfortable environment to promote speech unmonitored by the speaker. On the other hand, recording conversations can make the situation even worse: The interviewee might be influenced by the interviewer's presence, thereby causing changes in the language being absorbed. To minimise the observer's paradox, Wolfram (2018: 177) suggests that 'all interviews should be conducted in settings that are comfortable for the individual participants' (further discussion in Section 6.7.2. Overcoming Observer's Paradox).

In addition, a number of difficulties and challenges were also associated with implementing this corpus. These included size, balance (choosing informants), representativeness, and transcription protocol and annotation, as discussed below:

- **Size**: The size of the corpus depends heavily on the types of tasks to be investigated. Most researchers generally agree that bigger is better. However, it is possible to obtain useful data from a small corpus, particularly when investigating high-frequency items. In the case of transcribing spoken data, Evans (2007) points out that 'If you need to transcribe spoken data with a high degree of detail then it may only be feasible to work with thousands rather than millions of words'.

[4] 'A problem faced by field researchers whereby the presence of observers of language changes the linguistics context and thus possibly the language data collected' (Wolfram, 2018).

- **Balance**: In the case of collecting data for a spoken corpus, Evans (2007) suggests that to achieve balance in the data, it is necessary to consider the demographics of the people whom researchers use as informants. This was one of the obstacles I observed while conducting the interviews. It was difficult to find female GPA speakers who met all the requirements (i.e. spoke either Tagalog, Punjabi, Sinhala, Malayalam, Sunda, or Punjabi as their L1 and had lived in the Gulf for either five years or less, or 10 years or more) and convince them to participate in the interview.
- **Representativeness**: A corpus can be said to be representative if the findings from that corpus are generalisable to a language or a particular aspect of the language as a whole (Evans, 2007). Since part of the process of data collection was purposefully formulated to elicit variants of certain linguistic variables, this data elicitation procedure may not be fully representative of an informant's normal use of the TL. In addition, the informant will be more aware of aspects of linguistic communication and become more self-conscious. This was one of the common predicaments I faced when conducting sociolinguistic interviews. Bayley and Preston (1996: 2) describe this phenomenon as follows: 'the more aware respondents are that speech is being observed, the less natural their performance will be'. The following subsections explain the first step of creating the corpus: selecting informants.

#### *6.6.3.1. Sampling Strategy*

Obtaining a representative sample that closely matches the whole population is the primary goal of a sampling process, whatever method is chosen. Sampling is a method that allows researchers to examine a smaller group of subjects (a sample) and produce accurate generalisations about the larger group (a population). Ruane (2005: 105) states that 'samples can be very efficient devices –they allow us to look at the "few" in order to know about the many'. Importantly, Schilling-Estes (2007: 166-67) focuses more on the goals of the study, as the sampling strategy depends heavily on them. She states that to decide on the sample, one of

the first questions should be, ‘what counts as the speech community?’ Therefore, researchers must define the speech community, either in terms of empirical linguistics or via social analyses, to determine which people should be targeted for data collection. Since my study investigates GPA phenomena, the speech community consists of native GA people and Asian immigrant workers in the Gulf region (see Chapter 2 and Smart 1990). Buchstaller and Khattab (2013: 74) believe that:

> the most reliable method for finding out about the language use of a particular group of people would be to collect linguistic information from every single person in the population, which in the social sciences refers to all members of the community.

However, this method is impractical, expensive, and time consuming. Hence, the target here was to look for ‘some people in the group in such a way that their responses and characteristics reflect those of the group from which they are drawn ...This is the principle of sampling’ (De Vaus, 2001: 60). In this case and other cases, examining a whole population is extremely difficult; therefore, Ruane (2005) stated that it is important to select the number of individuals to be representative of the whole population.

In research methods, several different sampling techniques are available and can be subdivided into two groups: probabilistic sampling and non-probabilistic sampling (Berg, 2001; Puri, Watson, and Newing, 2011). The difference between the two is whether the selection of the sample is based on the idea of randomisation. The selection is random if every participant has an equal chance of being chosen and being part of the sample for study. In probabilistic sampling (e.g. random sampling, systematic sampling, stratified sampling, and cluster sampling), the researcher will be more able to generalise the results from their study, and it is thus thought to be appropriate for use in quantitative research. However, these sampling methods tend to be more time consuming and expensive than non-probabilistic sampling. In non-probabilistic (non-random) sampling, randomisation does not apply (i.e. in a non-probability sample, individuals are selected based on non-random criteria, and not every

individual has a chance of being included). This technique is more reliant on the ability of the researcher to identify specific participants for a sample and fits strongly with qualitative research (Punch and Oancea, 2014). The outcome of sampling might be biased, and some participants hence have no chance of being selected (Thompson, 2005).

For each technique, a number of sampling methods are used in language variation research, such as convenience sampling, random sampling, stratified sampling (judgment or quota), ethnographic sampling, and snowball sampling (see Shuy, Wolfram, and Riley, 1968; Chambers, Drinkwater, and Boath, 2004; Milroy and Gordon, 2008; Buchstaller and Khattab 2014). These sampling techniques mainly serve the purpose of avoiding a biased selection of the sample and, accordingly, employing a sample which represents the target speech community.

The most prevalent types of sampling tools used in early sociolinguistic research drew heavily on random sampling, which was the most popular technique in social sciences at the time (see Massey, 1994). De Vaus (2001: 60) argues that 'the surest way of providing equal probability of selection is to use the principle of random selection. This involves listing all members of the population (this list is called a sampling frame) and then...pulling their names out of that.' This effectively means that in a random sample of a group or community (e.g. indigenous people of the Arabian Gulf and South Asian immigrants working in that region), every member of that community has an equal chance of being chosen for participation in the study. In this sense, it is inherently difficult to give each individual of the speech community an equal chance of participation, such as by choosing people randomly out of telephone books, 'as this will eliminate low-income members of the speech community who do not have phone numbers and those who chose not to include their numbers in the phone director' (Almoaily, 2012: 120). Ray (1985: 141) also highlights the following difficulty with random sampling: '[t]reatises on sampling generally seem to assume that a random sample has been obtained. In real-life

sampling, however, this seems never to be so – due to rejections to cooperate on the part of some of those drawn'). In the case of GPA speakers, obtaining the consent of all the people who had been randomly chosen was not possible in this study for two reasons. First, no list or statistics are available from which to pull the names or total number of immigrant female GPA speakers in SA (see Sections 5.1 and 5.2). Even the general population census in SA (2010) displays the total number of foreign labourers based only on nationalities (refer to Figure 5.2). In addition, some female GPA speakers refused to take part in interviews for numerous reasons: (1) Some of them were worried that they might lose their jobs if they do not speak Arabic well and hence did not want to be recorded and linguistically observed; (2) some of them were busy and could not find the time, or were not interested in being interviewed; and (3) some were simply uncooperative without mentioning reasons for their refusal. Schilling-Estes (2007) points out that refusal to participate in sociolinguistic studies is a common trend in fieldwork investigations.

Random sampling is not without its problems; thus, an alternative approach was employed to pull out a sample from the GPA speech community. An example is the snowball sampling technique, also known as the social network technique[5]. This aims to investigate locally specific, participant-designed groups and 'does aim to examine quantitative variation across the group, but uses networks for the recruitment and sampling of participants' (Buchstaller and Khattab, 2014: 80). In other words, it is used to recruit further speakers from the same networks (Milroy and Gordon, 2008: 32; Scobbie and Stuart-Smith, 2012: 611). This snowball sampling method starts by asking the first selected member of the population, who is easily accessed, to recommend other members who will fit the description of the sample needed and be willing to

[5] If the population is hard to access, snowball sampling can be used to recruit participants via other participants. The number of people to which researchers have access 'snowballs' as they get in contact with more people (Buchstaller & Khattab, 2014).

participate in the study (to avoid repetition and redundancy, the snow sampling process is discussed in detail in the next subsection).

#### *6.6.3.1.1. Preparing an Interview*

Before my pilot fieldwork, which took place in Riyadh over the summer of 2017, I obtained permission from the University of Wolverhampoton, Faculty of Arts Ethics Committee to begin collecting data so as to ensure compliance with ethical guidelines for conducting research in social science (see Appendix E). Additionally, I obtained an official letter from my supervisor granting permission to conduct interviews for my sample population. I started looking in the places where I expected to find the majority of female GPA speakers. I also began to communicate with companies[6] that are in charge of employing female Asian workers. They directed me mainly to hospitals, beauty centres, labourers' residences, and shopping malls. First, I talked with their employers, seeking their permission and assistance to arrange interviews with volunteers who were willing to participate in the study and showing them the official letter I obtained from my supervisor. Then, I asked each female GPA speaker about her L1 and length of stay in the Gulf. If she met the requirements of this study (i.e. spoke either Tagalog, Punjabi, Sinhala, Malayalam, Sunda, or Punjabi as her L1 and had lived in the Gulf for either five years or less, or 10 years or more), then I introduced myself to her as a postgraduate student studying at the University of Wolverhampton conducting fieldwork on GPA, and I asked for her consent to participate and be recorded in the study. I explained to the participants that GPA is a variety of Arabic and that it is different to GA by providing examples of GPA, such as the following:

(1) ana yi-dris jama'ah
I PRS.3SG.M-study university

[6] Authorized companies recruiting and employing foreign female workers, such as domestic workers, nurses, cooks, babysitters, carers, and cleaners.

‘I studied in university.’

In Sentence (1), I demonstrated to them that the third-person singular masculine present form of the GA verb is used over other subjects, while in GA, the 1SG form of the verb is used with the 1SG pronoun *ana*. Thus, the GA form of sentence (1) would be

| | | | |
|---|---|---|---|
| ana | daras-t | fi | el- jama’ah |
| I | study-1SG.PST | in | DEF-university |

In addition, I explained to the informants that studying their Arabic variety could help in teaching them the Arabic language. The findings regarding the morphosyntax used by migrant workers, as well as the reasons behind its use, will help teachers to know how this should best be adapted to Standard Arabic, which in turn may help GPA speakers find better job opportunities.

The selected sampling process (snowball sampling) was initially chosen based on the reasons stated above in the sampling section; however, I faced the same problem that Almoaily mentioned in his study, namely, that ‘snowball sampling worked in very few cases and I had to start the process until I collected a sufficient amount of data’ (2013: 121). Thus, I adopted the second sampling procedure, which is similar to snowball sampling but differs from it in that the people I started with are locals who are not potential informants but are in contact with immigrant GPA speakers. I started by speaking with their bosses first, and I allowed them introduce me to the informants and explain to them the reasons, goals, and procedures of the study, while simultaneously giving the participants the opportunity to ask any questions related to the research. I found this procedure more effective and less time consuming. It seems that GPA speakers are more confident and less likely to feel that they are being personally monitored when the call to participate in the study is made through their employment.

The initial plan was to conduct interviews with 12 GPA-speaking female informants from six linguistic backgrounds (Tagalog, Punjabi, Sinhala, Malayalam, Sunda, and Bengali), half of whom had spent five years or less in the Gulf, while the other half had spent 10 years or more in the area at the time of interview. However, the ratio of data selection changed when data was collected due to circumstances beyond my control. I was unable to conduct interviews for two short-stay informants (one Bengali speaker and one Sunda speaker), as it was difficult to find them within the time limitations. Thus, I conducted interviews with only 10 GPA-speaking female informants, and I used a high-quality digital voice recorder[7]. To investigate whether the long-term residents have actually shifted towards GA, the structural patterns of GA that were collected from the newcomers of each language group were compared with those of long-term residents (e.g. newly settled Tagalog speakers vs. Tagalog speakers who had spent more than a decade in the Gulf). In other words, I compared their proportional use of GA tokens (Arabic definiteness markers, Arabic conjunction markers, object or possessive pronoun, GPA copula, and agreement in the VP and the NP) with those produced by their newly settled counterparts. In this phase, I also investigated whether there is any indication in the data that the participating informants use morphosyntactic features similar to the ones found in their L1s when they speak GPA. Section, 6.6.3.1.2 provides a detailed description of the informants participating in this study.

#### *6.6.3.1.2. Participants*

This section presents details of the exact length of each interview and the social background of every informant polled in this study. All the informants in this study were females who work in medium-income jobs (this is to make the sample of GPA speakers homogeneous in order to avoid any potential influence of social class discrepancies on linguistic production), and all the

[7] Olympus vn-7800pc

interviews were conducted in the Saudi Central Province where Najdi Arabic – a subdialect of GA – is spoken. However, as regards the informants' level of education, there was some variation (this is to avoid sampling bias, which limits the generalisability of findings because it is a threat to external validity). For instance, the informants who spoke Sinhala (S1) and Sunda (SU1) had not reached the secondary education level, while those who spoke Sinhala (S1), Malayalam (M2), and Bengali (B1) had completed their secondary education. Most Tagalog (1, 2), Punjabi (1, 2), and Malayalam (1) speakers, on the other hand, had completed their undergraduate studies (see Table 6.1). To achieve anonymity of the informants, the participants in this study have been labelled with one digit and one or two alphabetical codes. As Babbie (2015) states, pseudonyms should be used instead of their real names. The first letter of each label denotes the L1 of the informant: T = Tagalog, M = Malayalam, P = Punjabi, S = Sinhala, SU = Sunda, and B = Bengali. The number in the label distinguishes members of the same language group (e.g. T2 and T1 are two informants both speaking Tagalog). Table 6.1 lists the informants, their L1s, their age, the number of years spent in SA, the length of the interview, and the place of the interview. The data of long-term residents is shaded.

| Interviewee | L1 | L2 | Education | Age | Years in Saudi Arabia | Length of interview | Place of interview |
|---|---|---|---|---|---|---|---|
| **T2** | Tagalog | English | College | 39 | 10 | 16 min 48 s | Hospital in Riyadh |
| **T1** | Tagalog | English | College | 27 | 1.2 | 19 min 7 s | Hospital in Riyadh |
| **P2** | Punjabi | Urdu English | College | 46 | 15 | 18 min 36 s | Hospital in Riyadh |
| **P1** | Punjabi | Urdu | College | 27 | 1.5 | 22 min 41 s | Hospital in Riyadh |
| **S2** | Sinhala | | Primary | 50 | 13 | 19 min 46 s | Hospital in Riyadh |
| **S1** | Sinhala | Tamil English | Secondary | 33 | 8 m | 19 min 46 s | House maid in Riyadh |
| **M2** | Malayalam | Hindi | College | 47 | 25 | 15 min 25 s | Hospital in Riyadh |
| **M1** | Malayalam | Urdu | Secondary | 35 | 5 | 16 min 46 s | Hospital in Riyadh |
| **B2** | Bengali | Urdu English | Secondary | 35 | 10 | 21 min 45 s | Beauty centre in Riyadh |
| **SU2** | Sunda | Urdu English | Primary | 32 | 10 | 25 min 20 s | Beauty centre in Riyadh |

Table 6.1: Metadata of each informant in GPA corpus (all are females).

### 6.6.4. Conducting the Interviews – Pilot Phase

To conduct an interview, especially when human beings are involved as participants, there are ethical considerations. This section discusses two issues: ethics and the structure of the interviews (e.g. the type of questions to be asked and target data).

#### *6.6.4.1. Ethical Considerations*

One of the most essential issues for a researcher to consider before conducting face-to-face recorded interviews involving human beings is how to keep the participants safe and away from any potential harm (Abed, 2016). As reported by a number of researchers (Newman and Ratliff, 2001; Rice, 2006; Bera, 2004; Fraenkel & Wallen, 2006; Grey, 2014), it is the researcher's responsibility towards the research participants to ensure their confidentiality,

safety, and dignity, whilst taking into consideration that the participants have the right to withdraw from the study at any time. Rice (2006) believes that ethics in fieldwork studies are more significant than the goal of gaining new knowledge itself. In addition, according to Ruane (2005), the researcher must display high standards of professionality to ensure that all their informants agree to participate and be recorded in the study. Newing (2011) stated that researchers can find ethical principles and rules in professional academic institutions, such as the American Sociological Association's Code of Ethics (1997), to ensure that their research follows these rules. Moreover, the Data Protection Act (2018)[8] asserts that disclosing any personal details about individuals related to the study without their agreement is not permitted. Furthermore, participants should be notified that their personal information will be kept confidential.

I notified all participants (female GPA speakers) of the aim of the research in a simple and concise way, and I explained that participating in the interviews was completely voluntary, that they had the right not to take part in the study, and that they could stop the interview at any time. Once I received approval from their employers that the participants were allowed to participate in this study, I began to talk with them and obtained their verbal consent to take part in the interview. Since most of my speakers cannot read English or Arabic, the idea of signing a consent form leads to awkward situations and discomfort for my speakers; thus, there was no need to do it, as Naess (2008) also mentioned. To help participants relax and to encourage them to speak more, as well as more openly, I was sure to clarify that all their names and the names of their places of work will be anonymised throughout the study and that all the interview questions were unrelated to the nature of their work. Moreover, I informed them of the purpose

[8] Available at http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted

of this study and the way in which they could help me by agreeing to participate in the interviews. I explained that the intention was not to judge their Arabic speech, but instead to help them to speak Arabic as natives, improving their work opportunities in the Arab world or any place where the Arabic language is required. Naess (2008: 13) stated that in sociolinguistic research, it is common not to inform participants of the real purpose of the study in order to make them less aware of their speech. However, 'working on the assumption that the consultants' access to the regular Gulf Arabic register was limited, I felt that being honest with them would be unproblematic as well as ethically sound'.

As mentioned above, confidentiality was a primary concern within the study. All transcripts were only accessible to permitted individuals, and the interviewees were notified that all the recordings will be kept with the researcher to transcribe.

Another ethical point is from the Ethical Guidelines for Good Research Practice, composed by the Association of Social Anthropologists of the UK and Commonwealth[9]. Interviewers are obliged to pay back interviewees for their time and assistance: 'fair return should be made for their help and services' (cf. Grant and Sugarman, 2004). Almoaily adopted this method, which he found to be encouraging. Therefore, I decided to pay informants a small amount of money[10] upon completion of the interviews, as compensation for their time and cooperation during the interviews.

#### *6.6.4.2. Interview Protocol*

Before conducting the interviews, I prepared a list of questions[11] to ask the interviewees. These questions primarily made it possible to draw conclusions regarding gender and language variation in GPA resulting from the morphosyntactic differences in the female speakers' L1s

[9] Can be retrieved from https://www.theasa.org/downloads/ethics/Ethical_guidelines.pdf
[10] 10 SAR (approximately equal to 2 GBP).
[11] All the questions I used in this interview are in Appendix A.

and from their length of stay in the Gulf, both of which are potential factors leading to the emergence of GPA (i.e. universal and substratal factors). To achieve the goals of this research, Cohen, Manion, and Morrison (2007: 356) confirm that '[t]his needs to be done in such a way that the questions adequately reflect what it is the researcher is trying to find out'. Thus, the design of the questions was suitable and applicable to meet the researcher's objectives.

Part of the challenge of conducting an effective interview is composing the right interview questions. For the purpose of this study, the interview questions should be (1) relevant to the topic of research and to the research questions that the researcher seeks to answer; (2) designed to be open questions, as these help respondents to answer the questions in different ways and encourage them to explain or justify their answers, which in turn stimulates the respondents to provide long answers, giving the researcher more data to work with; (3) clear and simple questions to ensure that there will be no confusion or misunderstanding about the question, since respondents need to understand each question and know how to answer it, otherwise the researchers will not obtain the information they are seeking; (4) applicable, to the extent that the respondents have the knowledge to answer them and eliminate questions they cannot realistically answer; (5) unbiased, avoiding making any judgmental assumptions about either the subject of research or the respondent. Furthermore, the respondents should feel completely confident in answering the questions without any fear of a negative impact. I followed these guidelines[12] while writing the interview questions. Some of the questions were adapted from Almoaily (2013), Labov (1966), and Tagliamonte (2006), though some needed amending to better address female speakers, for example asking them if they can cook the same kinds of food here (in SA) as they did in their home countries, and how.

[12] Accessed on 9 January 2017 from https://courses.lumenlearning.com/ivytech-engl112/chapter/writing-effective-interview-questions/

The interview questions can be divided into two sections. The first set of questions is open-ended (qualitatively analysed), designed to stimulate the informants to answer them with a wide range of possible answers. The second set was purposefully formulated to elicit variants of a certain variable.

As suggested by Labov (1984), Kamal (1991), and Llamas (1999), the method of combining two types of questions (i.e. open-ended questions and questions eliciting certain linguistic features) is highly effective and widely adopted in language variation studies in the sociolinguistic field. Almoaily (2012) stated that 'it worked well for my purposes'. In addition, as one of my research aims is to explore gender variation in GPA by comparing my female corpus with Almoaily's male corpus, I tried to reach reliable results by sharing as similar a set of topics throughout the interviews as I possibly could.

The topics of these questions are varied, and some of them are about demographics, work life, school days, personal concerns, cultural traditions, and language (Tagliamonte, 2006). The first part of the interview is divided into the three subparts, according to Almoaily's (2012) classification: (a) identify the participants' demographic backgrounds, (b) identify their linguistic backgrounds, and (c) stimulate informants to produce long turns in the interviews. In addition to the open-ended questions, Almoaily also suggested preparing a PowerPoint presentation and asking informants to reflect on objects they saw in the presentation. This was a successful way to elicit tokens of linguistic phenomena, such as prepositions, gender, and the number of distinctions in demonstrative pronouns, which are expected to be rare in the informant's answers to the open questions.

Furthermore, the questions in the PowerPoint presentation were designed to elicit two tasks. In the first task, subjects were asked to name objects of different quantities and genders located at various distances using a demonstrative pronoun. The purpose of this task was to check the use of GA demonstrative pronouns by GPA speakers. In the second task, informants were asked

to mention the location of a subject positioned in various places on each slide. The purpose of the second task was to investigate GPA speakers' use of prepositions. The interviews began with the general questions in Section 1 (Parts I, II, III) and then moved on to Section 2 containing the PowerPoint slides[13] (Activity 1,2). The recorded interviews range from 16 to 27 minutes, with an overall time of 3 h 22 min (see Table 6.1). They were all recorded in MP3 audio format, using a high-quality digital recorder. The following sub sections discuss in detail the next step in building the corpus: transcribing the interviews.

### 6.6.5. Transcribing the Interviews – Pilot Phase

After conducting all the interviews, I began transcribing all the recordings made with the 10 female Asian workers. One of the most significant parts of a researcher's writing process is transcribing interviews. Although transcription is a challenge that most researchers in the field of linguistics face at some point, it is an efficient mode for data coding and analysis. It can help the researcher to code the data and to find or organise illustrative examples of code pieces better than the data from an audio or video format. Kvale and Brinkmann (2009: 178) emphasise the importance of transcribing recorded interviews, referring to the transcriptions as 'the solid rock-bottom empirical data of an interview project'. In addition, they demonstrate that transcribing interviews is time consuming – it might take up to five hours to transcribe one hour of speech, even with a highly skilled typist. Moreover, Powers (2005) assumes that in some cases, when there are many speakers in an interview, and when the researcher decides to perform phonetic transcription rather than etymological transcription, or simply in poor recording circumstances, the interview transcription process might take up to 24 hours to transcribe one hour. In addition, the Arabic language is even more challenging in that many conversations are carried out in dialects for which there is no standard written form. Lamel,

[13] Slides containing these two direct elicitation tasks are in Appendix A.

Gauvain, Adda, Adda-Decker, Canseco, Chen, and Schwenk (2004) expect that 'many more differences from broadcast speech and across languages will arise as our work progresses'.

In my case, I transcribed all the interviews myself; it took me nearly four hours to transcribe and revise only 10 minutes of speech. In fact, I tried a number of Arabic transcription/dictation software applications (free with no limit), such as *Speech to Text, Happy Transcribe, Talktyper, Speechlogger,* and *Speechnotes,* among others. Most natural language processing (NLP) research and tools are based on MSA, such as automatic speech recognition and language identification; however, I found these Arabic transcription tools to be inaccurate for non-Standard Arabic varieties or Arabic-based contact languages, and they were thus avoided in transcribing the data for the current project. Similarly, initial experiments with the FARASA toolkit showed that existing NLP tools (such as Arabic lemmatisers and POS taggers) were not useful for processing GPA. I used audio player software, such as Express Scribe Transcription Playback Software[14], to help me manually convert the spoken audio into text transcripts (Standard Arabic script). Although it was a challenging task, particularly where the recorded audio was imperfect (e.g. fast speech or multiple speakers talking simultaneously), I found it the best possible tool to enhance my abilities when transcribing the audio files. It was helpful in my transcription task, as it allowed me to focus on listening and recording the text accurately. Almoaily (2012: 127) suggests that 'the best procedure for storing and retrieving the data of P/C languages might be using the standard spelling of the lexifier language and supplementing that with digital audio recordings'. Therefore, in the case of GPA, the standard spelling of the lexifier language is Arabic. For that reason, the transcriptions of all the interviews are in Standard Arabic script (see Appendix C).

[14] Professional audio player software, with variable speed playback, designed to aid in the transcription of audio recordings (Free, cross platform).

As determined earlier in Section 6.3.3.2 and Figure 6.1, the standardised number of words in the corpus for the analysis was set at 1,000 words. This amount of data is more likely to be naturalistic and free from any factors that could affect the accuracy of the collected data, such as the effect of the observer's paradox, which was defined and discussed earlier in this chapter (Section 6.6.3); insufficient time to become accustomed to the interview's atmosphere and to feel comfortable; and interviewer or interviewee fatigue. All the transcriptions that contain irrelevant data, such as researcher turns and other turns produced by people who are not the interviewees (e.g. the interviewees' bosses; some patients, in cases where the interviews were conducted in the hospitals; clients; or customers), and any signs that do not refer to actual words, such as the laughter sign @ or researcher comments as a transcriber, should be deleted. The next subsection illustrates the processes of labelling, glossing, and counting the tokens.

### 6.6.6. Annotation and Counting the Tokens – Pilot Phase

This step is one of the greatest obstacles I faced when compiling and annotating the corpus written in Arabic script. Many dialects are written in different scripts, have no conventions for spelling, and have no large body of literature.

To count and retrieve the tokens from the transcribed interviews, I used the AntConc software[15], which is one of the best tools for analysing a corpus. Froehlich (2015) refers to AntConc as a useful toolkit for finding patterns in language that would be difficult to identify simply by reading the text.

As a first attempt, I labelled each variant of a variable with a unique Roman code (e.g. CONJ+ if the conjunction is used and CONJ- if the conjunction is dropped), as it allows for quick access to the required token. This attempt failed because the AntConc software was not able to detect the linguistic code-switching within Arabic script text accurately, since Arabic script

[15] AntConc, written by Laurence Anthony (2005). It is a freeware corpus analysis toolkit for concordance and text analysis.

goes from right to left, whereas English Roman script goes from left to right. To overcome these systematic changes in writing direction, I developed separate coding schema for each of the selected forms. Therefore, each form contains its own details about the coding schema and analyses. I retranscribed all of my corpus files in a unified spelling system by using an Arabic code instead of a Roman code for the annotation (e.g. روابط+ when the conjunction is used and - روابط when the conjunction is dropped). Figure 6.2 is as an example of a labelled interview before and after unifying the spelling system by using Arabic code instead of Roman code. This revised annotation worked well and was adopted in the main corpus. Some evidence exists that statistical NLP techniques for information retrieval (IR) on European languages generally do not transfer well to Arabic because of the nature of the language and its writing system (Yahya 1989; Hmeidi, Kanaan, and Evens 1997; De Roeck and Al-Fares 2000).

Here is an example of labelled interview before and after unify the spelling system by using Arabic code instead of Roman code:

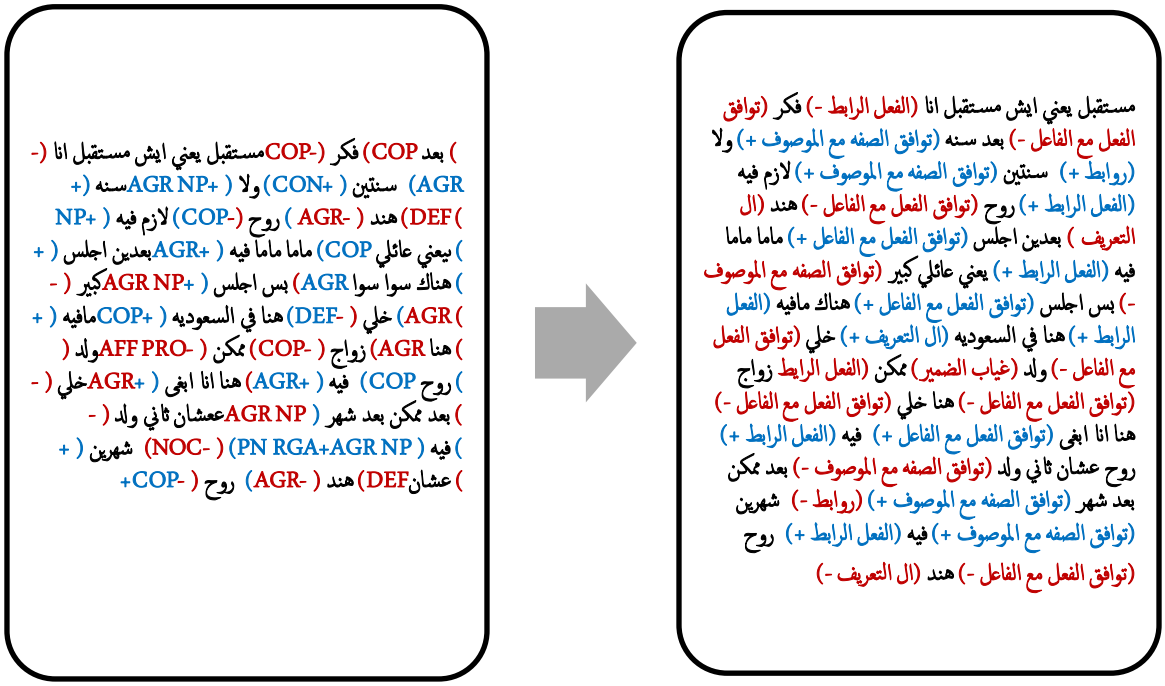



Figure 6.2: Example of labelled interview using Arabic code instead of Roman code from AntCoc.

In this excerpt, I asked M8 the following question: 'What plans do you have for the future?' She replied:

> What do you mean by future? Oh, my future. I think I have to go to India after one or two years to be with my mom and my big family. I will not be here, in SA, but I am going to leave my son here, in SA, because he might get married. I want to go to India also because of my other son. He will probably join the college after one or two months.

The new adopted codes I utilised to refer to each variant are illustrated in Table 6.2. These codes can also refer to tokens for GA features in the corpus of GPA, which I have denoted by the asterisk (*) next to their meaning, and the rest of codes/features[16] refer to GPA only.

| Code | Meaning |
|---|---|
| (AFF PRO +) منفصل ضمير | Object or possessive pronoun is used as a free morpheme |
| (AFF PRO –) متصل ضمير | Object or possessive pronoun is used as a bound morpheme * |
| (AFF PRO Ø) غياب الضمير | Possessive or object pronoun is dropped |
| (DEF +) التعريف + ال | The definite article is present * |
| (DEF –) التعريف - ال | The definite article is dropped |
| (CONJ) + روابط | The conjunction marker is present * |
| (– CONJ ) - روابط | The conjunction marker is dropped |
| (COP +) الرابط الفعل+ | The copula is used |
| (COP -) الرابط الفعل- | The copula is dropped * |
| (AGR +) الفاعل مع الفعل توافق+ | Verbal agreement is present * |
| (AGR -) الفاعل مع الفعل توافق- | The verbal agreement is missing |
| VØ غياب الفعل | The verb is dropped |
| (+ AGR NP) + توافق الصفه مع الموصوف | Agreement in the noun phrase or in the adjective phrase is present * |
| (- AGR NP) - توافق الصفه مع الموصوف | Agreement in the noun phrase or in the adjective phrase is missing |

Table 6.2: Codes and meanings of each linguistic feature.

[16] Some examples of the linguistic features of GPA labelled by these codes can be found in Section 4.1.2.

### 6.6.7. Quantification of Tokens – Pilot Phase

I used AntConc for analysing a corpus (see Figure 6.3 – a screenshot of spoken, transcribed, and labelled text for an Old Tagalog corpus). I tried to find the frequency of occurrence for every linguistic feature chosen in the study (e.g. *conjunction*). The concordance view showed me the chosen linguistic feature (e.g. *conjunction*) that appeared in my corpus (e.g. the Tagalog newcomers subcorpus), and some context thereof (such as a window of x words), which is called a 'key words in context' viewer. The (+/-) operator[17] (which finds zero or more characters) can also help – for instance, find both the present and the absent forms of the use of *conjunction* in the Tagalog subcorpus for newcomers (see Figure 6.4 – a screenshot of the old Tagalog corpus with the frequency of use of *conjunction* +روابط). I did the same for all corpus files that I had. Then, I calculated the percentage of tokens produced in every variant. For example, the variable conjunction in GPA has two variants: the prefix *wa* and Ø (i.e. dropping this prefix). Hence, if newly arrived Tagalog speakers drop the conjunction marker *wa* in 90% of the total number of tokens where they could use the conjunction marker, while long-staying Tagalog speakers drop it only in 60% percent of cases, then this can be taken as an indication that Tagalog speakers shift towards GA the longer they stay in the Gulf.

[17] Wildcard in AntConc search operators.

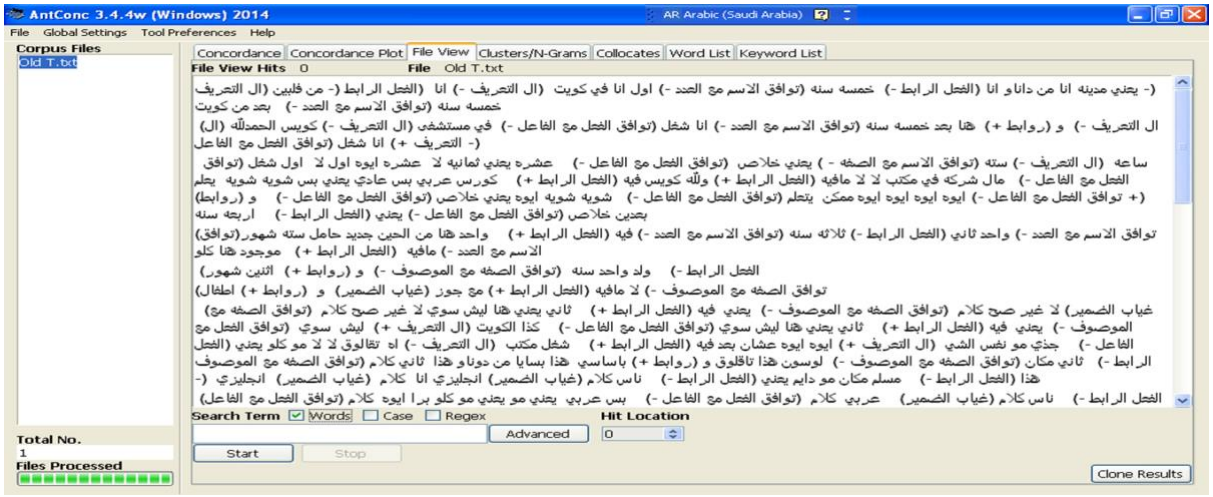



Figure 6.3: A screenshot of Old Tagalog corpus.

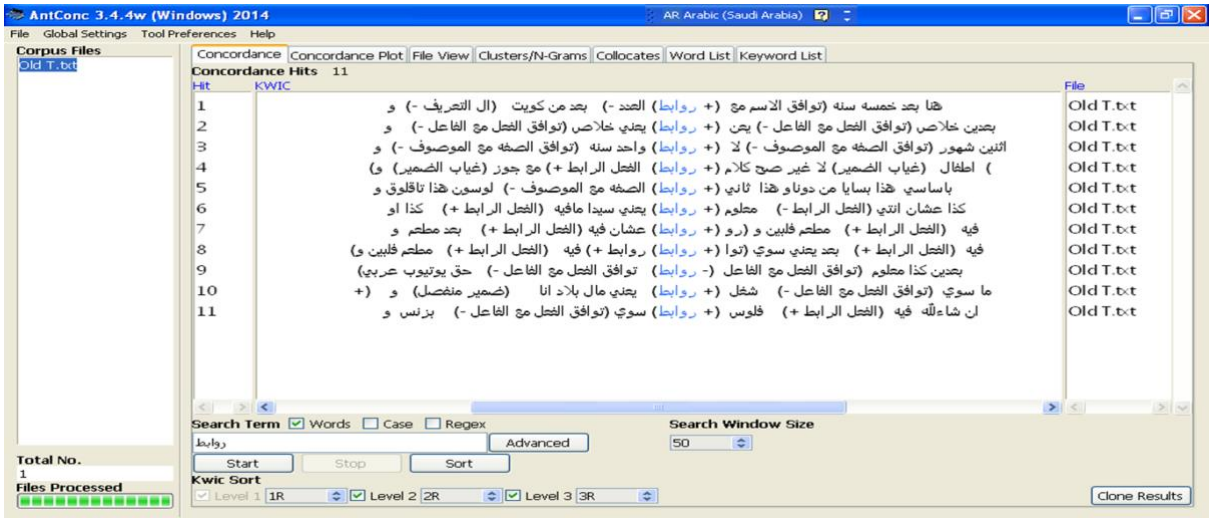



Figure 6.4: The use of conjunction marker *wa* (+روابط) concordance lines from a screenshot of an Old Tagalog corpus.

In terms of possible substrate languages' effect on GPA morphosyntax, if Punjabi speakers – who lack a copula in their L1, whereas Bengalis have one – are found to produce significantly fewer tokens of the GPA copula *fi* than Bengali speakers, this might be interpreted as a result of substrate influence. Regarding gender variation based on length of stay in the Gulf, I conducted a comparison analysis between my corpus (female GPA speakers) with that in Almoaily's corpus (male GPA speakers). If long-staying female Malayalam speakers in my study produce the definiteness marker *al-* more than long-staying male Malayalam speakers in Almoaily's corpus, then this is an indication that length of stay in the Gulf shows more

accommodation to standard GA in women than men. In addition, I compared the use of the given variant by members of a subgroup with that of other subgroups (e.g. newly settled female Tagalog speakers vs. long-term female Tagalog residents).

To statistically assess the significant differences between corpora, I used the Chi-squared test as one of the most powerful types of analysis. The advantage of this quantitative method is that it gives an idea of the use of a linguistic variant as compared to the other linguistic variants produced by a subgroup of speakers in the sample (see Lilliefors, 1967; Satorra and Bentler, 2001; Corder and Foreman, 2009).

To determine whether the hypotheses formulated in Section 6.2 can be accepted or rejected, **R** software[18] for statistical computing was used to run Chi-square tests. This test was used to establish the significance of the effect of the informants' L1 and years of residency in the Gulf on variation in GPA and on gender variation. The null hypothesis is rejected if the *p*-value is less than 0.05.

### 6.6.8. Reflections on the Pilot Study

The purpose of this section is to discuss the mechanics of analysis prior to the main study rather than statistical treatment, as the sample size of the pilot study in many cases was not sufficient to run any statistical tests. As detailed in Section 6.6.2, each language group was split into two groups based on informants' length of stay in SA or any other GA-speaking country (five years or less is referred to as 'New', and 10 years or more is referred to as 'Old'). Table 6.3 displays how the data is represented, and the number of GA linguistic tokens are compared to the GPA tokens that were produced by each informant per 1,000 words.

[18] R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. Page: 217.

| Interviewees (e.g. New Tagalog) | Variant 1 (GA feature) | Variant 2 (GPA feature) | Total |
|---|---|---|---|
| **Informant X** | Number of tokens (percentage[19]) | Number of tokens (percentage1) | Total of Variant 1 and Variant 2 tokens |
| **Informant Y** | Number of tokens (percentage) | Number of tokens (percentage) | Total of Variant 1 and Variant 2 tokens |
| **Average** | Average of tokens by X and Y (percentage) | Average of tokens by X and Y (percentage) | |

Table 6.3: Illustration of the results tables.

To contrast the proportionate use of GA variants as opposed to the proportionate use of GPA variants by each informant, I used the concordance program ***AntConc***. However, this package is not always optimal when it specifically deals with Arabic variety as I discussed above my experience with the FARASA toolkit showed that existing NLP tools were not useful for processing GPA: It may be inefficient, unclear, or missing steps, or it may produce typographic errors. The double-checking work enables me to improve or amend such processes and ensure accuracy and consistency. Indeed, I highly recommend supplementing this method with a manual one. In this regard, I performed the comparison of percentages of occurrence of each variable by myself.

In this section, I explain the process of outlining the observed linguistic variants for each linguistic feature under consideration, as Almoaily (2012) suggests, starting by listing the data of each GPA variant, for example definiteness, followed by conjunction markers; the copula; object and possessive; and agreement in the VP, in the NP, and in the ADJP. This process was applied on a series of tables which take the form exemplified in Table 6.3 above. Only a few tables are included in this section.

[19] The percentage in each cell represents the rate of occurrence of the token out of the total number of tokens for each variant.

#### *6.6.8.1.1. Variation in Definiteness*

GPA speakers use the GA definiteness marker (i.e. prefix *al-*) variably. The tables in this section (Tables 6.4 to 6.7) tabulate the rates of occurrence of the GA definiteness marker (presence versus absence) by L1 and length of stay in SA.

**1. Malayalam Informants**

The number of tokens where the GA definiteness marker is present and dropped in the data of Malayalam informants is presented in Table 6.4. The data reveals that there is a little variability between recently arrived and longer-term Malayalam residents (both ranging between 44.4% and 50%). However, Old Malayalam and New Malayalam speakers did not display any significant difference in the use of the GA definiteness marker (X-squared = 0.033144, df = 1, *p*-value = 0.8555).

| New Malayalam | Def. marker present (GA) | Def. marker missing (GPA) | Total | Old Malayalam | Def. marker present (GA) | Def. marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **M1** | 12 (44.4%) | 15 (55.5%) | 27 | M2 | 18 (50%) | 18 (50%) | 36 |

Table 6.4: Tokens of the definiteness marker *al-* by New and Old Malayalam informants.

**2. Punjabi Informants**

Tables 6.5 shows the use and dropping of the GA definiteness marker by the Punjabi informants. The data of the Punjabi informants reveals a noticeable difference in the use of the GA definiteness marker between new Punjabi and old Punjabi informants. P2 (old resident) used the GA definiteness marker less than P1 (new arrival) in 54.1% of cases compared to 29.2%.

| New Punjabi | Def. marker present (GA) | Def. marker missing (GPA) | Total | Old Malayalam | Def. marker present (GA) | Def. marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| P1 | 13 (54.1%) | 11 (46%) | 24 | P2 | 19 (29.2%) | 46 (71%) | 65 |

Table 6.5: Tokens of the definiteness marker *al-* by New and Old Punjabi informants.

### 3. Tagalog Informants

The instances of used and dropped GA *al-* markers by the Tagalog informants are displayed in Table 6.6. Tagalog speakers generally tend not to produce the definiteness marker *al-*. Note that the recently arrived Tagalog did not use the definiteness marker *al* at all (X-squared = 0.11272; df = 1; *p*-value = 0.7371, which was not statistically significant).

| New Tagalog | Def. marker present (GA) | Def. marker missing (GPA) | Total | Old Tagalog | Def. marker present (GA) | Def. marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| T1 | 0 (0%) | 5 (100%) | 5 | T2 | 4 (18%) | 18 (82%) | 22 |

Table 6.6: Tokens of the definiteness marker al- by New and Old Tagalog informants.

### 4. Sinhala Informants

Tables 6.7 lists the presence versus absence of the GA definite marker among the Sinhala participants. The data of the Sinhala informants also revealed that Sinhalese tend not to produce the definiteness marker *al*. Moreover, the recently arrived Sinhalese did not use the definiteness marker *al* at all (X-squared = NaN, df = 1, *p*-value = NA – the latter cannot be calculated due to division by zero).

| New Sinhalese | Def. marker present (GA) | Def. marker missing (GPA) | Total | Old Sinhalese | Def. marker present (GA) | Def. marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| S1 | 0 (0%) | 0 (0%) | 0 | S2 | 1 (3.3%) | 29 (97%) | 30 |

Table 6.7: Tokens of the definiteness marker *al-* by New and Old Sinhalese informants.

Overall, the data suggests that there is a slight variation in the definiteness marker among informants. A clear progression towards the use of the definiteness marker is observable in the data except for the Punjabi long-term resident (P2). In addition, with a small amount of data, the Chi-square test could not always be performed (it needs expected value counts of at least five per cell). These are just examples to demonstrate how the main data was handled. The same analysis process was adopted in the main phase for all five morphological features under investigation across members of the six L1 groups.

### 6.6.9. Conclusions and Recommendations

This section presents the conclusions reached on the basis of the pilot test, as well as my recommendations for Phase 2 of this study to extend the tested method to a larger and more representative sample of files. The key aim of this pilot test was to examine how to build and analyse a spoken GPA corpus for a sociolinguistic investigation. Indeed, I expected to face difficulties when deciding on size, balance, representativeness, and annotation of my spoken corpus. Compiling and analysing the corpus for this investigation were the most demanding and time-consuming tasks (see Section 6.3.3.3). First, choosing GPA speakers who meet certain criteria[20] and convincing them to participate in the interview was not easy. Many simply refused to be interviewed, and many others were too busy to take part in this study. Second, transcribing the interviews and choosing the appropriate transcription protocol for Arabic script presented great challenges. The strategy I employed to overcome or lessen the impact of these problems was to transcribe all of my corpus files in a unified spelling system by using Arabic code instead of Roman code. It was a fruitful technique (see Section 6.3.3.8).

To test 1) the influence of GPA speakers' L1 and length of stay in the Gulf on foreign female GPA speakers and 2) gender variation resulting from the morphosyntactic differences, in the main study I increased the sample and adopted all the recommendations mentioned above. A large sample enables more accurate tests of statistical significance.

## 6.7. The Main Data Collection Phase

The data for the current study (phase) comes from two sources: a primary source (sociolinguistic interviews) and a secondary source, namely, Almoaily's (2012) male corpus. The corpus of sociolinguistic interviews is comprised of spontaneous conversations that occurred in real-life settings between a) myself (the researcher), a native speaker of Najdi

[20] GPA should meet certain criteria (i.e. have spent five years or less, or 10 years or more in the Gulf and speak Bengali, Malayalam, Punjabi, Tagalog, Sinhala, or Sunda, as their first language).

Arabic and a fluent speaker of GPA, and b) female Asian workers. The two aforementioned corpora are detailed in the next subsection.

I followed the same procedures and steps used during the process of building the corpus in the pilot study, except for some changes, which are discussed in the following sections. The key purpose of the main data collection phase was to conduct real interviews that were developed during the pilot phase, to produce a more representative sample of female GPA speakers. This phase aimed to test all four hypotheses listed at the beginning of Chapter 6.

Based on my reflections from the pilot study, a number of factors were identified and modified in the design of the final phase of the current study. Van Teijlingen, Rennie, Hundley, and Graham (2002) stated that when pilot studies are conducted in academic papers and reports, researchers often claim that they have learned from undertaking the pilot study and have made the necessary changes, but they do so without offering the reader details on what exactly was learnt. Therefore, the following sections discuss the core changes that had to be made for the main data collection.

### 6.7.1. Increasing the Sample Size

During the process of compiling the GPA spoken corpus in the pilot phase, I faced a number of challenges that required rethinking approaches or redesigning experiments in the main phase. One of the main challenges was the large number of participants needed to produce a reasonable amount of data. The results of my initial quantitative analysis cannot be considered reliable, because of the small sample size. Since 'The appropriateness of the sample depends on factors including the purpose of the research and the domain within which the data is being collected' (Love, Dembry, Hardie, Brezina, and McEnery 2017), I increased the sample size to be able to test the statistical significance of the effect of the informants' L1 and years of residency in the Gulf on gender variation and language variation in GPA.

I also extended the period of the field study for data collection to cover a 3-month period, thereby allowing me to find all the participants who fulfilled certain criteria[21] and hence avoiding any distortion in the results. The field trip for the main study was undertaken in SA, Riyadh city, from February to April 2018 with a sample size of 72 interviews with GPA-speaking female informants, drawn from six linguistic backgrounds (Malayalam, Punjabi, Bengali, Tagalog, Sinhala, and Sunda). Half of the informants in the sample had spent five years or less in the Gulf, while the other half had spent 10 years or more in the area at the time they were interviewed.

Overall, the database contains 72,000 words: 2,000 words per substrate language, 1,000 of which were from recently settled informants and the remaining 1,000 words from long-term residents. The recorded interviews ranged from 16 to 35 minutes, with an overall time of 60 h 47 min (see Appendix B). The distribution of my data is illustrated in Figure 6.5, while Almoaily's data is shown in Figure 6.6.

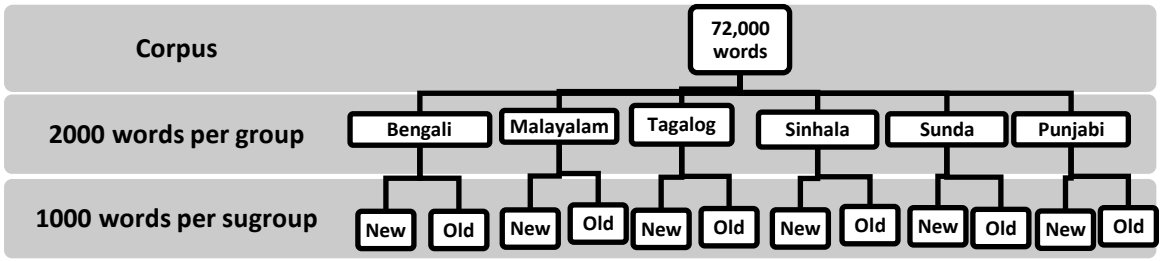



Figure 6.5: Distribution of my data.

[21] Female who speaks either Tagalog, Punjabi, Sinhala, Malayalam, Sunda, or Punjabi as her first language and has lived in the Gulf for either five years or less, or 10 years or more.

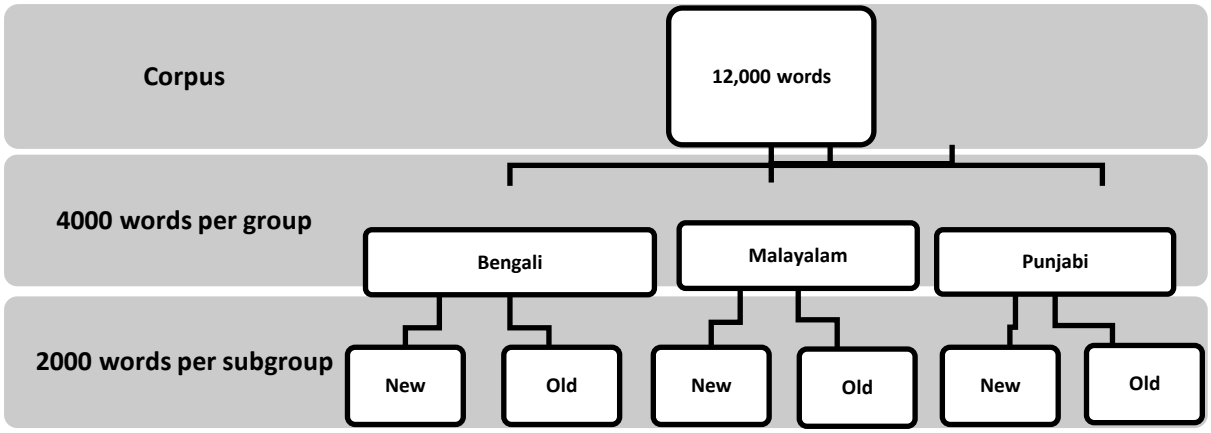


Figure 6.6: Distribution of Almoaily's data (adapted from Almoaily, 2012).

### 6.7.2. Overcoming Observer's Paradox

One of the most frequent problems that researchers face when conducting sociolinguistic interviews is the *observer's paradox*. Some new strategies and techniques I employed to overcome or lessen the impact of this phenomenon are discussed in this section.

As suggested by Huber (1999) and Labov (1972), one of the common ways to avoid the observer's paradox and motivate participants to take long turns speaking, is to ask questions that might cause participants to be less self-conscious and less aware that they are being recorded and linguistically observed. Labov (1972) provided some examples of questions, such as danger-of-death questions, which are listed at the end of this section. Although the reason for asking such a question was to determine the degree to which a person could remember the details of a horrible accident, this technique proved valid to make the subjects focus on remembering the details more than on their speech. These kinds of questions have been adopted by numerous researchers, such as Poplack (1989), Tagliamonte (2006), and Milroy and Gordon (2008).

Another technique to reduce the influence of the observer's paradox was suggested by Huber (ibid), who states that conducting interviews where participants work or live will relieve stress and anxiety and deliver more accurate data. Indeed, I found this method to be fruitful, since it makes interviewees feel more confident, especially when their bosses and their co-workers encourage them to be a part of the interviews.

A further method I used to reduce the impact of the observer's paradox was suggested by Almoaily (ibid): having a friendly open chat with the interviewee before starting to record the interview (e.g. asking them where they are from, what the name of their country or city is, whether their city is far away from SA, and how big this city is compared to where they live in SA). I found this method to be particularly helpful and effective for both the interviewer and the interviewee to familiarise themselves with each other before the actual start of the interview. In addition, engaging the interviewees on specific topics about the current events occurring in the region of SA proved to be helpful in reducing the effect of the observer's paradox as well as encouraging them to talk for longer. For example, my fieldwork trip took place after SA granted women the right to drive for the first time in decades, as SA had been unique in being the only country in the world where women were forbidden to drive motor vehicles (Laura, 2009). This topic was one of the most historic developments in Saudi women's lives. Hence, I added some questions to the interview schedule: 'Have you heard that SA lifted its ban on female drivers?', 'How do you feel about allowing women to drive? Are you happy or worried, and why?', and 'Do you plan to drive here in SA, or do you want to work as a driver?' The following subsections highlight the first step of creating the corpus: selecting informants.

### 6.7.3. Sampling

I concluded from the pilot study that the snowball sampling technique, which required asking permission and assistance from the informants' bosses to arrange interviews with those who were willing to participate in the study, is the best method out of the considered sampling strategies. Not only is this technique more effective and less time consuming, but GPA speakers are also more confident and less likely to feel that they are being personally monitored, especially when the call to participate in the study is made by their employers.

### 6.7.4. Stages of the Main Study

The fieldwork of the main data collection took place in Riyadh, the capital city of SA. I followed all the preparation processes that were adopted in the pilot study before and after conducting the interviews. I divided the process of finding the targeted participants and conducting the interviews into three stages during my stay in SA (i.e. 3 months spent in Riyadh). The first stage took place during the first month, when I visited hospitals, clinics, and health centres in which I expected to find a large number of female Asian workers. The second stage began at the start of the second month, when I went to shopping malls; beauty salons; some government facilities, such as schools and universities; and airports to conduct interviews. The last stage occurred in the third month of the data collection process, when the female participants were approached at their place of residence. I visited some residential buildings in West Riyadh where many female Asian workers reside, as well as many of my friends and family's houses where female domestic helpers work as nannies, housemaids, servants, and cooks. Nevertheless, some difficulties arose when conducting interviews. I expected that new speakers (informants who had been exposed to GPA for five years or less in the Gulf) would produce more pauses and speak slower than old speakers (informants who had been exposed to GPA for 10 years or more in the Gulf). Therefore, I increased the length of time spent in the interviews with new speakers by keeping the interview questions the same but allowing the interviewees to elaborate while responding to the questions. Another problem was that some people simply refused to be interviewed, and many others were too busy to take part in this study which made the process of finding the targeted participants still difficult.

### 6.7.5. Transcribing the Interviews – Main Phase

After conducting the interviews, I began transcribing all the recordings made with the 72 female Asian workers. As discussed in Section 6.6.5, the transcription of audio texts is one of the most challenging tasks a researcher can face. I transcribed all the interviews myself because GPA is

one of the Arabic dialects which is far removed from Standard Arabic, and thus no transcription or dictation software packages could be accurate to transcribe the data for the current project. Express Scribe Transcription Playback Software helped me manually convert spoken audio into text transcripts (for Standard Arabic script). I adopted standard spelling of the lexifier language, which is Arabic in the case of GPA, as suggested by Almoaily (2012: 127): 'the best procedure for storing and retrieving the data of P/C languages might be using the standard spelling of the lexifier language and supplementing that with digital audio recordings'. For that reason, the transcriptions of all interviews are in Standard Arabic script (see Appendix C). As determined earlier for the pilot study and in Figure 6.3 above, the standardised number of words in the corpus for the analysis was set at 1,000 words.

### 6.7.6. Annotation and Counting the Tokens – Main Phase

As was highlighted in the pilot study, annotating and counting the tokens in a corpus written in Arabic scripts is a significant challenge (see Section 6.3.3.9). This section presents the new technique that was tested in the pilot study to overcome the systematic changes in writing direction that AntConc software was not able to detect accurately. This problem was solved when retranscribing the corpus files in a unified spelling system by using an Arabic code instead of a Roman code for the annotation. This technique worked well in the pilot study and was thus adopted in my main corpus (refer to Table 6.2 for the codes and meanings associated with each linguistic feature).

### 6.7.7. Quantification of Tokens – Main Phase

To quantify the tokens, I followed all the recommendations and techniques used in the pilot study, employing **AntConc**, since it is one of the best tools for analysing a corpus. Furthermore, to statistically assess the differences between corpora, I used the Chi-squared test, which was

run by **R** (statistical computing software)[22]. This software could test whether the hypotheses formulated in Section 6.2 above are accepted or rejected. It was used to establish the significance of the effect of the informants' L1 and years of residency in the Gulf on language and gender variation in GPA. Next, Section 6.8 provides the conclusions of this chapter.

### 6.8. Conclusion to Chapter 6

The goal of this chapter was to outline the research method used to answer the research questions – both in the pilot study and in the main study. It also discussed the study participants, data collection, and interview questions. Both quantitative and qualitative approaches were employed to provide data that is as objective and accurate as possible. However, I experienced difficulty in mitigating the impact of some obstacles in the context of GPA. The first one was the difficulty in testing the amount of informants' exposure to GA (the superstrate language), which I expect to be one of the potential factors for language variation in GPA, as it might interfere with the results. Although I was aware of the need to select informants who are in direct contact with GA speakers (e.g. housemaids, nannies, and nurses), and avoided conducting interviews with those who have minimal or no exposure to GA (e.g. workers in companies or factories where English is the common language), differences in daily exposure to GA might still exist among the participants. Almoaily (ibid) states that 'polling from informants who have had exactly the same amount of exposure to Gulf Arabic, or even Gulf Pidgin Arabic, during their stay in SA seems impossible, especially for long-term residents'. The second obstacle relates to the characteristics of the participants themselves. I found it difficult to manage their personal traits, such as different aptitudes for language acquisition, different attitudes and willingness to learning GA, and different language-learning abilities. The results are presented in Chapter 7 and discussed in detail in Chapter 8.

---

[22] R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

# Chapter 7: Interview Results

In this chapter, I present the finding of my fieldwork. The chapter is divided into two main parts. The first part (Section 7.1) presents the corpus analysis results; the second part (Section 7.2) presents the findings from the corpus analysis as related to the research hypotheses presented in Section 6.2 and the obtained results listed in Section 7.1.

## 7.1. Corpus Analysis Results

To represent the findings of my fieldwork, I followed the same procedures described in Section 6.6.8. To statistically examine the differences between corpora, I used the **Chi-squared test**, which was run by **R** (statistical computing software). All resulting tables are presented in Sections 7.1, 7.2, and 7.3. This software tested whether the hypotheses are accepted or rejected, and it was used to establish the significance of the effect of the informants' L1 and years of residency in the Gulf on language and gender variation in GPA. In addition, I compared the percentages of occurrence of each variable manually.

The sections below outline the observed linguistic variants for each linguistic feature under consideration. The data contains information about each GPA variant for definiteness, followed by all other GPA variants; conjunction markers; the copula; object and possessive pronouns; and agreement in the VP, in the NP, and in the ADJP. The results from my data are shown in a series of tables which take the form exemplified in Table 6.3 above.

This section only outlines the data and briefly discusses some of the general patterns, Sections 7.2 and 7.3 of this chapter present the findings from the corpus analysis by running Chi-squared test in the light of the hypotheses.

### 7.1.1. Variation in Definiteness

GPA speakers variably produce the GA definiteness marker (i.e. the prefix *al)*. Tables 7.1 to 7.6 outline the rates of occurrence of the GA definiteness marker (presence versus absence) by ethnicity and length of stay in SA.

#### *7.1.1.1. Malayalam Informants*

The numbers of tokens where the GA definiteness marker is present and dropped in the data of Malayalam informants is presented in Table 7.1

| New Malayalam | Def. marker present (GA) | Def. marker missing (GPA) | Total | Old Malayalam | Def. marker present (GA) | Def. marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **M1** | 12 (44.4%) | 15 (55.5%) | 27 | **M7** | 27 (60%) | 18 (40%) | 45 |
| **M2** | 5 (20%) | 20 (80%) | 25 | **M8** | 42 (77.7%) | 12 (22.2%) | 54 |
| **M3** | 14 (25%) | 42 (75%) | 56 | **M9** | 16 (34.7%) | 30 (65.2%) | 46 |
| **M4** | 29(46.7%) | 33 (53.2%) | 62 | **M10** | 53 (70.6%) | 22 (29.3%) | 75 |
| **M5** | 18 (33.9%) | 35 (66.03%) | 53 | **M11** | 12 (26.6%) | 33 (73.3%) | 45 |
| **M6** | 30 (40.5%) | 44 (59.4%) | 74 | **M12** | 14 (34.1%) | 27 (65.8%) | 41 |
| **Average** | 18 (35.1%) | 31.5 (64.8%) | 49.5 | **Average** | 27.3(50.6%) | 23.6 (49.3) | 50.9 |

Table 7.1: Tokens of the definiteness marker al- by new and old Malayalam informants.

The data reveals that there is a little variability between the recently arrived and the longer-term Malayalam residents. However, the highest frequencies of the GA definiteness marker are produced by M10, a member of the old Malayalam group.

#### *7.1.1.2. Punjabi Informants*

The instances of used/dropped GA marker by the Punjabi informants is displayed in Table 7.2.

| New Punjabis | Def. marker present (GA) | Def. marker missing (GPA) | Total | Old Punjabis | Def. marker present (GA) | Def. marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **P1** | 11(20.7%) | 42(79.2%) | 53 | **P7** | 13(20.3 %) | 51(79.6%) | 64 |
| **P2** | 7(11.8%) | 52(88.1%) | 59 | **P8** | 7(13.7%) | 44(86.2%) | 51 |
| **P3** | 5(10.2%) | 44(89.7%) | 49 | **P9** | 6(10.9%) | 49(88%) | 55 |
| **P4** | 9(15.2%) | 50(84.7%) | 59 | **P10** | 8(12.6%) | 55(87.3%) | 63 |
| **P5** | 8(14.8%) | 46(85.1%) | 54 | **P11** | 10(18.5%) | 44(81.4%) | 54 |
| **P6** | 6 (15%) | 34 (85%) | 40 | **P12** | 8(12.6%) | 55(87.3%) | 63 |
| **Average** | 7.6(14.6) | 44.6(85.3%) | 52.2 | **Average** | 8.6 (14.8%) | 46.1 (85.1%) | 54.7 |

Table 7.2: Tokens of the definiteness marker *al-* by new and old Punjabi informants

The data of the Punjabi informants does not reveal noticeable variation between two the two groups of Punjabi informants.

### *7.1.1.3. Bengali Informants*

Table 7.3 depicts the use/dropping of the GA definiteness marker by the Bengali informants:

| New Bengalis | Def. marker present (GA) | Def. marker missing (GPA) | Total | Old Bengalis | Def. marker present (GA) | Def. marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **B1** | 2(4.8%) | 39(95.1%) | 41 | **B7** | 19(29.2%) | 46 (70.7%) | 65 |
| **B2** | 8(19.5%) | 33(80.4%) | 41 | **B8** | 12(26%) | 34(73.9%) | 46 |
| **B3** | 4(10.5%) | 34(89.4%) | 38 | **B9** | 22(30.9%) | 49(69%) | 71 |
| **B4** | 6(12.5%) | 42(87.5%) | 48 | **B10** | 18(35.2%) | 33(64.7%) | 51 |
| **B5** | 7(13.7%) | 44(86.2%) | 51 | **B11** | 15(28.8%) | 37(71.1%) | 52 |
| **B6** | 9(22.5%) | 31(77.5%) | 40 | **B12** | 24(35.8%) | 43(64.1%) | 67 |
| **Average** | 6 (14%) | 37(86.1%) | 43.1 | **Average** | 18.3(31%) | 40.3(68.9%) | 58.6 |

Table 7.3: Tokens of the definiteness marker *al-* by new and old Bengali informants

The data shows that Bengali speakers tend not to produce the definiteness marker *al-*. The highest rate of *al-* production is by B12 (36%). There is a difference between the newcomers and the long-term resident Bengalis in that old Bengalis use the definiteness marker more than new Bengalis.

### *7.1.1.4. Sinhala Informants*

The instances of used/dropped GA marker by the Sinhala informants is displayed in Table 7.4.

| New Sinhalese | Def. marker present (GA) | Def. marker missing (GPA) | Total | Old Sinhalese | Def. marker present (GA) | Def. marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **S1** | 0(0%) | 30(100%) | 30 | **S7** | 8(25.8%) | 23(74.1%) | 31 |
| **S1** | 5(12.8%) | 34(87.1%) | 39 | **S8** | 13(40.6%) | 19(59.3%) | 32 |
| **S3** | 2(4.6%) | 41(95.3%) | 43 | **S9** | 10(25%) | 30(75%) | 40 |
| **S4** | 4(9.3%) | 39(90.6%) | 43 | **S10** | 8(22.8%) | 27(77.1%) | 35 |
| **S5** | 11(26.1%) | 31(73.8%) | 42 | **S11** | 17(34.6%) | 32(65.3%) | 49 |
| **S6** | 8(16%) | 42(84%) | 50 | **S12** | 16(42.1%) | 22(57.8%) | 38 |
| **Average** | 5(11.4%) | 36.1(88.5%) | 41.1 | **Average** | 12(31.8%) | 25.5(68.1%) | 37.5 |

Table 7.4: Tokens of the definiteness marker *al-* by new and old Sinhala informants

The data of the Sinhala informants also reveals that old Sinhalese speakers tend to produce the definiteness marker *al-* more often than new Sinhalese speakers.

#### *7.1.1.5. Tagalog Informants*

The numbers of tokens where the GA definiteness marker is present and dropped in the data of Tagalog informants are presented in Table 7.5.

| New Tagalog | Def. marker present (GA) | Def. marker missing (GPA) | Total | Old Tagalog | Def. marker present (GA) | Def. marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **T1** | 1(4.3%) | 22(95.6%) | 23 | **T7** | 8(25%) | 24(75%) | 32 |
| **T2** | 0(0%) | 17(100%) | 17 | **T8** | 5(18.5%) | 22(81.4%) | 27 |
| **T3** | 0(0%) | 25(100%) | 25 | **T9** | 7(35%) | 13(65%) | 20 |
| **T4** | 3(7.5%) | 37(92.5%) | 40 | **T10** | 6(26%) | 17(73.9%) | 23 |
| **T5** | 2(10%) | 18(90%) | 20 | **T11** | 2(6.8%) | 27(93.1%) | 29 |
| **T6** | 2(7.6%) | 24(92.3%) | 26 | **T12** | 4(17.3%) | 19(82.6%) | 23 |
| **Average** | 1.3(4.9%) | 23.8(95%) | 25.1 | **Average** | 5.3(21.4%) | 20.3(78.5%) | 25.6 |

Table 7.5: Tokens of the definiteness marker *al-* by new and old Tagalog informants

Generally, Tagalog speakers tend to produce fewer tokens of the definiteness marker *al-* than speakers of the other languages. Note that the recently arrived Tagalog speakers T2 and T3 did not use the definiteness marker *al-* at all.

#### *7.1.1.6. Sundanese Informants*

Table 7.6 depicts the use/dropping of the GA definiteness marker by the Sundanese informants:

| New Sundanese | Def. marker present (GA) | Def. marker missing (GPA) | Total | Old Sundanese | Def. marker present (GA) | Def. marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **Su1** | 1(4.1%) | 23(95.8%) | 24 | **Su7** | 2(5.7%) | 33(94.2%) | 35 |
| **Su2** | 3(7.6%) | 36(92.3%) | 39 | **Su8** | 5(22.7%) | 17(77.2%) | 22 |
| **Su3** | 0(0%) | 22(100%) | 22 | **Su9** | 4(9.5%) | 38(90.4%) | 42 |
| **Su4** | 2(6.4%) | 29(93.5%) | 31 | **Su10** | 3(9.6%) | 28(90.3%) | 31 |
| **Su5** | 3(15%) | 17(85%) | 20 | **Su11** | 4(20%) | 16(80%) | 20 |
| **Su6** | 2(7.6%) | 24(92.3%) | 26 | **Su12** | 3(12%) | 22(88%) | 25 |
| **Average** | 1.8 (6.8%) | 25.1 (93.1%) | 26.9 | **Average** | 3.5(13.2%) | 25.6(86.7%) | 29.1 |

Table 7.6: Tokens of the definiteness marker *al-* by new and old Sinhalese informants

The data of the Sundanese informants also reveals that long-resident Sundanese speakers tend to produce the definiteness marker *al-* more compared with recently-arrived Sundanese speakers.

Overall, the results seem to confirm adaptation to GA. A clear progression towards use of the definiteness marker is observable especially amongst the Sinhalese, Tagalog, and Sundanese sample. In addition, Malayalam informants seem to use the definiteness marker slightly more than the other language sampls. These observations are discussed in more detail in Sections 8.2.2, 1 and 8.2.3, 1.

### 7.1.2. Variation in the Use of Conjunction Markers

This section discusses the use of conjunction markers amongst the GPA speakers in my corpus. Tables 7.8 to 7.13 list the instances where the informants used GA conjunction markers such as *aw* 'or' and *wa* 'and', compared to the number of cases where they produced juxtaposition.

#### *7.1.2.1. Malayalam Informants*

Table 7.7 lists the presence versus absence of GA conjunction markers in Malayalam language group.

| New Malayalam | Conj marker present (GA) | Conj marker missing (GPA) | Total | Old Malayalam | Conj marker present (GA) | Conj marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **M1** | 2 (7%) | 25 (92%) | 27 | **M7** | 2 (7.6%) | 24 (92.3%) | 26 |
| **M2** | 1 (5.2%) | 18 (94%) | 19 | **M8** | 7 (36.8%) | 12 (63.1%) | 19 |
| **M3** | 3 (11.1%) | 24 (88.8%) | 27 | **M9** | 4 (14%) | 23 (85.1%) | 27 |
| **M4** | 0 (0%) | 23(100%) | 23 | **M10** | 6 (18.7%) | 26(81.2%) | 32 |
| **M5** | 2 (8.6%) | 21(91%) | 23 | **M11** | 3 (10%) | 27(90%) | 30 |
| **M6** | 3(11.1%) | 24(88.8%) | 27 | **M12** | 5 (18.5%) | 22 (81.4%) | 27 |
| **Average** | 1.8 (7.2%) | 22.5 (92.5%) | 20.2 | **Average** | 4.5(17.7%) | 22.3(82.2%) | 21.8 |

Table 7.7: New and old Malayalam speakers' use of conjunction markers

The data reveals an increase in the use of conjunction markers by most of the old speakers except for M7 and M11.

### *7.1.2.2. Punjabi Informants*

Table 7.8 depicts the cases where Punjabi informants drop/use the GA conjunction markers.

| New Punjabis | Conj marker present (GA) | Conj marker missing (GPA) | Total | Old Punjabis | Conj marker present (GA) | Conj marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **P1** | 3 (11%) | 24 (89%) | 27 | **P7** | 7 (25.9%) | 20 (74%) | 27 |
| **P2** | 2 (8%) | 22 (92%) | 24 | **P8** | 11 (40.7%) | 16 (59.2%) | 27 |
| **P3** | 6 (38%) | 10 (63%) | 16 | **P9** | 10 (45.4%) | 12 (54.5%) | 22 |
| **P4** | 3 (14%) | 18 (86%) | 21 | **P10** | 5 (38.4%) | 8 (61.5%) | 13 |
| **P5** | 2 (6%) | 33 (94%) | 35 | **P11** | 2 (25%) | 6 (75%) | 8 |
| **P6** | 4 (13%) | 26 (87%) | 30 | **P12** | 6 (37.5%) | 10 (62.5%) | 16 |
| **Average** | 3.3 (15%) | 22.1 (84.9%) | 17.4 | **Average** | 6.8(35.5%) | 12 (64.4%) | 18.8 |

Table 7.8: New and old Punjabi speakers' use of conjunction markers

The data of the Punjabi informants also reveal noticeable variation within members of the same group. For example, P9 used the GA conjunction markers in 45% of all cases, while P7 and P11 only used them 25% of the time.

### *7.1.2.3. Bengali Informants*

Table 7.9 tabulates the tokens of dropping/uttering the GA conjunction marker among the Bengali language group:

| New Bengalis | Conj marker present (GA) | Conj marker missing (GPA) | Total | Old Bengalis | Conj marker present (GA) | Conj marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **B1** | 0 (0%) | 33 (100%) | 33 | **B7** | 8 (27%) | 22 (73%) | 30 |
| **B2** | 0 (0%) | 13 (100%) | 13 | **B8** | 9 (28%) | 23 (72%) | 32 |
| **B3** | 2 (10%) | 19 (90%) | 21 | **B9** | 5 (15%) | 29 (85%) | 34 |
| **B4** | 1 (5%) | 20 (95%) | 21 | **B10** | 10 (26%) | 28 (74%) | 38 |
| **B5** | 2 (7%) | 25 (93%) | 27 | **B11** | 2 (6%) | 34 (94%) | 36 |
| **B6** | 1 (4%) | 22 (96%) | 23 | **B12** | 6 (15%) | 33 (85%) | 39 |
| **Average** | 1 (4.3%) | 22 (95.6%) | 11.5 | **Average** | 6.6(19.4%) | 28.1(80.5%) | 19.8 |

Table 7.9: New and old Bengali speakers' use of conjunction markers

The data reveals an increase in the use of conjunction markers by the members of the old speakers except for B11, whereas the recently-arrived Bengali informants seem to produce very low token numbers of conjunction markers.

### *7.1.2.4. Sinhala Informants*

Table 7.10 lists the presence versus absence of GA conjunction markers in Sinhala language group.

| New Sinhalese | Conj marker present (GA) | Conj marker missing (GPA) | Total | Old Sinhalese | Conj marker present (GA) | Conj marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **S1** | 0 (0%) | 17 (100%) | 17 | **S7** | 4 (30.7%) | 9 (69.2%) | 13 |
| **S1** | 1(8.3%) | 11(91.6%) | 12 | **S8** | 6(37.5%) | 10(62.5%) | 16 |
| **S3** | 0(0%) | 14 (100%) | 14 | **S9** | 9(56.2%) | 7(43.7%) | 16 |
| **S4** | 2(20%) | 8(80%) | 10 | **S10** | 3(18.7%) | 13(81.2%) | 16 |
| **S5** | 2(18.1%) | 9 (81.8%) | 11 | **S11** | 5(55.5%) | 4(44.4%) | 9 |
| **S6** | 1(18.1%) | 13(92.8%) | 14 | **S12** | 3(25%) | 9(75%) | 12 |
| **Average** | 1(8.9%) | 12(91%) | 13 | **Average** | 5(37.3%) | 8.6(62.6%) | 13.6 |

Table 7.10: New and old Sinhalese speakers' use of conjunction markers

Both S1 and S3 did not use any of the GA conjunction markers, whereas the GA conjunction markers were used more by the old Sinhalese speakers.

### *7.1.2.5. Tagalog Informants*

Table 7.11 depicts the cases where Tagalog informants drop/use the GA conjunction markers.

| New Tagalog | Conj marker present (GA) | Conj marker missing (GPA) | Total | Old Tagalog | Conj marker present (GA) | Conj marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **T1** | 1 (3.2%) | 31(96.8%) | 32 | **T7** | 10 (58.8%) | 7(41.1%) | 17 |
| **T2** | 6(40%) | 9(60%) | 15 | **T8** | 6(27.2%) | 16(72.7%) | 22 |
| **T3** | 3(15%) | 17(85%) | 20 | **T9** | 21(77.7%) | 6(22.2%) | 27 |
| **T4** | 4(25%) | 12(75%) | 16 | **T10** | 5(21.7%) | 18(78.2%) | 23 |
| **T5** | 2(8.6%) | 21(91.3%) | 23 | **T11** | 12(46.1%) | 14(53.8%) | 26 |
| **T6** | 4(26.6%) | 11(73.3%) | 15 | **T12** | 17(65.3%) | 9(34.6%) | 26 |
| **Average** | 3.3(19.7%) | 16.8(80.2%) | 20.1 | **Average** | 11.8(49.5%) | 11.5(50.4%) | 23.3 |

Table 7.11: New andold Tagalog speakers 'use of conjunction markers

The data reveals a significant increase in the use of conjunction markers by long-term resident Tagalog speakers.

#### *7.1.2.6. Sundanese Informants*

Table 7.12 tabulates the tokens of dropping/uttering the GA conjunction marker among the Sundanese language group:

| New Sundanese | Conj marker present (GA) | Conj marker missing (GPA) | Total | Old Sundanese | Conj marker present (GA) | Conj marker missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **Su1** | 2 (9.5%) | 19(90.4%) | 21 | **Su7** | 4(26.6%) | 11(73.3%) | 15 |
| **Su2** | 3(15.7%) | 16(84.2%) | 19 | **Su8** | 6(66.6%) | 3(33.3%) | 9 |
| **Su3** | 6(42.8%) | 8(57.1%) | 14 | **Su9** | 5(35.7%) | 9(64.2%) | 14 |
| **Su4** | 5(31.2%) | 11(68.7%) | 16 | **Su10** | 7(63.6%) | 4(36.3%) | 11 |
| **Su5** | 8(66.6%) | 4(33.3%) | 12 | **Su11** | 4(33.3%) | 8(66.6%) | 12 |
| **Su6** | 4(23.5%) | 13(76.4%) | 17 | **Su12** | 5(33.3%) | 10(66.6%) | 15 |
| **Average** | 4.6(31.5%) | 11.8(68.3%) | 16.4 | **Average** | 5.1(43.2%) | 7.5(56.7%) | 12.6 |

Table 7.12: New and old Sundanese speakers' use of conjunction markers

Only Su5 produced a high percentage of conjunction markers similar to the percentage used by the old member Su10.

In general, the data in Tables 7.7 to 7.12 show a possible correlation between the length of stay and the use of conjunction markers in all the language groups.

### 7.1.3. Variation in Copula

As discussed in Chapter 4, while there is no copula in GA in the present tense, in GPA there is an optional copula *fi*. This section plots the occurrence of the copula *fi* across the speakers in my corpus, tabulated as presence versus absence. Tables 7.13 to 7.24 list the number of instances where the informants used the copula and compare this to the number of instances where the informants could have used the copula but did not use it.

#### *7.1.3.1. Malayalam Informants*

Tables 7.13 and 7.14 show the numbers of the use of/dropping the GPA copula *fi* in the Malayalam sample.

| New Malayalam | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | PRS | PST | PRS | PST | |
| **M1** | 33(56.8) | 15(68.1%) | 25(43.1%) | 7(31.8%) | PRS 58<br>PST 22 |
| **M2** | 49(56.3%) | 22(61.1%) | 38(43.6%) | 14(38.8%) | PRS 87<br>PST 36 |
| **M3** | 45(78.9%) | 5(71.4%) | 12 (21%) | 2(28.5%) | PRS 57<br>PST 7 |
| **M4** | 48(85.7%) | 10(100%) | 8(14.2%) | 0(0%) | PRS 56<br>PST 10 |
| **M5** | 33(82.5%) | 11 (10%) | 7(17.5%) | 0 (0%) | PRS 40<br>PST 11 |
| **M6** | 44(57.8%) | 21(65.6%) | 32(42.1%) | 11(34.3%) | PRS 76<br>PST32 |
| **Average** | 42(69.7%) | 14(77.7%) | 20.3(30.2%) | 5.6(22.2%) | 81.9 |

Table 7.13: New Malayalam speakers' use of the copula *fi.*

| Old Malayalam | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | PRS | PST | PRS | PST | |
| **M7** | 60(65.2%) | 25(69.4%) | 32(34.7%) | 11(30.5%) | PRS 92<br>PST 36 |
| **M8** | 77(71.2%) | 29(80.5%) | 31(28.7%) | 7(19.4%) | PRS108<br>PST36 |
| **M9** | 87(71.9%) | 15(53.5%) | 34(28%) | 13(46.4%) | PRS121<br>PST 28 |
| **M10** | 58(66.6%) | 17(62.9%) | 29(33.3%) | 10(37%) | PRS 87<br>PSP27 |
| **M11** | 67(65%) | 21(84%) | 36(34.9%) | 4(16%) | PRS103<br>PST 25 |
| **M12** | 88(71.5%) | 19(76%) | 35(28.4%) | 6(24%) | PRS123<br>PSP25 |
| **Average** | **72.8 (68.6%)** | **21(71%)** | **32.8 (31.3%)** | **8.5(28.9%)** | **135.1** |

Table 7.14: Old Malayalam speakers' use of the copula *fi.*

Dropping the copula both in the present and in the past tenses in the data of the Malayalam language group is more frequent than using it. In addition, it seems that there is no statistically significant difference in the way old and new speakers use the copula, as percentages are fairly similar.

### *7.1.3.2. Punjabi Informants*

The instances of dropping and retaining the GPA copula in the Punjabi language group is demonstrated in Tables 7.15 and 7.16 below:

| New Punjabis | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | PRS | PST | PRS | PST | |
| **P1** | 103(69.5%) | 22(78.5%) | 45(30.4%) | 6(21.4%) | PRS148 PST 28 |
| **P2** | 102(67.5%) | 14(63.6%) | 49(32.4%) | 8(36.3%) | PRS151 PST 22 |
| **P3** | 95(71.4%) | 33(89.1%) | 38(28.5%) | 4(10.8%) | PRS133 PST 37 |
| **P4** | 98(75.3%) | 30(88.2%) | 32(24.6%) | 4(11.7%) | PR130 PST 34 |
| **P5** | 105(70.4%) | 13(61.9%) | 44(29.5%) | 8(38%) | PRS149 PST 21 |
| **P6** | 88(73.9%) | 34(94.4%) | 31(26%) | 2(5.5%) | PRS119 PST 36 |
| **Average** | **98.5(71.3%)** | **24.3 (79.3%)** | **39.8 (28.6%)** | **5.3 (20.6%)** | **167.9** |

Table 7.15: New Punjabi speaker's use of the copula *fi.*

| Old Punjabis | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | PRS | PST | PRS | PST | |
| **P7** | 99(75%) | 29(87.8%) | 33(25%) | 4(12.1%) | PRS 132 PST 33 |
| **P8** | 101(72.6%) | 25(83.3%) | 38(27.3%) | 5(16.6%) | PRS 139 PST 30 |
| **P9** | 88(57.1%) | 19(59.3%) | 66(42.8%) | 13(40.6%) | PRS 154 PST 32 |
| **P10** | 78(56.5%) | 24(70.5%) | 60(43.4%) | 10(29.4%) | PRS 138 PST 34 |
| **P11** | 98(63.6%) | 26(68.4%) | 56(36.3%) | 12(31.5%) | PRS 154 PST 38 |
| **P12** | 77(57.8%) | 24(75%) | 56(42.1%) | 8(25%) | PRS 133 PST 32 |
| **Average** | **90.1(63.8%)** | **24.5(74%)** | **51.5(36.1%)** | **8.6(25.9%)** | **174.7** |

Table 7.16: Old Punjabi speaker's use of the copula *fi.*

The data reveals that members of the Punjabi language group drop the copula in more cases than they use it. The length of stay seems to have no influence on the informant on the use/dropping of the GPA copula.

### *7.1.3.3. Bengali Informants*

Tables 7.17 and 7.18 display the tokens of using and dropping the GPA copula by the Bengali informants.

| New Bengalis | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | PRS | PST | PRS | PST | |
| **B1** | 42(60.8%) | 12(80%) | 27(39.1%) | 3(20%) | PRS 69<br>PST 15 |
| **B2** | 35(63.6%) | 9(81.8%) | 20(36.3%) | 2(18.1%) | PRS 55<br>PST 11 |
| **B3** | 60(57.6%) | 10(62.5%) | 44(42.3%) | 6(37.5%) | PRS 104<br>PST 16 |
| **B4** | 102(76.1%) | 10(55.5%) | 32(23.8%) | 8(44.4%) | PRS 134<br>PST 18 |
| **B5** | 98(67.1%) | 7(50%) | 48(32.8%) | 7(50%) | PRS 146<br>PST 14 |
| **B6** | 37(62.7%) | 13(86.6%) | 22(37.2%) | 2(13.3%) | PRS 59<br>PST25 |
| **Average** | **62.3(64.6%)** | **10.1(69.4%)** | **32.1(35.3%)** | **4.6(30.5%)** | **141.2** |

Table 7.17: New Bengali speaker's use of the copula *fi*.

| Old Bengalis | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | **PRS** | **PST** | **PRS** | **PST** | |
| **B7** | 102(80.3%) | 16(66.65) | 25(19.6%) | 8(33.3%) | PRS 127<br>PST 24 |
| **B8** | 99(77.3%) | 26(83.8%) | 29(22.6%) | 5(16.1%) | PRS 128<br>PST 31 |
| **B9** | 96(77.4%) | 24(85.7%) | 28(22.5%) | 4(14.2%) | PRS 124<br>PST 28 |
| **B10** | 104(79.3%) | 14(63.6%) | 27(20.6%) | 8(36.3%) | PRS 131<br>PST 22 |
| **B11** | 112(83.5%) | 17(65.3%) | 22(16.4%) | 9(34.6%) | PRS 134<br>PST 26 |
| **B12** | 85(73.2%) | 22(88%) | 31(26.7%) | 3(12%) | PRS 116<br>PSP25 |
| **Average** | **99.6(78.5%)** | **19.8(75.5%)** | **27(21.4%)** | **6.1(24.4%)** | **152.5** |

Table 7.18: Old Bengali speaker's use of the copula *fi*.

The data for the Bengali informants show that the members of the old as well as the new group have a tendency to drop the copula both in the present and in the past tenses.

### *7.1.3.4. Sinhala Informants*

Tables 7.19 and 7.20 show the numbers of the use of/dropping the GPA copula *fi* in the Sinhala sample.

| New Sinhalese | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | PRS | PST | PRS | PST | |
| **S1** | 44(81.4%) | 1(100%) | 10(18.5%) | 0 (0%) | PRS 54<br>PST 1 |
| **S1** | 36(59%) | 3(37.5%) | 25(40.9%) | 5(62.5%) | PRS 61<br>PST 8 |
| **S3** | 52(53.6%) | 2(25%) | 45(46.3%) | 6(75%) | PRS 97<br>PST 8 |
| **S4** | 33(37.5%) | 1(100%) | 55(62.5%) | 0(0%) | PRS 88<br>PST 1 |
| **S5** | 37(57.8%) | 1(25%) | 27(42.1%) | 3(75%) | PRS 64<br>PST 4 |
| **S6** | 55(75.3%) | 4(66.6%) | 18(24.6%) | 2(33.3%) | PRS 73<br>PST 6 |
| **Average** | **42.8(60.7)** | **2(59%)** | **30(39.2%)** | **2.6(40.9%)** | **77.4** |

Table 7.19: New Sinha speaker's use of the copula *fi*.

| Old Sinhalese | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | PRS | PST | PRS | PST | |
| **S7** | 78(95.1%) | 7(77.7%) | 4(4.8%) | 2(22.2%) | PRS 82<br>PST 9 |
| **S8** | 66(85.7%) | 13(81.2%) | 11(14.2%) | 3(18.7%) | PRS 77<br>PST 16 |
| **S9** | 102(83.6%) | 6(75%) | 20(16.3%) | 2(25%) | PRS 122<br>PST 8 |
| **S10** | 69(74.1%) | 6(85.7%) | 24(25.8%) | 1(14.2%) | PRS 93<br>PST 7 |
| **S11** | 95(88.7%) | 10(83.3%) | 12(11.2%) | 2(16.6%) | PRS 107<br>PST 12 |
| **S12** | 111(92.5%) | 14(82.3%) | 9(7.5%) | 3(17.6%) | PRS 120<br>PST 17 |
| **Average** | **86.8(86.6%)** | **9.3(80.9%)** | **13.3(13.3%)** | **2.1(19%)** | **111.5** |

Table 7.20: Old Sinha speaker's use of the copula *fi*.

The data also shows a tendency to drop the copula in more cases than using it by the members of the Sinhala language group, both in the present and in the past tense. As regards the length of stay, it seems to have no effect on the informant on the use/dropping of the GPA copula.

### *7.1.3.5. Tagalog Informants*

The instances of dropping and retaining the GPA copula in the Tagalog language group is demonstrated in Tables 7.21 and 7. 22 below:

| New Tagalog | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | PRS | PST | PRS | PST | |
| **T1** | 38(67.8%) | 24(88.8%) | 18(32.1%) | 3(11.1%) | PRS 56<br>PST 27 |
| **T2** | 54(71%) | 4(66.6%) | 22(28.9%) | 2(33.3%) | PRS 76<br>PST 6 |
| **T3** | 49(79%) | 10(90.9%) | 13(20.9%) | 1(9%) | PRS 62<br>PST 11 |
| **T4** | 33(78.5%) | 18(90%) | 9(21.4%) | 2(10%) | PRS 42<br>PST 20 |
| **T5** | 52(89.6%) | 16(100%) | 6(10.3%) | 0(0%) | PRS 58<br>PST 16 |
| **T6** | 44(80%) | 12(75%) | 11(20%) | 4(25%) | PRS 55<br>PST 61 |
| **Average** | **45(77.6%)** | **14(85.2%)** | **13.1(22.3%)** | **2(14.7%)** | **74.1** |

Table 7.21: New Tagalog speaker's use of the copula *fi.*

| Old Tagalog | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | PRS | PST | PRS | PST | |
| **T7** | 44(60.2%) | 17(65.3%) | 29(39.7%) | 9(34.6%) | PRS 73<br>PST 26 |
| **T8** | 77(70%) | 22(78.5%) | 33(30%) | 6(21.4%) | PRS 110<br>PST 28 |
| **T9** | 64(75.2%) | 31(83.7%) | 21(24.7%) | 6(16.2%) | PRS 85<br>PST 37 |
| **T10** | 58(75.3%) | 14(77.7%) | 19(24.6%) | 4(22.2%) | PRS 77<br>PST 18 |
| **T11** | 62(69.6%) | 10(76.9%) | 27(30.3%) | 3(23%) | PRS 89<br>PST 13 |
| **T12** | 78(69%) | 24(96%) | 35(30.9%) | 1(4%) | PRS 113<br>PST 25 |
| **Average** | **63.8(69.9%)** | **19.6(79.7%)** | **27.3(30%)** | **4.8(20.2%)** | **115.5** |

Table 7.22: Old Tagalog speaker's use of the copula *fi.*

Both in the present and in the past tenses, the Tagalog informants seem to drop the copula more frequently than using it, although a minor increase was observed in the old informants' output.

### *7.1.3.6. Sundanese Informants*

Tables 7.23 and 7.24 display the tokens of using and dropping the GPA copula by the Sundanese informants.

| New Sundanese | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | PRS | PST | PRS | PST | |
| **Su1** | 38(65.5%) | 12(80%) | 20(34.4%) | 3(20%) | PRS 58<br>PST 15 |
| **Su2** | 54(71%) | 9(81.8%) | 22(28.9%) | 2(18.1%) | PRS 76<br>PST 11 |
| **Su3** | 60(58.8%) | 10(62.5%) | 42(41.1%) | 6(37.5%) | PRS 102<br>PST 16 |
| **Su4** | 101(75.3%) | 8(66.6%) | 33(24.6%) | 4(33.3%) | PRS 134<br>PST 12 |
| **Su5** | 52(54.1%) | 13(86.6%) | 44(45.8%) | 2(13.3%) | PRS 96<br>PST 15 |
| **Su6** | 96(75%) | 10(76.9%) | 32(25%) | 3(23%) | PRS 128<br>PST 31 |
| **Average** | **66.8(66.6%)** | **10.3(75.7%)** | **32.1(33.3%)** | **3.3(24.2%)** | **112.5** |

Table 7.23: New Sundanese speaker's use of the copula *fi*.

| Old Sundanese | Copula dropped (GA) | | Copula used (GPA) | | Total |
|---|---|---|---|---|---|
| | PRS | PST | PRS | PST | |
| **Su7** | 59(67.8%) | 24(85.7%) | 28(32.1%) | 4(14.2%) | PRS 87<br>PST 28 |
| **Su8** | 97(75.1%) | 23(82.1%) | 32(24.8%) | 5(17.8%) | PRS 129<br>PST 28 |
| **Su9** | 111(83.4%) | 31(91.1%) | 22(16.5%) | 3(8.8%) | PRS 133<br>PST 34 |
| **Su10** | 89(79.4%) | 26(86.6%) | 23(20.5%) | 4(13.3%) | PRS 112<br>PST 30 |
| **Su11** | 89(75.4%) | 22(78.5%) | 29(24.5%) | 6(21.4%) | PRS 118<br>PST 28 |
| **Su12** | 102(80.9%) | 26(89.6%) | 24(19%) | 3(10.3%) | PRS 126<br>PST 29 |
| **Average** | **91.1(77%)** | **25.3(85.6%)** | **26.3(22.9%)** | **4.1(14.3%)** | **146.8** |

Table 7.24: Old Sundanese speaker's use of the copula *fi*.

The data also shows a tendency to drop the copula in more cases than using it by the members of the Sundanese language group. There seems to be no effect of the length of stay of the informant on the use/dropping of the GPA copula.

In general, the data reveal that all informants drop the copula more often in the present tense, except S4, who uses the copula in 63% of the total number of cases and drops it in 38% of the cases. In the past tense, all informants, including S4, tend to drop the copula rather than retaining it. Both factors examined in this project – the linguistic background of the informant and their length of stay in Saudi Arabia – seem to have no effect on the use of the copula

among GPA speakers. More discussion on these findings can be found in Sections 8.2.2, 3 and 8.2.3, 3.

### 7.1.4. Variation in the Use of the Object and Possessive Pronouns

The use of object and possessive pronouns in GPA vary. There are four patterns of the possible structures in the pronominal system as Almoaily (2013: 140) suggests. To narrow them down; the presence or absence of the pronoun as well as the type of morphology (i.e. bound versus free). The four possible variants to which Almoaily refers are:

1. **AGR+ Bound**: The agreeing object or possessive pronoun is attached to the verb, noun, or preposition as a suffix, as in GA (e.g. kitab-***i*** 'book-**my'**).

2. **AGR-Bound pro:** A possessive or object pronoun is attached as a suffix but does not agree with the noun. (e.g. inta yiti-***ik*** *fuluus* 'you give-**you'** [instead of the GA suffixed pronoun ***-ni***: 'me'])

3. **Free morph:** The subject form of the object or possessive pronoun is used (subject forms are free morphemes), e.g. *Inti qoul **ana*** 'you telling I' (instead of *qoulti-li* 'told.PST-1SG.OJB PRO' in GA).

4. **Dropped**: The object or possessive pronoun is dropped, e.g. *zawj-**Ø** mawjod hina* 'husband-**Ø** is here [instead of *zawj-i mawjod hina* 'my husband is here'].

The instances of (1-4) found in my data are tabulated in Tables 7.25 to 7.36 below.

#### *7.1.4.1. Malayalam Informants*

Tables 7.25 and 7.26 show the instances of the four variants for the GPA possessive and object pronoun in the Malayalam language group (the percentages of object and possessive pronouns are calculated separately).

| New Malayalam | AGR+ Bound (GA) | | AGR- Bound pro (GPA) | | Free morph. (GPA) | | Dropped (GPA) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | POSS | OBJ | POSS | OBJ | POSS | OBJ | POSS | OBJ | |
| **M1** | 0(0%) | 2(66.6%) | 0(0%) | 0(0%) | 1(100%) | 0(0%) | 0(0%) | 1(33.3%) | OBJ 3 POSS 1 |
| **M2** | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 2(100%) | 0(0%) | 1(33.3%) | 1(100%) | OBJ 1 POSS 3 |
| **M3** | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 1(50%) | 0(0%) | 1(50%) | 2(100%) | OBJ 2 POSS 2 |
| **M4** | 1(16.6%) | 0 (0%) | 0(0%) | 0(0%) | 3(50%) | 0(0%) | 2(33.3%) | 2(100%) | OBJ 2 POSS 6 |
| **M5** | 0 (0%) | 0(0%) | 0(0%) | 0(0%) | 1(100%) | 0(0%) | 0(0%) | 3(100%) | OBJ 3 POSS 1 |
| **M6** | 1(7.6%) | 6 (75%) | 0(0%) | 0(0%) | 9(69.2%) | 1(12%) | 3(23%) | 1(12.5%) | OBJ 8 POSS13 |
| **Average** | 0.3(4%) | 1.3(23.6%) | 0(0%) | 0(0%) | 2.8(78.2%) | 0.1(2%) | 1.1(23%) | 1.6(74.3%) | 7.2 |

Table 7.25: New Malayalam speaker's use of object and possessive pronouns.

| Old Malayalam | AGR+ Bound (GA) | | AGR- Bound pro (GPA) | | Free morph. (GPA) | | Dropped (GPA) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | POSS | OBJ | POSS | OBJ | POSS | OBJ | POSS | OBJ | |
| **M7** | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 1(20%) | 0(0%) | 4(80%) | 0(0%) | OBJ 0 POSS 5 |
| **M8** | 3(42. %) | 0 (0%) | 0(0%) | 0(0%) | 4(57.1%) | 0(0%) | 0(0%) | 4(100%) | OBJ 4 POSS 7 |
| **M9** | 4(40%) | 2(33.3%) | 0(0%) | 2(33.3%) | 6(60%) | 1(16.6%) | 0(0%) | 1 (16.6%) | OBJ6 POSS10 |
| **M10** | 2(33. %) | 1(33.3%) | 0(0%) | 0(0%) | 4(66.6%) | 0(0%) | 0(0%) | 2 (66.6%) | OBJ3 POSS6 |
| **M11** | 2(18. %) | 2(66.6%) | 0(0%) | 1(33.3%) | 5(45.4%) | 0(0%) | 4(36.3%) | 0(0%) | OBJ3 POSS11 |
| **M12** | 3(37.5) | 1(20%) | 0(0%) | 2 (40%) | 3(37.5%) | 1(20%) | 2(25%) | 1(20%) | OBJ 5 POSS 8 |
| **Average** | 2.3 (29%) | 1 (25.5%) | 0 (0%) | 0.8 (18%) | 3.8 (47.7%) | 0.3 (6.1%) | 1.6 (23.5%) | 1.3 (33.8%) | 11.1 |

Table 7.26: old Malayalam speaker's use of object and possessive pronouns.

Although there is a great variation between members of the same group, wherefor instance, dropped object pronouns in the new group range between 100% and 33% while instances in the old group range between 100% and 66%, the numbers are extremely low, especially those in Table 7.25.

### *7.1.4.2. Punjabi Informants*

Tables 7.27 and 7.28 present the results for the Punjabi language group.

| New Punjabis | AGR+ Bound (GA) | | AGR- Bound pro (GPA) | | Free morph. (GPA) | | Dropped (GPA) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | **POSS** | **OBJ** | **POSS** | **OBJ** | **POSS** | **OBJ** | **POSS** | **OBJ** | |
| **P1** | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 7(53.8%) | 0(0%) | 6(46.1%) | 8(100%) | OBJ 8 POSS13 |
| **P2** | 1(16.6%) | 0(0%) | 0(0%) | 0(0%) | 2(33.3%) | 5(50%) | 3(50%) | 5(50%) | OBJ 10 POSS 6 |
| **P3** | 0 (0%) | 0 (0%) | 0(0%) | 0(0%) | 5(55.5%) | 0 (0%) | 4(44.4%) | 8(100%) | OBJ 8 POSS 9 |
| **P4** | 1(9%) | 0(0%) | 0(0%) | 0(0%) | 8(72.7%) | 2(25%) | 2(18.1%) | 6(75%) | OBJ 8 POSS11 |
| **P5** | 2(22.2%) | 1(8.3%) | 0(0%) | 0(0%) | 2(22.2%) | 2(16.6%) | 5(55.5%) | 9(75%) | OBJ 12 POSS 9 |
| **P6** | 1(11.1%) | 1(10%) | 0(0%) | 0(0%) | 4(44.4%) | 1(10%) | 4(44.4%) | 8(80%) | OBJ 10 POSS 9 |
| Average | 0.8(9.8%) | 0.3(3%) | 0(0%) | 0(0%) | 4.6(47%) | 1.6(16.9%) | 4(43.1%) | 7.3(80%) | 18.6 |

Table 7.27: New Punjabis speaker's use of object and possessive pronouns.

| Old Punjabis | AGR+ Bound (GA) | | AGR- Bound pro (GPA) | | Free morph. (GPA) | | Dropped (GPA) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | POSS | OBJ | POSS | OBJ | POSS | OBJ | POSS | OBJ | |
| **P7** | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 11(55%) | 2(40%) | 9(45%) | 3(60%) | OBJ 5 POSS10 |
| **P8** | 0(0%) | 0(0%) | 0(0%) | 1(33.3%) | 2(50%) | 2(66.6%) | 2(50%) | 0(0%) | OBJ 3 POSS 4 |
| **P9** | 2(9.5%) | 0(0%) | 0(0%) | 1(100%) | 10(47.6%) | 0(0%) | 9(42.8%) | 0(0%) | OBJ 1 POSS11 |
| **P10** | 2(9%) | 0(0%) | 0(0%) | 2(66.6%) | 7(31.8%) | 1(33.3%) | 13(59%) | 2(66.6%) | OBJ 5 POSS12 |
| **P11** | 1(6.25%) | 0(0%) | 0(0%) | 0(0%) | 13(81.2%) | 3(60%) | 2(12.5%) | 2(40) | OBJ 5 POSS16 |
| **P12** | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 3(27.2%) | 2(66.6%) | 8(72.7%) | 1(33.3%) | OBJ 3 POSS11 |
| Average | 0.8(4.1%) | 0(0%) | 0(0%) | 0.6(33.3%) | 7.6(48.8%) | 1.6(39.4%) | 7.1(47%) | 1.3(33.3%) | 19 |

Table 7.28: Old Punjabis speaker's use of object and possessive pronouns.

The data shows that the newly arrived Punjabi informants seem either to use possessive pronouns as free morphemes or drop them, and to drop object pronouns, while old Punjabis drop possessive pronouns more than using them as either free or bound morphemes. It looks as if no general pattern can be found for object pronouns in the old Punjabi informants' data. This result is also (at least partly) due to the fact that the numbers are so low.

### *7.1.4.3. Bengali Informants*

The occurrences of the four variants of the possessive and object pronouns by the Bengali informants are displayed in Tables 7.29 and 7.30:

| New Bengalis | AGR+ Bound (GA) | | AGR- Bound pro (GPA) | | Free morph. (GPA) | | Dropped (GPA) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | POSS | OBJ | POSS | OBJ | POSS | OBJ | POSS | OBJ | |
| **B1** | 2(18.1%) | 0(0%) | 0(0%) | 0(0%) | 8(72.7%) | 2(50%) | 1(9%) | 2(50%) | OBJ 4 POSS11 |
| **B2** | 1(10%) | 1(12.5%) | 0(0%) | 0(0%) | 5(50%) | 0(0%) | 4(40%) | 7(87.5%) | OBJ 8 POSS10 |
| **B3** | 1(20%) | 0(0%) | 0(0%) | 0(0%) | 4(80%) | 1(50%) | 0(0%) | 1(50%) | OBJ 2 POSS 5 |
| **B4** | 3(23%) | 1(12.5%) | 0(0%) | 0(0%) | 5(38.4%) | 1(12.5%) | 5(38.4%) | 6(75%) | OBJ 8 POSS13 |
| **B5** | 5(26.3%) | 0(0%) | 0(0%) | 0(0%) | 12(63.1%) | 3(50%) | 2(10.5%) | 3(50%) | OBJ 6 POSS19 |
| **B6** | 2(16.6%) | 0(0%) | 0(0%) | 0(0%) | 9(75%) | 0(0%) | 1(8.3%) | 1(100%) | OBJ 1 POSS12 |
| **Average** | 2.3(19%) | 0.3(4.1%) | 0(0%) | 0(0%) | 7.1(63.2%) | 1.1(21.7%) | 2.1(17.7%) | 3.3(68.7%) | 16.2 |

Table 7.29: New Bengalis speaker's use of object and possessive pronouns.

| Old Bengalis | AGR+ Bound (GA) | | AGR- Bound pro (GPA) | | Free morph. (GPA) | | Dropped (GPA) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | POSS | OBJ | POSS | OBJ | POSS | OBJ | POSS | OBJ | |
| **B7** | 3(33.3%) | 1(16.6%) | 0(0%) | 0(0%) | 4(44.4%) | 3(33.3%) | 2(22.2%) | 2(33.3%) | OBJ 6 POSS9 |
| **B8** | 3(42.8%) | 0(0%) | 0(0%) | 1(20%) | 3(42.8%) | 2(40%) | 1(14.2%) | 2(40%) | OBJ 5 POSS7 |
| **B9** | 2(50%) | 2(40%) | 0(0%) | 0(0%) | 2(50%) | 1(20%) | 0(0%) | 2(40%) | OBJ 5 POSS4 |
| **B10** | 7(58.3%) | 1(10%) | 0(0%) | 2(20%) | 3(25%) | 3(30%) | 2(16.6%) | 4(40%) | OBJ10 POSS12 |
| **B11** | 3(25%) | 1(12.5%) | 0(0%) | 1(12.5%) | 6(50%) | 4(50%) | 3(25%) | 2 (25%) | OBJ 8 POSS12 |
| **B12** | 4(33.3%) | 0(0%) | 0(0%) | 0(0%) | 5(41.6%) | 3(60%) | 3(25%) | 2(40%) | OBJ 5 POSS12 |
| **Average** | 3.6 (40.4%) | 0.8 (13.1%) | 0 (0%) | 0.6 (8.75%) | 3.8 (42.3%) | 2.6 (38.8%) | 1.8 (17.1%) | 2.3 (36.3%) | 16.2 |

Table 7.30: Old Bengali speaker's use of object and possessive pronouns.

New Bengali speakers tend to have demonstrated a tendency to drop object pronouns and the use of possessive pronouns as free morphemes, while the old members are more likely to use object and possessive pronouns as free morphemes rather than dropping them or using them as bound morphemes.

#### *7.1.4.4. Sinhalese Informants*

Tables 7.31 and 7.32 show the instances of the four variants for the GPA possessive and object pronoun in the Sinhala language group.

| **New Sinhalese** | **AGR+ Bound (GA)** | | **AGR- Bound pro (GPA)** | | **Free morph. (GPA)** | | **Dropped (GPA)** | | **Total** |
|---|---|---|---|---|---|---|---|---|---|
| | **POSS** | **OBJ** | **POSS** | **OBJ** | **POSS** | **OBJ** | **POSS** | **OBJ** | |
| **S1** | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 0(0%) | OBJ 0 POSS 0 |
| **S1** | 4(50%) | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 4(50%) | 0(0%) | OBJ 0 POSS 8 |
| **S3** | 3(33.3%) | 1(16.6%) | 0(0%) | 0(0%) | 4(44.4%) | 3(50%) | 2(22.2%) | 2(33.3%) | OBJ 6 POSS 9 |
| **S4** | 2(50%) | 2(40%) | 0(0%) | 0(0%) | 2(50%) | 1(20%) | 0(0%) | 2(40%) | OBJ 5 POSS 4 |
| **S5** | 3(33.3%) | 1(16.6%) | 0(0%) | 0(0%) | 4(44.4%) | 3(50%) | 2(22.2%) | 2(33.3%) | OBJ 6 POSS19 |
| **S6** | 1(20%) | 0(0%) | 0(0%) | 0(0%) | 4(80%) | 1(50%) | 0(0%) | 1(50%) | OBJ 2 1POSS5 |
| **Average** | 2.1(31.1%) | 0.6(12.2%) | 0(0%) | 0(0%) | 2.3(36.4%) | 1.3(28.3%) | 1.3(15.7%) | 1.1(26.1%) | 8.7 |

Table 7.31: New Sinhalese speakers' use of object and possessive pronouns.

| **Old Sinhalese** | **AGR+ Bound (GA)** | | **AGR- Bound pro (GPA)** | | **Free morph. (GPA)** | | **Dropped (GPA)** | | **Total** |
|---|---|---|---|---|---|---|---|---|---|
| | **POSS** | **OBJ** | **POSS** | **OBJ** | **POSS** | **OBJ** | **POSS** | **OBJ** | |
| S7 | 2(25%) | 1(10%) | 0(0%) | 0(0%) | 6(75%) | 7(87.5%) | 0(0%) | 2(20%) | OBJ 10 POSS 8 |
| S8 | 2(33.3%) | 0(0%) | 0(0%) | 1(33.3%) | 2(33.3%) | 1(33.3%) | 2(33.3%) | 1(33.3%) | OBJ 3 POSS 6 |
| S9 | 2(50%) | 0(0%) | 0(0%) | 0(0%) | 2(50%) | 1(50%) | 0(0%) | 1(50%) | OBJ 2 POSS 4 |
| S10 | 1(11.1%) | 1(10%) | 0(0%) | 1(10%) | 8(88.8%) | 5(50%) | 0(0%) | 3(30%) | OBJ 10 POSS 9 |
| S11 | 3(37.5%) | 0(0%) | 0(0%) | 1(33.3) | 4(50%) | 1(33.3%) | 1(12.5%) | 1(33.3%) | OBJ 3 POSS 8 |
| S12 | 3(33.3%) | 1(8.3%) | 0(0%) | 1(8.3%) | 6(66.6%) | 7(58.3%) | 0(0%) | 3(25%) | OBJ 12 1POSS9 |
| **Average** | 2.1(31.7%) | 0.5(4.7%) | 0(0%) | 0.6(14.1%) | 4.6(60.6%) | 3.6(52%) | 0.5(7.6%) | 1.8(31.9%) | 13.7 |

Table 7.32: Old Sinhalese speakers' use of object and possessive pronouns.

The long-term residents among the Sinhalese informants tend to use the possessive and object pronouns as free morphemes more than dropping them, while the new Sinhalese use possessive pronouns as free morphemes more the use of object pronouns.

### *7.1.4.5. Tagalog Informant*

Tables 7.33 and 7. 34 present the results for the Tagalog language group.

| New Tagalog | AGR+ Bound (GA) | | AGR- Bound pro (GPA) | | Free morph. (GPA) | | Dropped (GPA) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | POSS | OBJ | POSS | OBJ | POSS | OBJ | POSS | OBJ | |
| **T1** | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 1(100%) | 1(100%) | 0(0%) | OBJ 1 POSS 1 |
| **T2** | 1(14.2%) | 0(0%) | 0(0%) | 0(0%) | 5(71.4%) | 1(33.3%) | 1(14.2%) | 2(66.6%) | OBJ3 POSS 7 |
| **T3** | 3(33.3%) | 0(0%) | 0(0%) | 0(0%) | 3(33.3%) | 0(0%) | 3(33.3%) | 2(100%) | OBJ 2 POSS 9 |
| **T4** | 2(25%) | 0(0%) | 0(0%) | 0(0%) | 2(25%) | 1(33.3%) | 4(50%) | 2(66.6%) | OBJ3 POSS 8 |
| **T5** | 2(22.2%) | 0(0%) | 0(0%) | 0(0%) | 2(22.2%) | 1(50%) | 5(55.5%) | 1(50%) | OBJ2 POSS 9 |
| **T6** | 5(33.3%) | 0(0%) | 0(0%) | 0(0%) | 4(26.6%) | 0(0%) | 6(40%) | 3(100%) | OBJ3 POSS15 |
| Average | 2.1(21.3%) | 0(0%) | 0(0%) | 0(0%) | 2.6(29.7%) | 0.6(36.1%) | 3.3(48.8%) | 1.6(63.8%) | 10.2 |

Table 7.33: New Tagalog speakers' use of object and possessive pronouns.

| Old Tagalog | AGR+ Bound (GA) | | AGR- Bound pro (GPA) | | Free morph. (GPA) | | Dropped (GPA) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | POSS | OBJ | POSS | OBJ | POSS | OBJ | POSS | OBJ | |
| **T7** | 2(50%) | 0(0%) | 0(0%) | 0(0%) | 2(50%) | 1(25%) | 0(0%) | 4(80%) | OBJ 5 POSS 4 |
| **T8** | 3(18.7%) | 0(0%) | 0(0%) | 0(0%) | 8(50%) | 5(71.4%) | 5(31.2%) | 2(28.5%) | OBJ7 POSS16 |
| **T9** | 3(27.2%) | 1(14.2%) | 0(0%) | 1(14.2%) | 6(54.5%) | 4(57.1%) | 2(18.1%) | 1(14.2%) | OBJ7 POSS11 |
| **T20** | 2(14.2%) | 1(11.1%) | 0(0%) | 0(0%) | 8(57.1%) | 7(77.7%) | 4(28.5%) | 1(11.1%) | OBJ9 POSS14 |
| **T21** | 2(22.2%) | 1(20%) | 0(0%) | 1(20%) | 4(44.4%) | 1(20%) | 3(33.3%) | 2(40%) | OBJ5 POSS 9 |
| **T22** | 2(15.3%) | 1(16.6%) | 0(0%) | 0(0%) | 6(46.1%) | 2(33.3%) | 5(38.4%) | 3(50%) | OBJ6 POSS13 |
| Average | 2.6(24.6%) | 0.6(10.35) | 0(0%) | 0.3(5.7%) | 5.6(50.3%) | 3.3(47.4%) | 3.1(24.9%) | 2.1(37.3%) | 17.6 |

Table 7.34: Old Tagalog speakers' use of object and possessive pronouns.

A great variation also has been noticed between members of the same group. For instance, dropped object pronouns in the new group range between 100% and 66%.

### *7.1.4.6. Sundanese Informants*

The occurrences of the four variants of the possessive and object pronouns by the Sundanese informants are displayed in Tables 7.35 and 7.36:

| New Sundanese | AGR+ Bound (GA) | | AGR- Bound pro (GPA) | | Free morph. (GPA) | | Dropped (GPA) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | POSS | OBJ | POSS | OBJ | POSS | OBJ | POSS | OBJ | |
| **Su1** | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 3(60%) | 1(20%) | 2(40%) | 2(66.6%) | OBJ 3 POSS 5 |
| **Su2** | 1(16.6%) | 0(0%) | 0(0%) | 0(0%) | 3(50%) | 1(33.3%) | 2(33.3%) | 2(66.6%) | OBJ3 POSS 6 |
| **Su3** | 1(25%) | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 0(0%) | 3(75%) | 1(100%) | OBJ1 POSS 4 |
| **Su4** | 1(16.6%) | 0(0%) | 0(0%) | 0(0%) | 2(33.3%) | 2(100%) | 3(50%) | 0(0%) | OBJ2 POSS 6 |
| **Su5** | 2(20%) | 0(0%) | 0(0%) | 0(0%) | 4(40%) | 1(25%) | 4(40%) | 3(75%) | OBJ4 POSS10 |
| **Su6** | 2(25%) | 0(0%) | 0(0%) | 0(0%) | 3(37.5%) | 0(0%) | 3(37.5%) | 1(100%) | OBJ1 POSS 8 |
| **Average** | 1.1(17.2%) | 0(0%) | 0(0%) | 0(0%) | 2.5(36.8%) | 0.8(29.7%) | 2.8(45.9%) | 1.5(68%) | 8.7 |

Table 7.35: New Sundanese speakers' use of object and possessive pronouns.

| Old Sundanese | AGR+ Bound (GA) | | AGR- Bound pro (GPA) | | Free morph. (GPA) | | Dropped (GPA) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | POSS | OBJ | POSS | OBJ | POSS | OBJ | POSS | OBJ | |
| **Su7** | 3(33.3%) | 0(0%) | 0(0%) | 1(16.6%) | 4(44.4%) | 3(33.3%) | 2(22.2%) | 2(33.3%) | OBJ 6 POSS 9 |
| **Su8** | 2(28.5%) | 0(0%) | 0(0%) | 0(0%) | 3(42.8%) | 2(66.6%) | 2(28.5%) | 1(33.3%) | OBJ3 POSS 7 |
| **Su9** | 1(20%) | 0(0%) | 0(0%) | 0(0%) | 3(60%) | 6(66.6%) | 1(20%) | 3(33.3%) | OBJ 9 POSS 5 |
| **Su10** | 1(14.2%) | 1(10%) | 0(0%) | 1(10%) | 5(71.4%) | 7(70%) | 1(14.2%) | 1(10%) | OBJ10 POSS 7 |
| **Su11** | 4(36.3%) | 0(0%) | 0(0%) | 0(0%) | 4(36.3%) | 3(75%) | 3(27.2%) | 1(25%) | OBJ4 POSS11 |
| **Su12** | 2(16.6%) | 1(12.5%) | 0(0%) | 0(0%) | 8(66.6%) | 5(62.5%) | 2(16.6%) | 2(25%) | OBJ 8 POSS12 |
| **Average** | 2.1(24.8%) | 0.3(3.7%) | 0(0%) | 0.3(4.4%) | 4.5(53.6%) | 4.3(62.3%) | 1.8(21.5%) | 1.6(26.6%) | 14.9 |

Table 7.36: Old Sundanese speakers' use of object and possessive pronouns.

Both the long-term Sinhalese informants and newly arrived Sundanese informants tend to use the possessive and object pronouns as free morphemes more than dropping them.

Overall, the results in Tables 7.26 to 7.37 reveal that all the informants rarely use pronouns as bound morphemes. Pronouns in GPA are typically used either as free morphemes, or dropped. The fact that no reliable statistical inferences can be made based on the data due to the very low numbers observed in the data, especially those in Table 7.26. One or two instances cannot possibly constitute a pattern. In this case, increasing the size of the sample in the future would probably be highly beneficial in counter acting this effect, and may actually help observe trends and make generalisations. These findings will be discussed in more detail in Sections 8.2.2, 4 and 8.2.3, 4.

### 7.1.5. Variation in Agreement

#### *7.1.5.1. Verbal Agreement*

The GA verb agrees with the noun in gender, number, and person. The verb also inflects for tense, mood, and voice, as discussed in Chapter 4, Section 4.2. However, the GPA verb typically does not agree with the noun. Most GPA speakers use the GA third person singular masculine form to be used with all subjects, (Almoaily, 2013). As for tense in GPA, it is not marked by any verbal inflection. GPA speakers may use forms like *Ana maalom shougol* 'I know.PST job' (as opposed to the GA form: *Ana aʕarif astagoul* 'I know.PRS 1SG job), or *ana akhad rateb gabl youm* 'I had my salary yesterday' (as opposed to the GA form *akhadt* 'had-1.SG.PST). Furthermore, GPA speakers might also drop the verb completely when it will be understood directly throughout the context. Most of the data of GPA corpus regards verbal agreement are attested in Almoaily's classifications as follows:

Three verbal agreement variants are grouped as follows:

1. **AGR Present:** The inflected verb agrees with the subject in gender, number, and person.
2. **AGR Missing:** The verb is inflected but does not agree with the noun in person, number, or gender
3. **Verb Dropped:** The verb is dropped.

Tables 7.37 to 7 42 below list number the tokens for 1, 2, and 3 above:

#### *7.1.5.1.1. Malayalam Informants*

Table 37 shows the number of tokens for verbal agreement, missing agreement, and for verb drop in the Malayalam informants' data:

| New Malay-alam | AGR Present (GA) | AGR Missing (GPA) | Verb Dropped (GPA) | Total | Old Malay-alam | AGR Present (GA) | AGR Missing (GPA) | Verb Dropped (GPA) | Total |
|---|---|---|---|---|---|---|---|---|---|
| **M1** | 1(3.5%) | 15(53.5%) | 12(42.8%) | 28 | **M7** | 4(7.8%) | 42(82.3%) | 5(9.8%) | 51 |
| **M2** | 0(0%) | 9(33.3%) | 18(66.6%) | 27 | **M8** | 0(0%) | 21(72.4%) | 8(27.5%) | 29 |
| **M3** | 2(6.6%) | 16(53.3%) | 12(40%) | 30 | **M9** | 2(4.8%) | 33(80.4%) | 6(14.6%) | 41 |
| **M4** | 2(5.7%) | 20(57.1%) | 13(37.1%) | 35 | **M10** | 4(9.5%) | 34(80.9%) | 4(9.5%) | 42 |
| **M5** | 0(0%) | 7(29.1%) | 17(70%) | 24 | **M11** | 5(15.6%) | 20(62.5%) | 7(21.8%) | 32 |
| **M6** | 2(5.2%) | 17(44.7%) | 19(50%) | 38 | **M12** | 0(0%) | 31(88.5%) | 4(11.4%) | 35 |
| **Average** | 1.1 (3.5%) | 14 (45.2%) | 15.1 (51.2%) | 30.2 | **Average** | 2.5 (6.3%) | 30.1 (77.8%) | 5.6 (15.8%) | 38.2 |

Table 7.37: Verbal agreement in the new and old Malayalam speakers' data.

The data show that there are very few cases of subject-verb agreement in both groups. Overall, the old group tends to drop the verb in fewer cases than the new group.

#### *7.1.5.1.2. Punjabi Informants*

Table 7.38 demonstrates agreement/lack of agreement between the subject and the verb and verb drop by Punjabi informants:

| New Punjabis | AGR Present (GA) | AGR Missing (GPA) | Verb Dropped (GPA) | Total | Old Punjabis | AGR Present (GA) | AGR Missing (GPA) | Verb Dropped (GPA) | Total |
|---|---|---|---|---|---|---|---|---|---|
| **P1** | 0(0%) | 45(76.2%) | 14(23.7%) | 59 | **P7** | 2(6%) | 24(72.7%) | 7(21.2%) | 33 |
| **P2** | 1(3.2%) | 23(74.1%) | 7(22.5%) | 31 | **P8** | 3(5.8%) | 27(79.4%) | 4(11.7%) | 34 |
| **P3** | 0(0%) | 22(84.6%) | 4(15.3%) | 26 | **P9** | 0(0%) | 22(81.4%) | 5(18.5%) | 27 |
| **P4** | 2(5%) | 26(65%) | 12(30%) | 40 | **P10** | 2(4%) | 44(89.7%) | 3(6.1%) | 49 |
| **P5** | 0(0%) | 31(83.7%) | 6(16.2%) | 37 | **P11** | 4(8.7%) | 36(78.2%) | 6(13%) | 46 |
| **P6** | 2\|(7.1%) | 24(85.7%) | 2(7.1%) | 28 | **P12** | 5(12.5%) | 33(82.5%) | 2(5%) | 40 |
| **Average** | 0.8 (2.5%) | 28.5 (78.2%) | 7.5 (19.1%) | 36.8 | **Average** | 2.6 (6.6%) | 31 (80.6%) | 4.5 (12.6%) | 38.1 |

Table 7.38: Verbal agreement in the new and old Punjabi speakers' data.

The new Punjabi informants seem to produce very few tokens of subject-verb agreement compared to the old group as the old group also tend to drop the verb less frequently than the newly arrived speakers.

#### *7.1.5.1.3. Bengali Informants*

Table 7.39 demonstrates the use/absence of verbal agreement by the Bengali informants.

| **New Bengalis** | **AGR Present (GA)** | **AGR Missing (GPA)** | **Verb Dropped (GPA)** | **Total** | **Old Bengalis** | **AGR Present (GA)** | **AGR Missing (GPA)** | **Verb Dropped (GPA)** | **Total** |
|---|---|---|---|---|---|---|---|---|---|
| **B1** | 0(0%) | 9(30%) | 21(70%) | 30 | **B7** | 3(9%) | 18(54.5%) | 12(36.3%) | 33 |
| **B2** | 2(7.6%) | 14(53.8%) | 10(38.4%) | 26 | **B8** | 2(6.6%) | 22(73.3%) | 6(20%) | 30 |
| **B3** | 1(4.5%) | 12(54.5%) | 9(40.9%) | 22 | **B9** | 2(10%) | 13(65%) | 5(25%) | 20 |
| **B4** | 0(0%) | 8(26.6%) | 22(73.3%) | 30 | **B10** | 1(2.7%) | 24(66.6%) | 11(30.5%) | 36 |
| **B5** | 0(0%) | 6(24%) | 19(76%) | 25 | **B11** | 0(0%) | 20(71.4%) | 8(28.5%) | 28 |
| **B6** | 1(3%) | 9(27.2%) | 23(69.6%) | 33 | **B12** | 3(6.8%) | 27(61.3%) | 14(31.8%) | 44 |
| **Average** | 0.6 (2.5%) | 9.6 (36%) | 17.3 (61.4%) | 27.5 | **Average** | 1.8 (5.8%) | 20.6 (65.3%) | 9.3 (28.7%) | 31.7 |

Table 7.39: Verbal agreement in the new and old Bengali speakers' data.

Both newly arrived Bengali informants and long-term ones produce very few tokens of subject-verb agreement. The new Bengali informants seem to drop the verb more than the old informants. The old Bengali group, on the other hand, seem to move towards non-agreeing verbal forms.

#### *7.1.5.1.4. Sinhala Informant*

Table 7.40 shows the use/absence of verbal agreement by the Sinhalese informants.

| **New Sinhalese** | **AGR Present (GA)** | **AGR Missing (GPA)** | **Verb Dropped (GPA)** | **Total** | Old Sinhalese | **AGR Present (GA)** | **AGR Missing (GPA)** | **Verb Dropped (GPA)** | **Total** |
|---|---|---|---|---|---|---|---|---|---|
| **S1** | 0(0%) | 12(63.1%) | 7(36.8%) | 19 | S7 | 5(10.8%) | 39(84.7%) | 2(4.3%) | 46 |
| **S1** | 0(0%) | 24(88.8%) | 3(11.1%) | 27 | S8 | 2(3.3%) | 54(90%) | 4(6.6%) | 60 |
| **S3** | 2(8%) | 17(68%) | 6(24%) | 25 | S9 | 4(7.5%) | 44(83%) | 5(9.4%) | 53 |
| **S4** | 0(0%) | 22(81.4%) | 5(18.5%) | 27 | S10 | 6(8.1%) | 62(83.7%) | 6(8.1%) | 74 |
| **S5** | 1(2.7%) | 26(72.2%) | 9(25%) | 36 | S11 | 8(8%) | 89(89.8%) | 2(2%) | 99 |
| **S6** | 1(4.1%) | 15(62.5%) | 8(33.3%) | 24 | S12 | 4(7.2%) | 48(87.2%) | 3(5.4%) | 55 |
| **Average** | 1.8 (5.8%) | 20.6 (65%) | 9.3 (28 %) | 31.7 | Average | 4.8 (7.5%) | 56 (86.4%) | 3.6 (6%) | 64.4 |

Table 7.40: Verbal agreement in the new and old Sinhalese speakers' data.

The data show that there are very few cases of subject-verb agreement in newly arrived Sinhalese speakers. Both groups tend to drop the verb, while the old group is more likely to not produce fully inflected verb forms.

#### *7.1.5.1.5. Tagalog Informant*

Table 7.41 demonstrates agreement/lack of agreement between the subject and the verb and verb drop by Tagalog informants:

| New Tagalog | AGR Present (GA) | AGR Missing (GPA) | Verb Dropped (GPA) | Total | Old Tagalog | AGR Present (GA) | AGR Missing (GPA) | Verb Dropped (GPA) | Total |
|---|---|---|---|---|---|---|---|---|---|
| **T1** | 1(3.1%) | 18(56.2%) | 13(40.6%) | 32 | **T7** | 1(3.5%) | 21(75%) | 6(21.4%) | 28 |
| **T2** | 0(0%) | 7(26.9%) | 19(73%) | 26 | **T8** | 0(0%) | 28(82.3%) | 6(17.6%) | 34 |
| **T3** | 1(2.9%) | 13(38.2%) | 20(58.8%) | 34 | **T9** | 2(8.3%) | 17(70.8%) | 5(20.8%) | 24 |
| **T4** | 0(0%) | 9(31%) | 20(68.9%) | 29 | **T10** | 2(9.5%) | 16(76.1%) | 3(14.2%) | 21 |
| **T5** | 1(3.8%) | 8(30.7%) | 17(65.3%) | 26 | **T11** | 0(0%) | 26(86.6%) | 4(13.3%) | 30 |
| **T6** | 0(0%) | 15(41.6%) | 21(58.3%) | 36 | **T12** | 1(4%) | 16(64%) | 8(32%) | 25 |
| **Average** | 0.5 (1.6 %) | 11.6 (37 %) | 18.3 (60 %) | 30.5 | **Average** | 1 (4.1%) | 20.6 (75 %) | 5.3 (19 %) | 27 |

Table 7.41: Verbal agreement in the new and old Tagalog speakers' data.

The new Tagalog informants seem to drop the verb more frequently than the old group counterparts who seem to use more inflected, but fewer agreeing forms.

#### *7.1.5.1.6. Sundanese Informants*

Table 7.42 shows the number of tokens for verbal agreement, missing agreement, and for verb drop in the Sundanese informants' data:

| New Sundanese | AGR Present (GA) | AGR Missing (GPA) | Verb Dropped (GPA) | Total | Old Sundanese | AGR Present (GA) | AGR Missing (GPA) | Verb Dropped (GPA) | Total |
|---|---|---|---|---|---|---|---|---|---|
| **Su1** | 0(0%) | 6(85.7%) | 1(14.2%) | 7 | **Su7** | 2(7.6%) | 22(84.6%) | 2(7.6%) | 26 |
| **Su2** | 0(0%) | 11(55%) | 9(45%) | 20 | **Su8** | 6(15.3%) | 33(84.6%) | 0(0%) | 39 |
| **Su3** | 3(11.1%) | 22(81.4%) | 2(7.4%) | 27 | **Su9** | 5(9.4%) | 44(83%) | 4(7.5%) | 53 |
| **Su4** | 0(0%) | 4(30.7%) | 9(69.2%) | 13 | **Su10** | 3(9%) | 23(69.6%) | 7(21.2%) | 33 |
| **Su5** | 2(10%) | 7(35%) | 11(55%) | 20 | **Su11** | 7(15.9%) | 34(77.2%) | 3(6.8%) | 44 |
| **Su6** | 2(7.6%) | 18(69.2%) | 6(23%) | 26 | **Su12** | 6(14.2%) | 33(78.5%) | 3(7.1%) | 42 |
| **Average** | 1.1 (4.8%) | 11.3 (59%) | 6.3 (35%) | 18.8 | **Average** | 4.8 (11.9%) | 31.5 (79%) | 3.1 (8.4%) | 39.4 |

Table 7.42: Verbal agreement in the new and old Sundanese speakers' data.

The new Sundanese informants seem rarely to produce fully inflected verb forms which suggest that length of stay seems to have a minor effect on the Sundanese sample. However, they use more inflected but fewer agreeing forms. Also, the new Sundanese informants seem to drop the verb more frequently than the long-term speakers. The informants' length of stay in SA, however, seems to have a significant effect on the Sundanese informants.

In general, the data in tables 7.37 to 7.42 reveal that all the informants rarely produce fully inflected verb forms that agree with the subject (i.e. the form used in GA). The data also suggests that all informants show a length of-stay related development in the use of verbs: all new informants seem to drop verbs more frequently than their 'old group' counterparts, who seem to use more inflected, but fewer agreeing forms.

#### *7.1.5.2. Agreement in the NP and in the ADJP*

The data in GPA corpus revealed that informants usually produced adjectives that typically do not agree with their noun in gender and number. Furthermore, Almoaily (2013) adds that the singular masculine form is used with all nouns. In addition, my corpus revealed that, in some cases, informants produced adjectives which agree with the noun in number and gender. As regards the use of demonstratives, GPA speakers tend to use the singular masculine as an unmarked form to all nouns and adjectives regardless of their gender. However, some informants in few cases use demonstrative which agrees with the noun in number and gender. To quantify the GPA system of agreement in the NP and in the ADJP, I followed Almoaily's classification which falls into two categories:

**1. Agreement present:**

Almoaily (2013: 148) suggests that instances of agreement of the unmarked form (singular masculine) should not been counted because these instances do not reflect whether the informants actually apply GA NP/ ADJP agreement or whether they use an invariant form. Thus, it takes any of the following:

a) Either: The adjective agrees with the noun in gender and number.

b) Or: The demonstrative agrees with the noun in number and gender.

c) Or: The noun or adjective agrees with the numeral (for numbers between 3 and 10).

**2. Agreement missing:** There is no agreement in gender and/or number in the adjective phrase (ADJP) or in the noun phrase (NP).

Tables 7.43 to 7.48 tabulate the number of tokens for agreement/lack of agreement in the NP and in the ADJP for every informant.

#### *7.1.5.2.1. Malayalam Informants*

The instances of the presence/absence of nominal agreement in the Malayalam informants' data are shown in Table 7.43.

| New Malayalam | Agreement present (GA) | Agreement missing (GPA) | Total | Old Malayalam | Agreement present (GA)) | Agreement missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **M1** | 4 (15.3%) | 22 (84.6% ) | 26 | **M7** | 3 (13%) | 20(86.9%) | 23 |
| **M2** | 5 (16.1%) | 26 (83.8%) | 31 | **M8** | 2 (8%) | 23 (92%) | 25 |
| **M3** | 0 (0%) | 18 (100%) | 18 | **M9** | 5 (13.1%) | 33(86.8%) | 38 |
| **M4** | 3 (15.7%) | 16 (84.2%) | 19 | **M10** | 7 (23.3%) | 23(76.6%) | 30 |
| **M5** | 0 (0%) | 21 (100%) | 21 | **M11** | 4 (28.5%) | 10(71.4%) | 14 |
| **M6** | 2 (10%) | 18 (90%) | 20 | **M12** | 6 (21.4%) | 22(78.5%) | 28 |
| Average | 2.3 (9.5%) | 20.1 (90.4%) | 22.4 | **Average** | 4.5(17.9%) | 21.8(82%) | 26.3 |

Table 7.43: Agreement in the NP and in the ADJP, new and old Malayalam informants.

In the Malayalam language group, old members show some development in the accommodation to the GA nominal agreement system.

#### *7.1.5.2.2. Punjabi Informants*

Table 7.44 presents the tokens of nominal agreement/lack of agreement in the Punjabi speakers' data.

| New Punjabis | Agreement present (GA) | Agreement missing (GPA) | Total | Old Punjabis | Agreement present (GA)) | Agreement missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **P1** | 5(25%) | 15(75%) | 20 | **P7** | 7 (25.9%) | 20 (74%) | 27 |
| **P2** | 3(20%) | 12 (80%) | 15 | **P8** | 4(20%) | 16 (80%) | 20 |
| **P3** | 2(16.6%) | 10 (83.3%) | 12 | **P9** | 4(25%) | 12 (75%) | 16 |
| **P4** | 0 (0) | 18(100%) | 18 | **P10** | 2(20%) | 8 (80%) | 10 |
| **P5** | 4(28.5%) | 10 (71.4%) | 14 | **P11** | 5(45.4%) | 6(54.5%) | 11 |
| **P6** | 0 (0%) | 9 (100%) | 9 | **P12** | 6(37.5%) | 10 (62.5%) | 16 |
| **Average** | 2.3(13.4%) | 12.3(86.5%) | 14.6 | **Average** | 4.6 (28.9%) | 12 (71%) | 16.6 |

Table 7.44: Agreement in the NP and in the ADJP, new and old Punjabi informants.

The data revealed that there is variation among the members of the new Punjabi group. For instance, P5 produces the nominal agreement form 28.5% of the time while P6 and P4 and P2 produce it in 0% and in 17% of the cases, respectively. Note that the old Punjabi members show some development in the acquisition of the GA nominal agreement system.

### *7.1.5.2.3. Bengali Informants*

The tokens for agreement in the NP and in the ADJP in the data of Bengali speakers are displayed in Table 7.45.

| New Bengalis | Agreement present (GA) | Agreement missing (GPA) | Total | Old Bengalis | Agreement present (GA)) | Agreement missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **B1** | 1(7.6%) | 12 (92.3%) | 13 | **B7** | 2(28.5%) | 5(71.4%) | 7 |
| **B2** | 0(0%) | 13(100%) | 13 | **B8** | 1(10%) | 9(90%) | 10 |
| **B3** | 0(0%) | 9(100%) | 9 | **B9** | 2(11.7%) | 15(88.2%) | 17 |
| **B4** | 2(16.6%) | 10(83.3%) | 12 | **B10** | 2(10%) | 18(90%) | 20 |
| **B5** | 1(11.1%) | 8(88.8%) | 9 | **B11** | 4(21%) | 15(78.9%) | 19 |
| **B6** | 1(8.3%) | 11(91.6%) | 12 | **B12** | 3(12%) | 22(88%) | 25 |
| **Average** | 0.8(7.3%) | 11.3(92.6%) | 12.1 | **Average** | 2.3(15.5%) | 14 (84.4%) | 16.3 |

Table 7.45: Agreement in the NP and in the ADJP, new and old Punjabi informants.

In the Bengali language group, the old members show some development in the accommodation of the GA nominal agreement system. However, the predominant form is missing agreement.

### *7.1.5.2.4. Sinhala Informant*

The instances of the presence/absence of nominal agreement in the Sinhalese informants' data are shown in Table 7.46.

| New Sinhalese | Agreement present (GA) | Agreement missing (GPA) | Total | Old Sinhalese | Agreement present (GA)) | Agreement missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **S1** | 0 (0%) | 17 (100%) | 17 | **S7** | 3 (25%) | 9 (75%) | 12 |
| **S1** | 3 (21.4%) | 11 (78.5%) | 14 | **S8** | 4 (28.5%) | 10 (71.4%) | 14 |
| **S3** | 2 (12.5%) | 14 (87.5%) | 16 | **S9** | 6 (46.1%) | 7 (53.8%) | 13 |
| **S4** | 0 (0%) | 20 (100%) | 20 | **S10** | 4 (23.5%) | 13 (76.4%) | 17 |
| **S5** | 1 (4%) | 24 (96%) | 25 | **S11** | 3 (42.8%) | 4(57.1%) | 7 |
| **S6** | 2 (9.5%) | 19 (90.4%) | 21 | **S12** | 5(35.7%) | 9(64.2%) | 14 |
| **Average** | 1.3 (7.9%) | 17.5 (92%) | 18.8 | Average | 4.1(33.6%) | 8.6 (66.3%) | 12.7 |

Table 7.46: Agreement in the NP and in the ADJP, new and old Sinhalese informants.

Sinhalese new informants produce very few tokens of agreement between the noun and the number, adjective, or the demonstrative, while the old members are more likely to use agreement in the NP and in the ADJP.

### *7.1.5.2.5. Tagalog Informant*

Table 7.47 presents the tokens of nominal agreement/lack of agreement in the 5) Tagalog speakers' data.

| New Tagalog | Agreement present (GA) | Agreement missing (GPA) | Total | Old Tagalog | Agreement present (GA)) | Agreement missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **T1** | 0 (0%) | 31 (100%) | 31 | **T7** | 5 (41.6%) | 7 (58.3%) | 12 |
| **T2** | 1 (10%) | 9 (90%) | 10 | **T8** | 4 (20%) | 16 (80%) | 20 |
| **T3** | 3 (15%) | 17 (85%) | 20 | **T9** | 5 (45.4%) | 6 (54.5%) | 11 |
| **T4** | 1 (7.6%) | 12 (92.3%) | 13 | **T10** | 3 (14.2%) | 18 (85.7%) | 21 |
| **T5** | 0 (0%) | 21 (100%) | 21 | **T11** | 6 (30%) | 14 (70%) | 20 |
| **T6** | 3 (21.4%) | 11(78.5%) | 14 | **T12** | 4 (30.7%) | 9 (69.2%) | 13 |
| **Average** | 1.3(9%) | 16.8 (90.9%) | 18.1 | **Average** | 4.5 (30%) | 11.6 (69%) | 16.1 |

Table 7.47: Agreement in the NP and in the ADJP, new and old Tagalog informants.

The data in Table 7.47 indicates that there is an increase in the use of agreement in the NP and in the ADJP among the old Tagalog group compared to the newly arrived Tagalog group.

#### *7.1.5.2.6. Sundanese Informants*

Table 7.48 presents the tokens of nominal agreement and lack of agreement in the Sundanese speakers' data. The data of the Sundanese language group suggests that there is an increase in the use of agreement in the NP and in the ADJP among the old Sundanese group compared to the newly arrived Sundanese.

| New Sundanese | Agreement present (GA) | Agreement missing (GPA) | Total | Old Sundanese | Agreement present (GA) | Agreement missing (GPA) | Total |
|---|---|---|---|---|---|---|---|
| **Su1** | 0 (0%) | 19 (100%) | 19 | **Su7** | 3 (21.4%) | 11(78.5%) | 14 |
| **Su2** | 3(15.7%) | 16 (84.2%) | 19 | **Su8** | 5(62.5%) | 3(37.5%) | 8 |
| **Su3** | 2 (20%) | 8 (80%) | 10 | **Su9** | 4(30.7%) | 9(69.2%) | 13 |
| **Su4** | 1 (8.3%) | 11 (91.6%) | 12 | **Su10** | 3(42.8%) | 4(57.1%) | 7 |
| **Su5** | 1(20%) | 4(80%) | 5 | **Su11** | 6(42.8%) | 8(57.1%) | 14 |
| **Su6** | 2(13.3%) | 13 (86.6%) | 15 | **Su12** | 3(23%) | 10(76.9%) | 13 |
| **Average** | 1.5 (12.9%) | 11.8 (87%) | 13.3 | **Average** | 4(37.2%) | 7.5(62.7%) | 11.5 |

Table 7.48: Agreement in the NP and in the ADJP, new and old Sundanese informants.

The data for the Sundanese language group shows that that there is an increase in the use of agreement in the NP and in the ADJP among the old Sundanese group compared to the newly arrived Sundanese group.

To sum up the results for agreement in the NP and ADJP, the data in Tables 7.43 to 7.48 shows that the number of years the informants have stayed in the Gulf seems to have only a slight positive effect on the occurrence of agreement in the NP and in the ADJP for all language groups. The next sections present the corpus analysis results obtained on the basis of the Chi-squared test analysis.

## 7.2. The Extent to which Language Variation between Gulf Pidgin Arabic Speakers is Significant

The following tables were created to determine whether *L1* and *length of stay in the Gulf* are potential factors for *language* and *gender* variation in GPA.

### 7.2.1. Substrate-Language-Based Comparison (Substrate-Language-Based Variation)

In this section, I compare speakers coming from different language groups (i.e. Malayalam, Punjabi, Bengali, Sinhala, Tagalog, and Sunda) and their choice among 10 morphosyntactic features: presence or absence of the Arabic definiteness markers, presence or absence of Arabic conjunction markers, presence or absence of the GPA copula, free or bound object or possessive pronoun, and presence or absence of agreement in the VP and the NP. According to my calculations and results, the conclusion of the features investigated in this study, based on potential substrate influence on language variation in GPA, is displayed in Table 7.49 (the data of *p values* lower than 0.05 is shaded). The table shows that Malayalam speakers, at 43%, use the definiteness marker more than the other language groups, despite the absence of definiteness markers in their language. As regards the conjunction markers, the Bengali language group produced fewer tokens of the conjunction markers (only 12%) than the other groups.

| Substrate language / Linguistic features | Malayalam | Punjabi | Bengali | Sinhala | Tagalog | Sunda | *p*-value |
|---|---|---|---|---|---|---|---|
| **Definiteness** | **43%** | **15%** | **33%** | **22%** | **13.1%** | **10%** | ***p* = 0.06** |
| **Conjunction** | **17%** | **26%** | **12%** | **23.1%** | **35%** | **29%** | ***p* = 0.001** |
| **Copula** | **70%** | **68%** | **72%** | **74%** | **74%** | **72%** | ***p* = 0.08** |
| **Pronoun** | **20.4%** | **4.2%** | **19.1%** | **20%** | **14%** | **11.4%** | ***p* = 0.07** |
| **Verbal AGR** | **5%** | **5%** | **4.1%** | **7%** | **3%** | **8%** | ***p* = 0.08** |
| **NP and ADJP AGR** | **14%** | **21.1%** | **11.4%** | **21%** | **20%** | **25%** | ***p* = 0.0008** |

Table 7.49: Summary of substrate language influence.

The chi-square test revealed that the difference between speakers of the six groups is significant at a *p*-value of 0.001. This is likely due conjunction markers in Sunda, Tagalog, and Punjabi being free morphemes, like GA conjunction, whereas they are suffixes attached to the noun in Malayalam and Sinhala, and optional in Bengali. For the use of the copula, it seems that there are differences across the substrate languages. These differences are expected to have an effect on the informants' use of the GPA copula, *fi*, since there is no copula in the superstrate

language, GA, in the present tense; the consideration will only focus on the use of the copula *fi* in the present tense in GPA. On average, the use of bound morpheme cases were far fewer than the tokens of dropped or free pronouns. However, it seems that no reliable generalisations can be made based on the data due to the low numbers observed in the data. Furthermore, the results in Table 7.49 suggest that all of the informants rarely produced the subject-verb agreement form as it is used in GA (i.e. fully inflected verb forms that are marked for TMA and agree with the subject in number, gender, and person). As for the noun/adjective agreement, the data indicates that Malayalam and Punjabi speakers show more cases of noun-adjective agreement than the other four languages. The chi-square test further revealed that the difference between speakers of the six groups is significant at a *p*-value of 0.0008. The results are depicted in Figure 7.1.

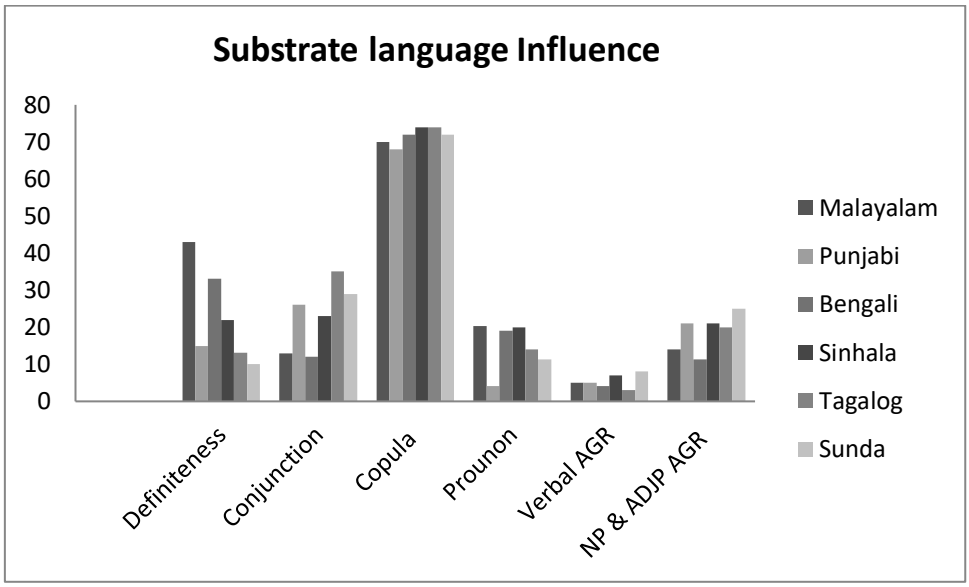



Figure 7.1: Percentage of substrate language influence on female GPA speakers.

### 7.2.2. Production of Linguistic Features by the Newcomers Versus Long-Term Residents

In this section, I compare newly settled informants with long-term residents for every feature. Table 7.50 displays the features investigated in this study and indicates whether a shift occurred

towards the superstrate language, GA. Furthermore, Figure 7.2 illustrates female GPA speakers' shift towards GA based on length of stay in the Gulf.

| Linguistic feature | Female GPA speakers | | *p*-value |
|---|---|---|---|
| | New | Old | |
| **Definiteness** | 11% | 34% | $p = 0.002$ |
| **Conjunction** | 13% | 42% | $p = 0.001$ |
| **Copula** | 68.4% | 74% | $p = 0.35$ |
| **Pronoun** | 18% | 23% | $p = 0.0001$ |
| **Verbal AGR** | 5.2% | 8.4% | $p = 0.22$ |
| **NP and ADJP AGR** | 19.3% | 27% | $p = 0.08$ |

Table 7.50: Summary of informants' shift towards GA.

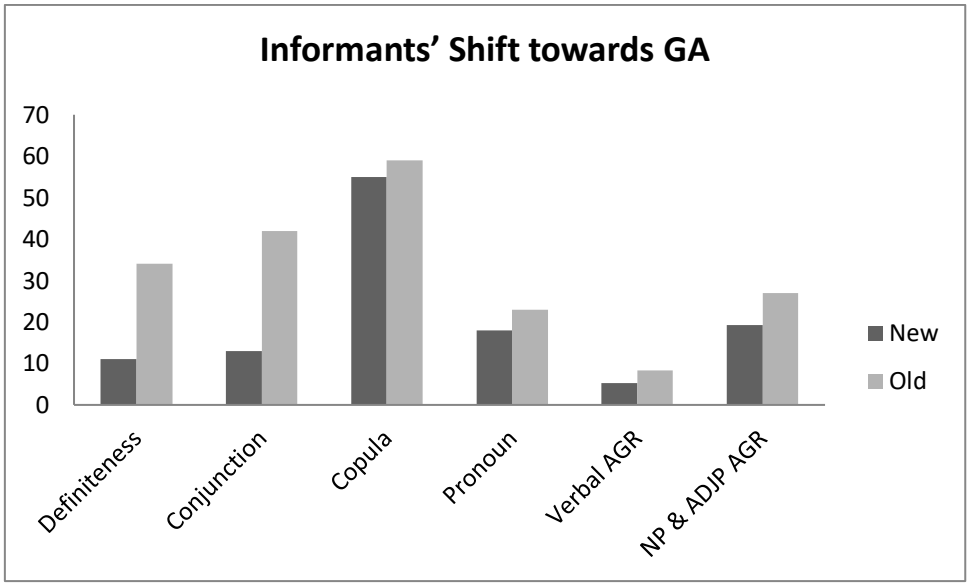



Figure 7.2: Female GPA speakers' shift towards GA based on Length of Stay in the Gulf.

### 7.2.3. Number of Years of Residency Based on Gender Variation

In this section, I draw a comparison between the five morphosyntactic features used by *male* and *female* GPA speakers. I investigated the influence of female GPA speakers'[1] length of stay on linguistic production, and I compared my results with Almoaily's (2013), which revealed that Asian male workers, after spending some time in the Gulf, are hardly affected in terms of their choices of morphosyntactic features. This allowed me to investigate the question of whether there is a difference in rates of accommodation in the Gulf by gender. Table 7.51 displays the differences in linguistic features investigated in this study – shown as a comparison between GPA men and women speakers based on length of stay in the Gulf – and Figure 7.3 depicts these differences in graph form.

| **Linguistic feature** | **GPA speakers** | | | | ***p*-value** |
|---|---|---|---|---|---|
| | **Men** | | **Women** | | |
| | **New** | **Old** | **New** | **Old** | |
| **Definiteness** | 19.1 | 30 | 11 | 34 | $p = 0.06$ |
| **Conjunction Markers** | 7.6 | 23.3 | 13 | 42 | $p = 0.005$ |
| **Copula** | 68.2 | 72 | 55 | 59 | $p = 0.22$ |
| **pronoun** | 11.5 | 23 | 18 | 23 | $p = 0.07$ |
| **Verb Agreement** | 3.1 | 4 | 5.2 | 8.4 | $p = 0.06$ |
| **Agreement in NP and ADJP** | 6.2 | 15 | 19.3 | 27 | $p = 0.008$ |

Table 7.51: Summary of differences in linguistics features between men (adopted from Almoaily's male corpus) and women, based on length of stay in the Gulf.

[1] That is, female GPA speakers who spent five years or less in the Gulf at the time I interviewed them, compared with long-term female residents in the Gulf (i.e. those who spent 10 years or more in the area at the time they were interviewed) who come from three different linguistic backgrounds (Malayalam, Bengali, and Punjabi).

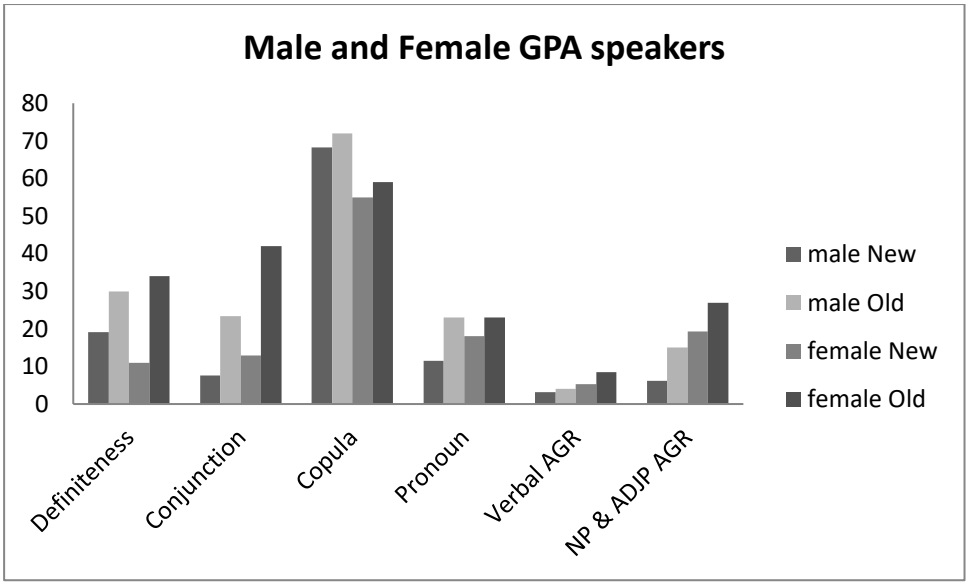



Figure 7.3: Differences in linguistics features between men and women based on length of stay in the Gulf.

The rather general discussion of the five morphological features under investigation across members of the six L1 groups is concluded in Section 7.3 below. Thereafter, Chapter 8 provides a more detailed discussion of these findings.

### 7.3. Conclusion to Chapter 7

This chapter presented the results of my fieldwork. As detailed in Sections 7.2.1–7.2.5, every language group was split into two groups based on their length of stay in the Gulf (five years or less as 'New' speakers, and 10 years or more as 'Old'). Each section tabulated the observed linguistic variants for each linguistic feature under consideration. In addition, Section 7.2 laid out the findings from the corpus analysis as related to the research hypotheses presented in Section 6.2. The results of the statistical analysis in Section 7.3 revealed that both the amount of time the informants have stayed in the Gulf and superstrate influence seem to have a slight positive effect on female GPA speakers' choices of morphosyntactic features. At the same time, female GPA speakers display a greater tendency to shift towards GA than men after spending more than 10 years in the Gulf. Next, Chapter 8 provides a more concrete analysis of language variation in GPA resulting from different morphosyntactic structures of male and female GPA

speakers, the substrate languages of female GPA speakers, and their duration of stay in the Gulf.

# Chapter 8: Discussion of Findings

*Make things as simple as possible, but not simpler. (Albert Einstein)*

## 8.1. Introduction to Chapter 8

As the GPA female speakers' population continues to grow in size (963,723 foreign domestic workers in the Kingdom by the end of 2017[1] accounting for 76% of the total domestic workforce), understanding the way in which they acquired and speak GA is important to promote healthy working environments. In particular, it is important to consider the account of variation in GPA conditioned by the substrate language, length of stay, and gender variation in order to shed light on potential factors conditioning language variation in GPA. This would be the first step to opening up the possibility for further research on this group of women who use GPA and understanding the importance of social and developmental factors in language production.

This research aims to provide a qualitative and a quantitative analysis of language variation in GPA resulting from different morphosyntactic structures of male and female GPA speakers, the substrate languages of female GPA speakers, and their duration of stay in the Gulf. Moreover, it aims to determine the extent to which a less-studied non-Indo-European-based Arabic-lexifier pidgin, namely, GPA, complies with the proposed universal typological features of P/Cs. At the end of this chapter, a discussion of the competing theories of the genesis

[1] According to statistics issued by the Ministry of Labor and Social Development. Retrieved on 13 December 2017 from https://saudigazette.com.sa/article/534105/SAUDI-ARABIA/14m-house-drivers-964000-domestic-workers-in-2017-mdash-ministry-statistics?fbclid=IwAR2xQHlOyLl96U2nifHRS2gRZwHSd9GChGqCRJKY8kglbV1utVb_kaYkydQ

of contact languages (i.e. the substratist and the universalist, including the imperfect SLA) is presented to explain the various aspects of the emergence of GPA (Section 8.4).

## 8.2. Corpus Analysis Findings

This section discusses the findings from the corpus analysis as related to the research hypotheses presented in Chapter 6 and the obtained results listed in Chapter 7. As a reminder, the investigated research hypotheses, along with the methods for testing them, are as follows: Different members of language groups were compared (e.g. the relative number of tokens produced by the Bengali sample vs. the relative number of instances produced by the Malayalam respondents). The analysis was based on the informants' use of the variants of five selected morphosyntactic phenomena: presence or absence of the Arabic definiteness marker, presence or absence of Arabic conjunction markers, presence or absence of the GPA copula, free or bound object or possessive pronoun, and presence or absence of agreement in the VP and the NP.

### 8.2.1. Research Hypothesis 1 Findings

The first hypothesis, namely, that there are differences between standard GA and GPA, was qualitatively tested by conducting a cross-linguistic comparison of the morphosyntax of GPA and GA in an attempt to investigate the differences in the six morphosyntactic variables: (1) presence or absence of the GA definiteness marker; (2) free, bound, or dropped object and possessive pronouns; (3) use of coordinating conjunction or juxtaposition; (4) use or dropping of the copula in the present tense; (5) presence or absence of nominal agreement; and finally (6) verb dropping, or presence or absence of verbal agreement in GPA (the pidgin under investigation in this project) and its lexifier, GA. Analysing the differences in the interview data displayed in Table 5.2, Chapter 4 supports Hypothesis 1; that is, there are significant differences in the morphosyntactic systems of GPA and GA. For example, the use of the copula

*fi* is optional in GPA, while in GA it is used overtly only in past and future sentences, but is covert in the present tense.

### 8.2.2. Research Hypothesis 2 Findings

The second research hypothesis was that a difference exists between the GPA spoken by speakers with different L1s, and the results are presented in Tables 7.2–7.51 in Chapter 7. These results of the observed linguistic variants for each linguistic feature under consideration are discussed in the following subsections numbered from 1 to 6. (Note that Section 4.1 was devoted to a descriptive and interpretative analysis account of all the five linguistic features of GPA.)

1. **Discussion of Findings – Definiteness**

As discussed in Chapter 5, markers for definiteness can be found in the substrate languages Bengali, Tagalog, and Sundanese, whereas these markers are missing in the rest of the languages. After examining the results in Tables 7.1 to 7.6 in Chapter 7, it was found that Malayalam speakers use the definiteness marker more than the other language groups, despite the absence of definiteness markers in their L1. Comparing the rate of occurrence of the GA definiteness marker *al-* in the data of Malayalam speakers (42.5%) with that produced by the Bengali (22.4%), Sinhala (21.6%), Punjabi (14.7%), Tagalog (13.1%), and Sundanese (10%) speakers, the chi-square test revealed that the difference between GPA speakers with different L1s is not statistically significant *(p*-value = 0.06). This shows that the influence of the informants' L1s on their choice of GPA definiteness variants is weak, and there was thus no substratal effect on the production of the GA definiteness by GPA speakers.

2. **Discussion of Findings – Conjunction Markers**

The discussion in Chapter 5 demonstrates that the use of conjunction markers is optional in Bengali and Sinhala, whereas it is compulsory in Malayalam, Punjabi, Tagalog, and Sunda.

Hypothesis 2 was tested on the use of conjunction markers among the six linguistic backgrounds.

The data in Chapter 7 has indeed revealed that – from a substratal point of view– the Bengali language group and the Sinhala sample produced fewer tokens of the conjunction markers (only 10.2% and 12.9%, respectively) compared to the Sunda speakers (24%), the Tagalog speakers (23.3%), the Punjabi speakers (22.7%), and the Malayalam speakers (19.3%). Thus, the hypothesis that speakers with different L1s speak GPA differently is confirmed. The chi-square test revealed that the difference between the speakers of the six groups is significant at a *p*-value of 0.001. It is also clear from these figures that Sunda, Tagalog, and Punjabi speakers produce a remarkably higher number of conjunction markers than speakers of the remaining three languages. This is probably due to the fact that conjunction markers *atwa*, *o*, and *jar* in Sunda, Tagalog, and Punjabi are free morphemes – they are like the GA coordinating conjunction *wa* ('or') – whereas they are suffixes attached to the noun in Malayalam and Sinhala, and optional in Bengali. This result is in parallel with Almoaily's (2012) study of male GPA speakers.

### 3. Discussion of Findings – The Copula *fi*

As discussed in Chapter 5, Sinhala is the only substrate language of the six in which the copula is usually dropped, and it is not used in spoken Sinhala (Gair, 1968). However, there are variations in the use of the copula across the substrate languages. In Malayalam, for instance, the copula is used with all predicates (see Asher and Kumari, 1997) while in Punjabi, it is used only with positive sentences (see Bhatia, 2013). On the other hand, the copula in Bengali is only used with temporary predicates (see Finch, 2001), and in Tagalog, the use of the copula is optional in SVO sentences (see Schachter and Otanes, 1983). With these differences, it would be possible to expect an effect on the use of the copula between GPA speakers. However, in GA, since there is no copula in the present tense, Almoaily (2013) suggests discussing the use

of the copula *fi* only in the present tense in GPA. If long-term residents are more likely to drop the copula in the present tense more often than newcomers, this might indicate more accommodation to GA. If not, then old GPA speakers might produce a form of GPA which is different from GA.

Comparing the percentages of each of the six variants revealed slight differences between language groups. According to the data in Section 7.1, the Tagalog sample dropped the copula *fi* in 73.7% of the total number of cases where a copula could have been used in the present tense. Likewise, Sinhala speakers dropped the GPA copula in 73.6% of cases, although the Sinhala sample is expected to drop the copula more than other languages group. Moreover, Sundanese informants dropped the copula *fi* in 71.8% of cases, and similarly, Bengali informants dropped the GPA copula in 71.5% of cases. In contrast, Punjabi and Malayalam speakers were found to drop the copula in the present tense in 69.5% and 69.1% of cases, respectively, where a copula could have been used. Hence, the hypothesis that there is a difference in terms of use the copula *fi* between the GPA spoken by speakers with different L1s can be rejected, as the data revealed that Sinhala speakers produce tokens of the copula which are almost equal to Tagalog speakers, despite the fact that they do not use the copula in their spoken language. Overall, the difference between the six groups is negligible and fails to reach significance ($p$-value = 0.08). This leads to support for the null hypothesis that there is no difference between the GPA spoken by speakers with different L1s in terms of the frequency of the use of the copula.

### 4. Discussion of Findings – Pronouns

Section 4.1.2 provided a descriptive account of GPA pronouns. In terms of the use of object and possessive pronouns, GPA has four variant forms in the construction of the object and

possessive pronouns: (1) using the GA bound pronoun[2] which agrees with the noun in person, number, and gender; (2) using the GA bound pronoun which does not agree with the noun; (3) using the free pronoun[3]; and (4) dropping the object or possessive pronoun. Based on the description of the six substrate languages in Chapter 5, the uses of object and possessive pronouns are free in all six substrate languages studied in the current project. It was initially expected that the informants' L1s would yield only minor differences in their selection among the four GPA variants for object and possessive pronouns. Almoaily (2013) referred to this as the transfer effect, where speakers of the six languages are expected to produce high rates of free pronouns compared to their production of bound pronouns.

Data in Section 7.1 revealed that, on average, participants seem to produce higher rates of tokens of dropped or free pronouns as compared to instances of bound morphemes. For example, in the data of Punjabi informants, free pronouns were chosen in 43.8% of cases, and the pronouns were dropped in 55.5% of cases, while the total frequency of the two variants *agreeing* and *non-agreeing bound pronouns* was only 7.2%. Similarly, the Sundanese informants dropped pronouns and used free pronouns in 53.3% and 47.3% of cases, respectively, while bound pronouns were used only 17.1% of the time when they could have used them. Likewise, Tagalog informants chose free morphemes in 52.1% of all tokens and 20.2% dropped pronouns, while they used bound possessive and object pronouns (both agreeing and non-agreeing) in a total of 20.4% of cases. Moreover, in the data of the Bengali speakers, free pronouns were chosen in 46.1% of cases, and the pronouns were dropped in 29.3% of cases, whereas bound pronouns were used in 24.2% of cases. Interestingly, none of the substrate languages under investigation have bound pronouns, although Malayalam informants, with some exceptions (informants M2, M3, M5, and M7 did not use GA bound

[2] GA object and possessive pronouns are always bound morphemes which are inflected for person, number, and gender (Qafisheh, 1977).
[3] There is no distinction in the use of subject pronouns as object or possessive pronouns (Almoaily, 2012).

pronouns), displayed more improvement in using the GA bound pronouns compared to the rest of the informants. They used bound pronouns in 38.2% of the total times where they could have used them. Hence, the data invalidates the hypothesis that there is a difference in terms of the use of pronouns between the GPA spoken by speakers with different L1s, especially since none of the substrate languages under investigation have bound pronouns. Although the difference fails to reach significance (*p*-value = 0.07), and with low frequencies, it seems that there is a development in the acquisition of bound pronouns, specifically with Malayalam informants.

## 5. Discussion of Findings – Subject-Verb Agreement

Chapter 5 discussed subject-verb agreement in all six languages, and demonstrated that Malayalam and Sinhalese are the only two substrate languages of the six that lack this agreement (Asher and Kumari, 1997; Müller-Gotama, 200). In Punjabi, Bhatia (2013) mentioned that the verb agrees with the subject in person, number, and gender. In Bengali, Ray (1966) stated that verb agrees with the subject in person. In Tagalog, the verb inflects for tense and aspect (Schachter and Otanes, 1983). In Sundanese, the verb agrees with subject in number and person (Hardjadibrata, 1985). After examining the GPA female corpus and Almoaily's (2012) suggestions, subject-verb agreement in GPA can take three variants: (1) agreeing GA verb[4], (2) non-agreeing GA verb, where the verb does not agree with the noun, although agreement markers are used; and (3) dropped verb.

Most of the informants in the current project did not use a GA verb where the verb is fully inflected for TMA markers and agrees with the subject in number, gender, and person. For instance, the percentage of AGR (agreement) present tokens (i.e. the verb agrees with the noun in gender, number, and person) was 4.9% among Malayalam informants, 4.5% among Punjabi

[4] In GA, the verb agrees with the noun in gender, number, and person (Feghali, 2004; Holes, 1990; Qafisheh, 1977).

informants, 4.1% among Bengali speakers, 6.6% in the Sinhalese sample, 2.8% among Tagalog informants, and 8.3% in the Sundanese data. Since Malayalam and Sinhalese are the only substrate languages in which the verb does not agree with the subject, Malayalam and Sinhalese informants were expected to use less subject-verb agreement. Thus, the hypothesis that there is a difference in terms of subject-verb agreement between the GPA spoken by speakers with different L1s can be rejected, as the data revealed that they use the GA fully inflected verb forms slightly more than members of the Punjabi, Bengali, and Tagalog language groups, although the difference is still not statistically significant (*p*-value = 0.08). A close look at the data suggests that substrate language groups seem to use more inflected verbs, but few are agreeing forms in more than half of the total number of tokens (65.5%), while the overall percentage of AGR-present tokens is only 5.2%.

## 6. Discussion of Findings – Nominal Agreement

Adjectives in GA agree with the nouns they modify in gender and number. The demonstratives are also distinguished by gender, number, and proximity (see Section 4.1.1). On the other hand, in GPA, adjectives are only rarely modified with gender and number markers, except for the unmarked form of the singular masculine adjective (see Section 4.2.5.2). I noticed that GPA female speakers only used one form of GA demonstrative – *hatha* ('this.M') – with all objects, irrespective of their gender, number, and distance, which Almoaily (2012) also noticed in his GPA male corpus. Thus, if long-term residents produce more forms of noun-adjective agreement in number and gender, and more forms of feminine or plural demonstratives, then this might show accommodation to standard GA.

Regarding the substrate-language effect, Asher and Kumari (1997) stated that the Malayalam predicative adjective agrees with its modified noun in person, number, and gender, and Bhatia (2013) pointed out that the Punjabi adjective agrees with its modified noun in number and

gender, but does not agree in loan words. On the other hand, in the Bengali[5], Sinhalese, Tagalog, and Sundanese languages, adjectives do not inflect for number or gender (Milne, 1993; Gair, 1968; Gair and Karunatilaka, 1974; Schachter and Otanes, 1983; Hardjadibrata, 1985). Thus, I initially expected to find more tokens of noun-adjective agreement in Malayalam and Punjabi speakers than for the rest of the six languages.

The data in Chapter 7 revealed that the Malayalam and Punjabi samples showed better accommodation towards acquiring the GA nominal agreement system. Malayalam informants and Punjabi informants produced instances of noun-adjective agreement in 18.7% and 17.2% of the total number of tokens, respectively. This result can be compared with the percentages relating to the Bengali-speaking (9.4%), Sinhalese-speaking (7.3%), Tagalog-speaking (7.6%), and Sundanese-speaking (6.1%) samples. Hence, the data supports the hypothesis that a difference exists in the use of nominal agreement between the GPA spoken by speakers with different L1s. The chi-square test revealed that the difference between speakers of the six groups is significant at a *p*-value of 0.008. It is also clear from these figures that only the Malayalam and Punjabi language groups show a better performance in acquiring the GA nominal agreement system, as the adjectives in their L1 agree with the modifier in person, number, and gender. Hence, the data is in parallel with the hypothesis that there is a difference between the GPA spoken by speakers with different L1s.

#### *8.2.2.1. A summary of Hypothesis 2 Findings*

**Research Hypothesis 2: There is a difference between the GPA spoken by speakers with different L1s.**

[5] Bengali uses the singular masculine form with all nouns; adjectives do not show any inflection for number or gender (Milne, 1993).

This section provides a brief discussion on the second hypothesis of this study related to the influence of speakers' L1. Table 8.1 summarises the hypotheses on the substrate-based effect and indicates whether the data of this investigation supports them.

| The data revealed that the following hypotheses can be | accepted | rejected |
|---|---|---|
| **There is a difference in terms of the use of the definiteness marker between the GPA spoken by speakers with different L1s** | | ✓ |
| **There is a difference in terms of the use of the conjunction markers between the GPA spoken by speakers with different L1s** | ✓ | |
| **There is a difference in terms of the use of the copula *fi* between the GPA spoken by speakers with different L1s** | | ✓ |
| **There is a difference in terms of the use of the object and possessive pronouns between the GPA spoken by speakers with different L1s** | | ✓ |
| **There is a difference in terms of the use of the subject-verb agreement between the GPA spoken by speakers with different L1s** | | ✓ |
| **There is a difference in terms of the use of the nominal agreement between the GPA spoken by speakers with different L1s** | ✓ | |

Table 8.1: Summary of substrate-language-based hypotheses

As a result, only conjunction markers and nominal agreement (agreement in the NP and in the ADJP) show slight positive correlations between the morphosyntactic features of speakers' (informants') L1 (substrate) and their choice among the available variants of the variable. However, this substrate effect seems to be weak, and it could be interpreted in terms of theories of the emergence of contact languages (of P/C genesis), which claim that contact languages emerge as a result of universal parameters, rather than due to substratal influence (see Ferguson, 1971; Todd, 2003; Bickerton, 1981; Muysken and Veenstra, 1995; and Singh 2000).

### 8.2.3. Research Hypothesis 3 Findings

In testing Hypothesis 3 – that female GPA speakers tend to shift towards GA after spending more than 10 years in the Gulf – the five selected linguistic features on which the findings are

based were presented in Tables 7.1 to 7.49 (Chapter 7). The discussion is in the following subsections, numbered 1 to 6.

### 1. Hypothesis 3 Findings – Variation in Definiteness

A possible link was noticed between the length of stay in GA-speaking countries and the use of the definiteness marker *al-*. This shift towards GA was observed in all six language groups. The newly arrived GPA speakers produced the definiteness marker in 10.7% of the cases, whereas the old members produced them in 33.7% of cases. The chi-square test revealed that the difference between the new informants and those who stayed longer in the Gulf in producing definiteness markers is significant at a *p*-value of 0.002.This noticeable shift towards using the GA definiteness marker among the long-term residents could potentially be a result of the fact that definiteness in GA is one of the easiest morphosyntactic features to learn, as it only involves adding the prefix *al* – or one of its allophones (e.g. *al-*, *el-*, and /ɪ /) – to the target noun.

### 2. Hypothesis 3 Findings – Variation in the Use of Conjunction Markers

The data indicated a major shift towards GA in the use of conjunction markers. This effect was observed in all six language groups. The newly arrived GPA speakers produced conjunction markers in 12.9% of the cases, whereas the old informants produced them in 41.5% of cases. The chi-squared test revealed that the difference between the new informants and those who stayed longer in the Gulf in producing conjunction markers is significant at a *p*-value of 0.001. This significant difference could be due to the fact that learning GA conjunction markers is not difficult. GA conjunction markers are free morphemes (e.g. *wa* 'and', and *aw* 'or'). This result is in parallel with Almoaily's (2013) study of male GPA speakers.

### 3. Hypothesis 3 Findings – Variation in the Use of the Copula

In GA, there is no copula in the present tense. Thus, the focus here is on the use of the copula *fi* in the present tense in GPA. If long-term residents were found to drop the copula more than

the newcomers, this might be an indication of a shift towards GA. The data revealed that the relation between the years of stay and the shift towards GA seems to be slightly negative at a *p*-value = 0.35. Overall, no significant shift towards GA was observed in the data of speakers participating in this study regarding the use of a copula, as new speakers dropped it on average 68.4% of the time, and old speakers dropped it on average 74% of the time.

Length of stay in the Gulf seems to have no effect on the use of the GPA copula *fi* in the data of the Tagalog speakers, bearing in mind that the more the participants drop the copula, the closer they are to the superstrate language. The Tagalog- and Punjabi-speaking newcomer informants dropped the copula on average 77.6% and 71.3% of the time, respectively, whereas old Tagalog and old Punjabi speakers dropped it 69.9% and 63.8% of the time, respectively, which can be interpreted as what Almoaily (2013) refers to as a 'shift to a GPA-internal norm' (i.e. features that did not show a noticeable development, neither towards GPA nor GA). As regards the Malayalam informants, there was a minor difference among them. Newly settled Malayalam speakers in the Gulf dropped the copula 69.7% of the time, and long-term Malayalam residents produced null-copula utterances in 68.6% of the times where a copula could have been used in the present tense. Furthermore, on average, Bengali, Sinhala, and Sunda informants are in contrast to the first three language groups, as the data revealed a possible link between the years of stay and the shift towards GA. All three newly settled language groups dropped the copula in 64.6%, 60.7%, and 66.6% of the time, respectively, while the old Bengali, Sinhala, and Sunda groups dropped it in 78.3%, 86.6%, and 77% of cases, respectively. The more participants dropped the copula, the closer they were to GA.

Overall, the new informants dropped the copula in 68.4% of the total number of cases, while their long-term counterparts dropped it in 74% of the cases. Although the difference is not statistically significant, an indication of development can be seen, especially when comparing

the data of Punjabi, Sinhala, and Sunda speakers. More data, and hence larger numbers, might achieve significant results in this area.

### 4. Hypothesis 3 Findings – Variation in the Use of the Object and Possessive Pronouns

As has been mentioned in Section 4.2.2 about the forms of pronouns in GPA, there are four variants for object and possessive pronouns: the GA bound pronoun which agrees with the noun in person, number, and gender; the GA bound pronoun which does not agree with the noun; the free pronoun; and the dropped object or possessive pronoun. The findings show that on average, newly settled informants in all six language groups produced bound object (both agreeing and non-agreeing) and possessive pronouns in 24.2% of cases, while the long-term residents produced them in 49.2% of cases. The difference is significant at a *p*-value of 0.0001. Note that the newcomers produced tokens of pronouns in free forms 71% of the time, and their old group counterparts produced them in 75.5% of cases. The high rate of free object and possessive pronouns and the fact that free pronouns are even higher in the data of old speakers supports the view that the target here is clearly not towards GA (bound pronouns) but GPA (free pronouns). Since this feature (the free pronoun) is found in all informants' L1s, it could likely have had some influence on the GPA speakers and led them to learn it at the first stage (Section 8.3 provides theories which discuss the substratal influence on the emergence of GPA).

As can be seen in the descriptive account of pronouns in Section 4.1.2, both object and possessive pronouns are bound morphemes in GA and are free ones in all six L1 language groups. Thus, minor differences were expected between the participants in this study as regards their choices among available variants for object and possessive pronouns. For example, new Malayalam speakers dropped object pronouns in an average of 74.3% out of the total number of object pronouns (i.e. AGR+ OBJ, AGR- OBJ, free OBJ, and dropped OBJ) and used them

as free morphemes in only 2% of cases. Possessive pronouns, however, were dropped in 23% of cases and were used as free morphemes 78.2% of the time (see Table 7.19 in Chapter 7). Evidence was also found for the different choices in the data of the old Sinhalese sample, whose members used possessive pronouns as free morphemes in 60.6% of the total number of tokens of possessive pronouns and dropped them in 7.6% of cases where they could have been used. Object pronouns, on the other hand, were dropped in 31.9% of cases and were used as free morphemes only 3.6% of the time (see Table 7.22 in Chapter 7). Differences in the use of object and possessive pronouns were also observed in the data of new Bengalese. They dropped object pronouns and possessive pronouns in 68.7% and 17.7% of cases, respectively. Almoayli (2012) stated that due to these variable choices of using pronouns across almost all language groups, he found it difficult to generalise informants' preferences regarding whether to drop possessive pronouns and use object pronouns as free morphemes or vice versa. Overall, there is a tendency to use possessive pronouns as free morphemes and to drop object pronouns. This is compatible with Almoayli's conclusion that this tendency is a GPA internal development that cannot be explained with recourse to substrate or superstrate languages of GPA. He also refers to this phenomenon (pro-drop) as a common feature in the early stages of SLA (see Eubank, 1991; Towell and Hawkins, 1994; and Montrul, 2004). This could support the potential role of language acquisition in the genesis of GPA.

### 5. Hypothesis 3 Findings – Variation in Subject-Verb Agreement

As has been discussed in Chapter 4, subject-verb agreement can take three variants which have been evidenced in the data. They are (1) agreeing verb, where the verb agrees with the subject in gender, number, and person; (2) non-agreeing verb, where the verb does not agree with the noun although agreement markers are used; and (3) dropped verb, where the verb is dropped and therefore no account of agreement can be given.

In relation to length of stay, the data shows that there is a positive development related to the informant's length of stay in the use of verbs: members of the new group tend to drop verbs more frequently (35.6%) than their old group counterparts (8.4%). The rate of dropping the verb is significantly higher in the data of new informants, with a *p*-value of 0.0002. However, it seems that there is no development in acquiring agreement in the GA verbal system.

Overall, the data revealed that all of the informants rarely produced the form of the verb that is used in GA (i.e. fully inflected verb forms that agree with the subject in number, gender, and person). The overall percentage of newcomers who produced a fully inflected GA verb (i.e. the verb agrees with the noun in gender, number, and person) was only 5.2% of the total number of tokens, while for old informants, it was produced in 8.4% of cases. Yet, the difference was not significant (*p*-value = 0.22). Note, the overall shift is clearly not towards GA, as the use of forms of verb markers which do not agree with the noun in gender, number, and person are predominant in the data of both new and old GPA speakers.

## 6. Hypothesis 3 Findings – Agreement in the Noun Phrase and in the Adjective Phrase

The data revealed a slight positive improvement in the occurrence of nominal agreement among participants who stay long in the Gulf as compared to their new counterparts. The percentages of nominal agreement among members of new Malayalam and Punjabi speakers were 14.5% and 13.4%, while those for old Malayalam and old Punjabi groups were 22.9% and 21.1%. In the Bengali sample, the newcomers only produced nominal agreement in 4% of the cases, while long-term Bengali participants produced tokens of nominal agreement 14% of the time. Similarly, newly arrived Sinhalese speakers produced nominal agreement in 3.4% of cases, while members of the old Sinhalese group produced instances of nominal agreement 11.2% of the time. Moreover, the newly arrived Tagalog members produced nominal agreement with an average of 4.8%, whereas their counterparts who stayed in the Gulf for 10 or more years

produced it 10.4% of the time. The percentage of nominal agreement among new Sundanese-speaking informants was 4.6%, while for old Sundanese informants, it was 7.7%. The new informants generally produced AGR+ tokens in 7.4% of the total number of cases, while their long-term counterparts produced it in 14.5% of cases. Though the difference fails to reach statistical significance (*p*-value = 0.08), a slight development can be noticed, especially in the data of Malayalam and Punjabi speakers. In addition, a vast amount of variation exists within the groups, and there seems to be a trend towards the acquisition of GA.

#### *8.2.3.1. A summary of Hypothesis 3 Findings*

**Hypothesis 3: Length of stay in the Gulf produces accommodation to standard GA.**

Regarding the hypothesis of the current project related to the number of years spent in the Gulf, Table 8.2 displays the investigated features. It shows both whether a shift occurred towards the superstrate language, namely, GA, or towards GPA internal developments.

| Linguistic features showing significant shift towards GA | Yes | No | Linguistic features showing significant shift towards internal GPA | yes | no |
|---|---|---|---|---|---|
| **Acquiring the GA definiteness marker *al*** | ✓ | | **Pronominal system (from dropping to free OBJ and POSS pronouns)** | ✓ | |
| **Acquiring the GA conjunction markers** | ✓ | | **S–V agreement (from dropping the verb to AGR form)** | ✓ | |
| **Dropping the GPA copula *fi*** | | ✓ | | | |
| **Acquiring the GA S-V agreement** | | ✓ | | | |
| **Acquiring the GA nominal agreement** | | ✓ | | | |

Table 8.2: Summary of informants' shift towards GA versus internal GPA

### 8.2.4. Research Hypothesis 4 Findings

In testing Hypothesis 4 – that length of stay in the Gulf produces more accommodation to standard GA in women than men – the five selected linguistic features on which the findings are based (Table 3.8) were compared with those in Almoaily's (2013) study of male GPA data. As a reminder, the database for Almoaily's study (2013) was established via interviews with 16 informants from three linguistic backgrounds: Malayalam, Bengali, and Punjabi. My comparative analysis is thus restricted to those three linguistic groups only.

### 1. Hypothesis 4 Findings – Variation in Definiteness

According to the data, newly arrived female GPA speakers produced a 21.5% average of definite markers, and the old members produced them in 33.5% of cases, whereas the data in Almoaily's (2012) study showed that recent male GPA speakers produced the definiteness marker on average 19.1% of the time, and male speakers who stayed in the Gulf for 10 years or more produced them in 29.6% of cases. There are no statistically significant differences here between men and women's speech (*p*-value = 0.06). Hence, the hypothesis that length of stay in the Gulf produces more accommodation to standard GA in women than men cannot be accepted, as the difference between the two groups is negligible. This demonstrates that gender groups do not display great differences in terms of accommodation to standard GA. These slight differences in the use of the GA definiteness marker between long-term male and female residents could be a result of the fact that the GA definiteness marker *al-* is not difficult to learn, as it requires adding the prefix *al* – or one of its allophones – to the target noun.

### 2. Hypothesis 4 Findings – Variation in the Use of Conjunction Markers

To test the rates of accommodation in the Gulf by gender, I compared the average rate of occurrence of the GA conjunction markers (e.g. *aw* and *ya* 'or' and *wa* 'and') in the data of both newly arrived female GPA speakers (18.9%) and their old counterparts (40.5%) with that produced by newly arrived male speakers (7.6%) and long-term male residents (23.3%.). The difference between the two groups (men and women) in their production of conjunction markers is observable. Women informants tend to use more conjunction markers in their speech than men. The difference between the two groups based on their length of stay in the Gulf in producing conjunction markers is significant at a *p*-value of 0. 0.005. Thus, the hypothesis that length of stay in the Gulf produces more accommodation to standard GA in women than men seems valid.

As has been explained before, GA conjunction markers are easy to learn for both men and women. However, GPA females are more likely exposed to local speech than GPA males due to the cultural and social norms of the country, which allow women to regularly interact with locals, as most of them working as house maids live with locals, and they mostly use GA when communicating with them. On the other hand, GPA male speakers are rarely in social contact with locals; most of them have a brief business conversation with natives and spend most of their time interacting with their co-workers, who come from different linguistic backgrounds using GPA.

### 3. Hypothesis 4 Findings – Variation in the Use of the Copula

The number of years of residency for both genders seems to have no difference on the use of the GPA copula *fi* in the data of men and women speakers. As a reminder, there is no copula in the GA present tense, thus the focus here is on the use of the copula *fi* in the present tense in GPA. If long-term residents were found to drop the copula more than newcomers, this might be an indication of a shift towards GA. On average, the new female-speaking informants dropped the copula 69% of the time, and old female GPA speakers dropped it in 70.3% of cases, whereas new male speakers dropped it on average 68.2% of the time, and old male speakers dropped it in 69% of cases. The difference between male and female accommodation towards GA fails to reach significance ($p$-value = 0.22). Hence, the hypothesis that length of stay in the Gulf produces more accommodation to standard GA in women than men cannot be accepted, as the difference between the two groups is negligible.

### 4. Hypothesis 4 Findings – Variation in the Use of the Object and Possessive Pronouns

As a reminder, pronouns have four different forms in GPA for object and possessive pronouns: the GA bound pronoun which agrees with the noun in person, number, and gender; the GA bound pronoun which does not agree with the noun; the free pronoun; and the dropped object

or possessive pronoun. The data revealed slight differences in the use of GA bound pronouns between newly settled male and female GPA speakers who spent five years or less in GA-speaking countries. This difference was evidenced by new female informants, who used bound pronouns in 17.3% of cases and more than new male informants (11.5%). This is in contrast to their long-term counterparts, who exhibited no differences in the use of GA bound pronouns. On average, old female and old male GPA speakers produced them in 22.7% and 24.9% of cases, respectively. The difference between male and female accommodation towards GA fails to reach significance (*p*-value = 0.07). Hence, the hypothesis that length of stay in the Gulf produces more accommodation in the use of the object and possessive pronouns to standard GA in women than men cannot be accepted, as the difference between the two groups is negligible.

### 5. Hypothesis 4 Findings – Variation in Subject-Verb Agreement

The agreement between subject and verb in GPA can take three variants, which have been found in the data: (1) agreeing verb, where the verb agrees with the subject in gender, number, and person; (2) non-agreeing verb, where the verb does not agree with the noun although agreement markers are used; and (3) dropped verb, where the verb is dropped and no account of agreement can thus be given. The data revealed no differences in the acquisition of agreement in the GA verbal system based on length of stay among male and female GPA speakers. On average, the new female informants produced a fully inflected GA verb only in 3.7% of the cases, and new male informants produced it in 3.1% of cases, while old female and old male informants produced it in 3.11% and 3.8% of cases, respectively. The difference between male and female accommodation towards GA subject-verb agreement fails to reach significance (*p*-value = 0.07). Again, the hypothesis that length of stay in the Gulf produces more accommodation to standard GA in women than men cannot be accepted, as the difference between the two groups is negligible.

## 6. Hypothesis 4 Findings – Agreement in the Noun Phrase and in the Adjective Phrase

As for the gender comparison of variation due to the length of stay, the data revealed a noticeable positive effect on the occurrence of nominal agreement for female over male informants. On average, the new female informants produced AGR+ tokens in 36% of the total number of cases, whereas they were produced by new male informants in 12% of cases. In contrast, long-term female speakers produced AGR+ tokens in 69% of the total number of cases, and long-term male speakers produced them in 25% of cases. Although the difference misses statistical significance (*p*-value = 0.09), an indication of development can be seen in GPA female production. Yet again, the hypothesis that length of stay in the Gulf produces more accommodation to standard GA nominal agreement in women than men is rejected. More male data, and hence larger numbers, might achieve significant results in this area.

### *8.2.4.1. A summary of Hypothesis 4 Findings*

**Hypothesis 4: Length of stay in the Gulf produces more accommodation to standard GA in women than men.**

This section provides a brief discussion on the hypothesis of the current project related to the gender variation based on number of years spent in the Gulf. Tables 8.3 and 8.4 illustrate the data comparison between male and female GPA speakers based on the potential influence of years of residency in a GA-speaking country. Table 8.3 displays the investigated features in this study and indicates those where female GPA speakers significantly shift towards the superstrate language, GA, and well as for the male GPA speakers in Almoaily's data. Table 8.4 lists the linguistic features for which female GPA speakers show more accommodation to GA than men based on their length of stay in the GA country. It also displays a summary of differences between men (in Almoaily's GPA male corpus) and women (in my corpus) based on length of stay.

| Linguistic features showing significant shift towards GA among women | yes | no | Linguistic features showing significant shift towards GA among men in Almoaily's study | Yes | No |
|---|---|---|---|---|---|
| **Acquiring the GA definiteness marker *al*** | ✓($p$ = 0.002) | | | | ✓($p$ = 0.08) |
| **Acquiring GA conjunction markers** | ✓($p$ = 0.001) | | | ✓($p$ = 0.002) | |
| **Dropping the GPA copula *fi*** | | ✓($p$ = 0.35) | | | No reportable |
| **Acquiring the GA pronominal system** | | ✓($p$ = 0.0001)[6] | | | ✓($p$ = 0.03)[7] |
| **Acquiring GA S-V agreement** | | ✓($p$ = 0.0.07) | | | No development |
| **Acquiring GA nominal agreement** | | ✓($p$ = 0.08) | | | ✓($p$ = 0.054) |

Table 8.3: Summary of statistical significance in men and women based on length of stay.

| Linguistic features showing a significantly greater shift towards GA among women than men | Yes | No |
|---|---|---|
| **Acquiring the GA definiteness marker *al*** | | ✓ |
| **Acquiring GA conjunction markers** | ✓ | |
| **Dropping the GPA copula *fi*** | | ✓ |
| **Acquiring the GA pronominal system** | | |
| **Acquiring GA S-V agreement** | | ✓ |
| **Acquiring GA nominal agreement** | | ✓ |

Table 8.4: Length of stay in the Gulf produces more accommodation to standard GA in women than men.

Possible differences can be found in the rate of use of the five selected features of GA morphosyntax between the genders. In this study, the rate of use of morphosyntactic features was higher in female than in male GPA speakers. Statistically, only one feature (conjunction markers) displays a higher tendency to accommodate to standard GA in women than men if they stay longer in the Gulf countries.

[6] Although it is significant, the overall shift is not towards GA, it is a development towards GPA norms.
[7] Although it is significant, the overall shift is not towards GA, it is a development towards GPA norms.

An overall result is that only conjunction markers and nominal agreement (agreement in the NP and in the ADJP) show slight positive correlations between morphosyntactic features of speakers' (informants') L1 (substrate) and their choice among the available variants of a variable. However, this substrate effect seems to be weak, and it could be interpreted in terms of theories of the emergence of contact languages (of P/C genesis), which claim that contact languages emerge as a result of universal parameters, rather than due to substratal influence (see Ferguson, 1971; Todd, 2003; Bickerton, 1981; Muysken and Veenstra, 1995; and Singh, 2000). In subsections 8.2.4.2, 8.2.4.3, and 8.2.4.4, the gender variation of GPA is considered from a sociolinguistic point of view.

#### *8.2.4.2. What the Findings Revealed in Regard to Gender Variation*

The results from Table 7.51 in Section 7.2.3 suggest that women GPA speakers shift towards standard GA more than men in their rate of choices among all six linguistic variables. However, this shift is significant in only one feature, namely, the conjunction (*p*-value = 0.005). The following subsections provide possible explanations for this result, which fall into two points of view: why women show more accommodation towards GA than men, and why it is significant in only one feature.

#### *8.2.4.3. Why Women Show more Accommodation Towards Gulf Arabic than Men*

In a community where social structures and norms control people's behaviour, the way they think reflects in the language they use. Since our knowledge and beliefs constitute our culture, they are habitually encoded and transmitted into language. In turn, language becomes a medium of communication through which the members in the community participate in a variety of social activities; this relationship is described as follows (Alqasem, 2017):

> Speakers of different languages and cultures living in such speech community and trying to acquire the language of the community may feel unsatisfied about the level of proficiency they reach because of the shortage of particular kind of words that they don't hear or they may hear it rarely. This may influence and limits the kind of language they acquire and which they use in their speech. Absence of kind of words and some

> linguistic structures of the language constitute an obstacle in a person's utterances that he/she consciously or unconsciously makes up for the shortage by using other words that he/she thinks they are right. (p. 44)

This situation affects the acquisition of an L2, and Terdal (1993: 25) emphasises the role of social interaction in SLA theories by stating that 'language is learned through social interaction'. For example, the highly conservative Saudi society is controlled by some social norms and restrictions, such as gender segregation. This constraint (i.e. separation between the two genders – male and female) applies to all men and women who live in this community, including foreigners. Alqasem (2017) suggests that this restriction has a great impact on GPA language acquisition when GPA speakers come into contact with a particular community such as the Saudi one (see detailed discussion in Chapter 2, Section 2.4).

Based on the observed situation in SA, there are several potential explanations for why GPA women show more accommodation towards standard GA than men in their rate of choices among all six linguistic variables. First, the majority of GPA male speakers work in places such as restaurants, shopping malls, and hospitals that require them to interact with both males and females. They deal mostly with men, and when it comes to interacting with women, Asian foreigners do not feel confident, become cautious, and use only a few words which they repeat in nearly all cases (Alqasem, 2017). While GPA female workers, most of whom are domestic helpers (or housemaids), also have roles that require them to interact with two genders, the difference is that those housemaids live with the family that employs them and grow accustomed to daily interactions with the whole family, thus leading them to feel more confident and less worried. This in turn enables them to freely communicate with their employers (who are native speakers), rapidly learn the language of the host community, and effortlessly adapt to the system of GA (the lexifier language).

The second reason is that accommodation towards standard GA basically depends on the GPA speaker's desire to learn the language. GA becomes a TL for GPA speakers in some cases. (1)

This occurs in rare occasions of marriages between immigrants and natives (intermarriages). Intermarried immigrant females have on average better education, and they learn language skills and the culture of the home country better than non-intermarried immigrants (see Nottmeyer, 2015; Chiswick and Houseworth, 2011; Furtado, 2012; Meng and Gregory, 2005; Bakkir, 2010). This intermarriage 'closes the socioeconomic gap' between minority and majority groups (see Section 2.8, Bakker, 2008 and Basu, 2017 for more details about the social gap in the dominant community of the locals), which enables them to adapt better to a new social environment and facilitates communication with natives. Intermarriage also increases language proficiency and the respective individuals' number of years of residence in the Gulf States. (2) The nature of female GPA speakers' work is better for defining the TL compared to that of their male counterparts . GA can be the TL of female GPA speakers when they work as domestic helpers. Local families use GA to communicate with their female domestic helpers who live with and work for them for long periods. This interaction between native Arabic speakers of the Gulf and immigrant workers from South and South-East Asia affords Asian female workers a prime opportunity to practice new cognitive skills and learn GA in real-life settings, unlike male workers who have limited social opportunities to interact in GA with their bosses (see Section 8.9). In fact, this scenario appears to be the case for the vast majority of female GPA speakers, including the participants polled in this study. (3) Some Gulf families prefer their female housemaids and nannies to be Muslims, since these females live with them and spend long hours caring for their children, although Muslim workers mostly have at least sufficient knowledge of Arabic, which allows them to recite and read the Quran, and this in turn enables them to attain a certain level of fluency, as I noticed among my informants.

Finally, GPA speakers tend to replicate the expressions and patterns to which they are often exposed in their speech (see detailed discussion about females' ability to learn an L2 in Section 2.10.2). In this case, hearing and communicating with females in the society is the authentic

input that supports the recognition of the language, which in some aspects is geared towards females and by which females are addressed (Alqasem, 2017).

#### *8.2.4.4. Why the Results were Significant for only One Feature (the Conjunction)*

Although female GPA speakers shift to GA more than male GPA speakers in all linguistic features after spending 10 years or more in the Gulf, this shift is significant in only one feature, namely, the conjunction (*p*-value = 0.005). This finding leads to two separate, but related questions: First, why was the conjunction marker the only statistically significant feature, and second, why was there only one significant feature?

One possible explanation for why women produced significantly more instances of GA conjunction markers than men is that these markers are easy to learn; they are free morphemes, and most of them are one-syllable words (e.g. *aw* and *ya* 'or', *wa* 'and'). Another possible reason is the quality of the input that GPA speakers receive from locals. This fact is supported by Almoaily (2008), who wrote that 'conjunction was the only feature where locals use its GA variant more than its GPA variant (72.2% versus 27.8%)'.

A possible answer to the second question (why there is only one significant feature among the six linguistic features investigated in this study) is that all of the polled females in the current study and males in Almaoily's study were from SA. As discussed above, the nature of the construction of the Saudi community could limit the contact between the two sexes in almost all aspects of life. Men are dominant in the society in most daily life activities, such as trade, administration, and business, while women perform work that men are not allowed to carry out, even in trade or medical industries, and all rules and norms that apply to locals also apply to foreigners, as they live in that community (Alqasem, 2017). The language of daily life activities in Saudi society is mostly gender-based. This language is used to communicate among the local speech community and with the foreigners who work in different fields. Immigrants may spend several years in the community without any contact with or exposure

to the language of the opposite sex, except in a few cases. Since the language of one gender is absent from that of the other gender, misunderstandings may occur in different aspects of the language, thereby causing differences in the recognition of some aspects of the acquired language, such as pronouns, verb agreement, and noun and adjective agreement, as all of them should agree in number, person, and gender with their nouns and modifiers. In turn, this absence of one gender's language will influence the speech of the learners, causing misunderstandings and confusion around various aspects of language when FWs use the language to address either natives or their co-workers, such as when using pronouns[8] (Alqasem, 2017). Alqasem (2017) indicates that the speakers of Saudi Arabic pidgin mostly cannot correctly distinguish between the use of a pronoun and its suffix[9] in the case of a particular gender – male or female. The example below highlights that GPA speakers can use the pronoun *inta* to address both males and females regardless of the fact that the pronoun is masculine:

(1) inta kalam keda keda (S7)

you.SGM speech this this

'You tell me to do so and so'

In fact, the limitation of using any structure in a language or misusing that structure could cause the learner to make errors. Thus, GPA speakers, whether male or female, tend to use an alternative word or structured thinking that will fulfil the shortage in their language. The next sections investigate the potential factors which may lead to the genesis of GPA, namely, substratal and universal factors.

[8] Pronouns in GA are mostly inflectional and can be distinguished in the case of feminine and masculine by the suffixes that are normally added to the noun.

[9] The second-person singular pronoun in GA inflects for gender (i.e. *int* 'you.SGM' and *inti* 'you.SGF'). In Arabic, possessive pronouns are suffixes added to the noun.

### 8.3. Theories on the Genesis of Pidgins and Creoles that Can Be Applied Based on the Results of the Current Study

Section 3.1 introduced theories pertaining to the origin of both pidgins and creoles. These theories of genesis are restricted to the purposes of the current study. First, the theory advocating that P/Cs consist of restructured varieties of their substrate language is called substratal influence. Second, the theory that creoles consist of language varieties reflecting the properties of universal grammar is the universalist theory (referred to in the literature as the LBH). Third, the theory stating that P/Cs are imperfect L2 varieties of their lexifier languages is called *imperfect SLA*. These current theories of P/C genesis are mainly based on the assumption that P/Cs were created from the contact situation in Africa[10].

The problem of the origin of P/C genesis has been addressed from different points of view over the last century and is still subject to controversy. As discussed in the literature review, the number of non-European-based P/C studies is limited compared to the number of European-based P/C studies. One of the aims of this study was to increase knowledge and awareness of the nature of P/C languages – such as the less-studied Arabic-based pidgin, GPA – and help to propose theories and assumptions that describe the emergence and development of contact languages as well as to list the typological features in the world's P/Cs.

The six languages tested in this study (i.e. Malayalam, Punjabi, Bengali, Sinhala, Tagalog, and Sunda) are typologically dissimilar (they have different lexical affiliations). This allowed me to test, for instance, Almoaily's suggestion that when comparing languages which are typologically similar (i.e. Bengali, Malayalam, and Punjabi), no substrate effects may be found, unlike typologically dissimilar languages, such as the ones investigated in this study (Tagalog, Sinhala, and Sunda), which might produce more visible effects. It is advantageous to seek more

[10] A situation where contact languages took place in Africa in the pre-colonial period, colonial period, and post-colonial period among people who need to communicate but do not share a common language (Isa, Halilu, and Ahmed, 2015).

evidence by using the data of less-studied P/Cs, such as the Arabic-based pidgin language. If the present study revealed that the participants produced morphosyntactic features which can be traced to their L1s and which show a significant effect on their choice among GPA variants, then it would support substratist theories of genesis. Likewise, if participants produced universal linguistic features which cannot be traced to their L1s and did not show a significant effect on their choice among GPA variants, then this would support universalist theories of genesis. The next sections determine which of the two potential factors (i.e. substrate influence and universal influence) that lead to GPA evolution is more significant, based on the findings reported in Chapter 7.

### 8.3.1. The Influence of Substrate Languages on Gulf Pidgin Arabic

The finding is significant for two GPA features (conjunction and nominal agreement) which show a potential substratal influence. From a substratal point of view, the linguistic features of *definiteness* and *verbal agreement* showed no relation between speakers' L1 and GPA variants. Despite the fact that object and possessive pronouns are freely used by speakers of all six substrate languages and consequently no difference in their use was expected between the GPA speakers, I found that Malayalam speakers produced a significantly higher number of bound pronouns. This leads to the finding that there is no substrate influence on the use of pronouns among the GPA speakers participating in this study.

In summary, a significant correlation between the informants' L1 and their choice among GPA variants was found in two features: *conjunction* and *nominal agreement.* The other three features (i.e. definiteness, copula, and verbal agreement) showed no relation. Even though a significant correlation has been manifested in only two linguistic features, a substrate influence is statistically evident among the dissimilar systems of speakers' L1s. This conclusion validates Almoaily's suggestion that languages which are typologically dissimilar might have more

visible substrate effects. In Table 8.5, I list the GPA features which show a potential substratal influence.

| **Feature** | **Conjunction** | **nominal agreement** |
|---|---|---|
| ***p*-value** | ($p < 0.001$) | ($p < 0.0008$) |

Table 8.5: Potential substrate influence on language variation in GPA.

The table suggests that substrate influence is statistically evident in two morphosyntactic features. However, more evidence for substrate influence on GPA variation might surface if a larger corpus could be constructed.

### 8.3.2. The Universal Influence on the Emergence of Gulf Pidgin Arabic

As discussed in Section 3.1.1.4, universalist theory discusses the universal similarities in grammar that both pidgins and creoles share, irrespective of their different contributing languages, by relating them to 'universal aspects of the human linguistic capacity' (Muysken and Veenstra, 1995: 121). Moreover, what distinguishes this theory is that it focuses on how humans create languages which are similar as somehow due to universal tendencies among the languages' speakers rather than the influence of the languages themselves. Hugo Schuchardt, in the chapter devoted to the language of the Saramaccans, reported by Knapik (2012), evokes the notion of universal principles:

> Creole dialects have not yet been fully appreciated for their general linguistic significance. They are customarily regarded as products of very peculiar or extreme mixture, but what distinguishes them is, rather, if I dare say so, their universal linguistic features. (Schuchardt [1909] 1979: 73)

Todd (2003) also writes about the role of innate tacit knowledge in the process of P/C formation:

> There are universal patterns of linguistic behavior appropriate to contact situations pidgins and creoles are alike because, fundamentally, languages are alike, and simplification processes are alike human beings are biologically programmed to acquire

> *Language* rather than any particular language, and ... the programming includes an innate ability to dredge one's linguistic behavior of superficial redundancies.

Common features have been manifested in the description of GPA (see Smart, 1990; Naess, 2008; Almoaily, 2008–2012; Alshammari, 2010; Bakir, 2010; Albaqawi, 2016; and Section 2.1.2 of this thesis). Comparing these features to their source languages, pidgins exhibit a characteristic simplification in their linguistic structure that relates to all aspects of grammar: lexicon, phonology, syntax, semantics, and morphology. Since this study focuses on the level of morphosyntax, my investigations were only on the morphosyntactic features of this contact language. However, McMahon (1994: 254) points out that 'pidgins do not simplify the grammars of their superstrate and substrate languages, but also restructure them to produce a new linguistic system'.

To answer this question regarding which particular structures are likely to show up in a pidgin, Sebba (1997) made a brief but comprehensive summary describing this reduced structural system by means of universal principles that were involved in the process of pidginisation (refer to Section 3.3 for details of the linguistic features discussed in this thesis). Some of these principles are also valid for creoles: lack of grammatical and morphological complexity – pidgins in general show a preference for semantic transparency as well as some characteristic reduction in vocabulary. Moreover, another characteristic feature[1] of pidgin languages is their higher tendency to simplify and reduce phonology (refer to Smart, 1990; Almoaily, 2008; Naess, 2008; Albaqawi, 2016; and Section 3.3.1 of this study for a full inventory of GPA and GA phonemes). This feature will not be discussed, as the main purpose of this research is to deduct GPA morphosyntax.

[1] This feature is not mentioned by Sebba (1997), but it is another general characteristic feature of pidgin languages.

Contact languages generally display structural similarities even though these languages arise in geographically different areas and are based on different lexifier languages where the structural features are common to P/Cs but not in their L1s. The next subsections discuss all the universal principles (see Section 3.3.2) ascribed to pidgin grammars which are also found in GPA.

### 1. Lack of Grammatical Complexity

The syntactic system of pidgins reveals a lack of surface grammatical complexity. P/C languages are documented to be simplified structural versions of full-fledged languages; this is observable evidence when it comes to grammar (Siegel, 2008). Moreover, Siegel (2004: 140) refers to the simplicity in P/C languages by stating that '[i]n P/C studies, the evidence given for simplicity in a pidgin or creole most commonly includes characteristics such as the absence of inflectional morphology, a low number of marked grammatical categories, small lexicon, or few stylistic options'. GPA, as discussed in Section 4.2, is characterised by a lack of grammatical features, such as no definite article; a reduction in overall sentence complexity; and tense, aspect, modality, and negation marked externally to the verb. These features could exist in GPA due to the universal principle of language simplification. Next, I discuss some GPA features displaying a lack of grammatical complexity.

- **No Definite or Indefinite Article**

There is typically no definite or indefinite article in pidgin languages. GPA speakers normally drop the GA definiteness marker. Moreover, it has been recorded that in GPA, there is no observable marker for definiteness (see Smart, 1990; Almoaily, 2008 and 2012; and Section 4.2.1 of this thesis). The notion of definiteness and indefiniteness seems to be fully contextualised in GPA. The GA definiteness marker *al-* was occasionally expressed by informants in the current study (refer to Chapter 7 for the number of occurrences for each

informant). Examples (2) and (3) below from my data show occasions where the informants sometimes dropped the GA definiteness marker *al-* and sometimes retained it, respectively:

(2) *awal* *ana* *šughul* *fi-al-ʕiyadh* (M9)
first I work in-DEF-clinic
'I used to work in the clinic'

(3) b*ʕdyeen* *yi-ji* Ø *mustašfa*
Then 1SG.F.PRS.come Ø hospital
'After that I came to work in the hospital'

- **Reduced Sentence Complexity**

Pidgin languages display a general tendency to reduce the overall complexity in the structure of sentences, such as subordination and embedded sentences. Alghamdi (2014) statistically reports that an SVO sentence structure is undoubtedly preferred to other sentence structures and patterns such as VSO, while Almoaily (2012) indicates that SOV sentences are mutually used. Additionally, Jenkins (2003), for example, points out another linguistic feature that characterises sentences in jargon or pre-pidgin, namely, articulation and reiteration, which refers to the tendency to express ideas in lengthy sentences. Moreover, Ansaldo and Matthews (2004) highlight the idea of stressing meaning by repeating the same word as also one common linguistic feature in contact languages. In fact, I found that my informants frequently repeated the same word to emphasise its meaning. For example, the word *miy:ah*, which means 'one hundred' in English, is repeated and reduplicated by GPA speakers (i.e. *miy:ah miy:ah*) to create a new meaning 'perfect'. Similarly, the word *kiða* ('this way') is repeated (i.e. *kiða kiða*) to indicate direction or behaviour ('do not go left or right or do not manipulate with me'). Moreover, *nos* ('half') is repeated (i.e. *nos nos*) to convey the following meanings: 'not so

good' and 'not complete'. Similarly, the word *sawa* ('together') is repeated (i.e. *sawa sawa*) to form a new meaning 'two or more similar things'.

- **Tense, Aspect, Modality, and Negation Marked Externally to the Verb**

A common linguistic feature in contact languages that characteristically marks the verb is TAM and negation. As discussed above (in Section 4.2.5.1), in P/C languages, the verb is usually uninflected, and it is indicated by individual and invariant lexical items preceding the verb. In the GPA data, I noticed that speakers could use both tenses (past and present) without a unified temporal reference. GPA speakers could use a reference in the present to express a verb in the past interchangeably**.** If the reference to the tense is not clear from the context, then GPA speakers tend to use time adverbial items such as *awwal* ('first'), *alhen* ('now'), *qabel* ('before'), and *baaden* ('later'). Næss (2008) indicates that the common GPA negation would be expressed by the negation particle *mafi*, which is used for non-verbal negation as well as for imperative verbs. Some examples from my data are demonstrated below where *mafi* is used with verbs to mark aspect (past tense):

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| (4) | *ana* | *hena* | *thalatha* | *sana* | *mafi* | *ruh* | (P6) |
| | I | here | three | years | NEG | go | |

'I have been here (in Saudi Arabia) for three years and I did not go home'

**2. Lack of Morphological Complexity**

One of the most common characteristic features in the field of pidgin typology is the lack of morphological complexity: The reduced morphological complexity of pidgins is due to the lack of inflections, which may even be completely absent in pidgin languages (refer to Todd, 1980; Drecshel, 1996; Hudson, 1996; Holm, 2000; Siegel, 2008; Bakker and Matras, 2013; and many other researchers). GPA displays a heavily reduced number of inflections in number (singular/plural), tense, gender, grammatical agreement (verbal and nominal system), or

grammatical cases (nominative, accusative, genitive) compared to its lexifier language (see Section 3.3.2.2). In GPA, nouns and pronouns have reduced inflections in the following:

- **Number**: The plural of nouns is not marked by inflection (i.e. the noun remains uninflected). Instead, any noun or adjective following a number always comes in the indefinite, singular, masculine form, regardless of the number. Example (5) from my data exemplifies the singular form of the noun *sana* (‘year’), which is used with the feminine form of the number *thamanyah* (‘eight’) when talking about age:

(5) *Omer walad ana thamany-ah sana* (S6)

Age boy me eight-F year.SG.F

‘My son is eight years old ‘

- **Case:** In GPA, the object form is not marked. The subject pronoun *ana* functions as an object pronoun. For example,

(6) *Kulo šughul bayt ana* (T5)

Everything work home I

‘All work in the home is done by me’

- **Gender:** Gender distinctions are usually missing. For example, in GPA, *huwa* is used to stand for both male and female referents: *hum* (‘they.M’) and *hin* (‘they.F’). Likewise, GA the second-person singular pronoun inflects for gender (i.e. *int* ‘you.SGM’ and *inti* ‘you.SGF’). In GPA, however, *inta* can be used with male and female speakers.
- **Possession:** In GA, object and possessive pronouns are suffixes attached to the noun or to the verb, while in GPA they are free morphemes. Moreover, the pronoun, for example, *ana* (‘I’), is used as subject, object, and possessive pronoun (see also Smart, 1990 and Almoaily, 2008 and 2012).

- **Tense:** Lack of morphological complexity is also reflected in the use of verbs. They are typically not inflected to mark tense. The verb usually stays uninflected, and separate adverbial lexical elements are used to mark the tense of the verb and to specify temporal reference. Some of these adverbial elements are *awwal* (‘first’), *alhen* (‘now’), and *baaden* (‘later or after’).

(7) *Mam ana kalam inta ruuh alhaeen Sri Lanka baaden θalaθah shaher yi-ji* (S7)
Mother I said you go.SGM.IMP now Sri Lanka then three month 3SG.M.PRS.come
‘My mother said you can now go to Sri Lanka and stay there three months then after that you can go back to Saudi Arabia’.

### 3. Preference for Semantic Transparency

Pidgins show a general tendency towards semantic transparency. One of the universal principles of pidgin languages is that they compound their lexicons in a semantically transparent way. The meaning of a word can thus be determined easily from the meanings of its constituents from which the word is built (see Baayen and Schreuder, 2003 and the discussion in Section 3.3.2.3, C). GPA seems to share this characteristic with the universal principle of pidgin languages. In GPA, gender is indicated by the use of a separate word: *walad* for male and *bent* for female:

*besa* = cat
*besa walad* = *GA.bes* ‘male cat’ (Su11)
*besa bent* = GA.besa ‘female cat’ (Su11)

Moreover, GPA speakers combine the GA words *saghiir* (‘small’) and *kabiir* (‘big’) with other words to create new GPA meanings:

*Mustashfa saghiir* =GA.*mustawsaf* ‘small hospital’ (T7)
*Saiyara kabiir* = GA. *Shahena* ‘truck’ (M4)
*Hurma kabiir* = *GA. adjuz* ‘old woman’ (B12)

### 4. Reduction in Vocabulary

Overall, the lexical system in pidgins has a small stock of words (Singh, 2000; Mühlhäusler, 1997; Samarin, 1971; Romaine, 1988; and Siegel, 2008). The reduced number of content

words, function words, compounds, and prepositions in the pidgin lexicon is a feature indicating simplicity in GPA. However, even though pidgins have a small size of vocabulary, it is sufficient to enable pidgin speakers to convey the meanings (Romaine, 1988 and 1992). As Sebba (1997) indicated, this is because pidgin languages contain a minimal number of synonyms. For example, the word 'yesterday' can be translated into GA as *ams*, *albarh*, *albarha,* and *alliyla almadiyah*, whereas in GPA it can be translated as *awal.* The GA existential particle *fi*(*h*)[2] also has multiple meanings in GPA. For example,

(8) The copula *fi* means 'there is':

*Ana fih θalaθah sana* (P9)

'I work here for three years'

(8) The following example shows the use of *fi* as 'is there':

*Fi hurma souq*? (B12)

'Is there a woman driver?'

(10) This example indicates that *fi* is used as a verb:

*ana ashouf eash fii akal* (B12)

'I will look then decide what I will eat'.

The discussion above concludes that GPA involves a mechanism that explains the striking structural similarities among P/C languages on the basis of the assumption that all humans are characterised by an innate ability to simplify language. These structural similarities are the result of universal strategies for language simplification which are shared by all humans around the world despite the different contributing languages. These features can be taken as evidence

[2] It means 'there' (Smart, 1990).

in favour of universalist theories of P/C genesis, since they cannot be linked to the linguistic systems of the substrate languages. As I previously mentioned, the six languages tested in this study are typologically dissimilar and may show some substratal effects when comparing them. For instance, although Malayalam and Sinhalese are typologically dissimilar and the only substrate languages of the six which lack subject-verb agreement in their morphosyntactic systems (as I have described in Chapter 5), the data of this study demonstrates that the two substrate languages of GPA use the GA fully inflected verb forms slightly more than members of the other substrate languages. Thus, GPA exhibits features typical of pidgin languages that cannot be linked to its substrate languages but could support universalists' claims that there are universal parameters leading to the emergence of pidgins. Yet, the data revealed that the effect of substrate languages on my informants' choices as regards the studied morphosyntactic features is significant only in two features: conjunction markers and definiteness. Thus, the potential substratal role in the emergence of GPA cannot be totally rejected. In addition to the previous sections that discussed all the universal principles found in GPA grammar, the emergence of GPA must also be considered from a language acquisition point of view, which is done in the next subsection.

### 8.3.3. Is the Process of Learning Gulf Pidgin Arabic Endorsed by the Foreigner Talk Theory or the Imperfect Second-Language Acquisition Theory?

The FT theory proposes that lexifier speakers deliberately attempt to simplify their languages when interacting with speakers of other languages; in other words, FT refers to the way in which native speakers intuitively alter their language when communicating with a non-native speaker who is not fluent in their language. It is thus the native speakers who produce this kind of speech, not the foreigners. The idea behind the FT theory is that non-natives are provided with a simplified model of the lexifier language. This theory therefore explains that pidgins have simple linguistic features compared to their lexifiers because of the type of input that was

given (Velupillai, 2015). This theory hypothesises that contact varieties are the result of constant attempts by lexifier speakers to simplify their language to suit their non-native partner. As discussed in Section 3.1.2.1, some researchers, such as Ferguson (1971), Miller (2003), and Singh (2000), link P/C genesis with FT theory, while others, such as Smart (1990) and Bakir (2010), believe that GPA is not an FT variety and instead refer to it as a pidgin. Smart (1990) suggests that GPA is a simplified form of speech used as a means of communication not only between lexifier speakers and non-native workers, but also among non-native workers themselves who do not share a common language. This strongly suggests that FT is not used solely by one group – native speakers of the lexifier language – but used by several linguistic groups in contact in the Gulf region. Moreover, the findings of the current study show that GPA has rule-governed morphosyntactic features, suggesting that this variety is not merely FT (further discussion can be found in Section 3.2). In addition, Lefebvre (2004) criticises the FT theory for failing to explain certain points. First, this theory does not clarify why P/Cs are manifested only among some linguistic groups. Second, the theory does not emphasise the idea of having a lingua franca for less prestigious languages. Finally, it does not explain the abrupt emergence of pidgin languages.

Unlike the FT theory, the imperfect L2 learning theory suggests that non-native speakers of the TL, the lexifier, aim to acquire the TL; however, due to circumstances of the contact situations, such as social differences, they use only limited knowledge of the TL and are not able to fully acquire their model target, and '[t]heir unsuccessful efforts result[ed] in the pidgin: a grammatically impoverished version of the lexifier with a very restricted vocabulary' (Sabba 1997: 79). This could be another reason that Asian foreigners are not able to learn the language systematically as one would when attending language courses, for example. This may be particularly relevant for morphologically rich languages such as Arabic (inflections are notoriously difficult and must often be learnt systematically and by heart). Indeed, some P/C

researchers, such as Alleyne (1980), Mufwene (1990), Schumann (1986), Siegel (2008b), and Thomason and Kaufman (1992), link the pidginisation process with the imperfect SLA process. Although many characteristics in P/C languages cannot be explained by the imperfect L2 learning theory, its role is obvious in the emergence of P/Cs (Singh, 2000). Moreover, the findings and the discussions of the current study revealed that adults acquire this pidgin only as an L2, as it seems that both contact languages and the process of imperfect L2 learning could be a result of simplified linguistic rules. Yet, researchers still argue about the relation between pidgin-formation and SLA processes. For example, Siegel (2008b: 208) writes, 'while more creolists today may agree about the involvement of processes of SLA in P/C genesis, there is no consensus about exactly what these processes are and how and when they apply'. Another view was tested in Klein and Perdue's study (1997). They pointed out that although *pidgins* and the *basic variety* (i.e. the language necessary for communication as they define it) are acquired as 'adult second language learners outside the classroom and they universally develop a well-structured, efficient, and simple form of language' (1997: 301), learning a pidgin and learning an L2 are still two distinct things. Finally, Klein and Perdue (1997: 340) write that 'there are certainly similarities, but it is quite unclear how far-reaching these are'.

Almoaily (2013) emphasises the idea of the nature of acquiring an L2 as a challenging keystone linking the emergence of P/C languages and SLA with two concepts: input and motivation. First, the quality of input from the superstrate language in pidgin speakers might not be similar to the input given to L2 learners. Indeed, to some extent, the situation for both male and female GPA speakers could be the same as for those L2 learners. Based on the results of this study, female GPA speakers show more accommodation towards GA than male GPA speakers. Although it is only significant in conjunction markers (*p*-value = 0. 0.005), there is noticeably more variation in women than men (Table 8.6 lists the GPA features which show a shift towards GA among male and female GPA speakers in two features: definiteness and conjunction).

Furthermore, some development has still occurred; the long-term female residents produced the GA definiteness marker in 33.5% of the total number of cases, while male long-term residents produced them in 29.6% of cases. Even GA nominal agreement was used 19.6% of the time by old female speakers who spent 10 years or more in the Gulf and was used 14.5% of the time by old male speakers. There seems to be a trend towards the acquisition of GA norms among female GPA speakers. A possible explanation is that female GPA speakers might have more GA input compared to male GPA speakers. This might be due to the nature of their work, as most of the females are maids or housekeepers with constant contact with their GA employers, while males are only occasionally in contact with them and mostly use GPA rather than GA when they communicate with their co-workers. In addition, since male and female GPA speakers work in a segregated society (e.g. in SA), females are expected to spend most of their time in contact with GA family members, as they live with them for no less than two years (see work permit regulation discussed in Section 2.5). This integration in the TL community causes them to have more corrective feedback than male GPA speakers who are relatively socially isolated from the host culture.

Second, pidgin speakers' attitudes and motivations towards learning the TL might vary from those of L2 learners. This situation might also apply to male and female GPA speakers. Attitudes, motivations, and needs are other reasons that women tend to shift to GA more than men. Abed (2017) mentioned that most of his GPA male interviewees lacked motivation and interest to learn Arabic, which is the TL. They believe that learning the basic linguistic level was sufficient to work and live in the country and that there was no need to learn the local variety, as it was not their target. Their motivation for learning how to speak GPA proficiently in their field is to be able to perform their work efficiently. Based on the literature of P/C studies, Field (2004: 135) stated that 'functional proficiency rather than native-like mastery may thus have been the target'. The situation is quite the opposite for female GPA speakers.

Based on the data collected from various interviewees who worked as domestic helpers, such as nannies, cleaners, housemaids, and cooks, the intention to learn the host language was mainly focused on being able to speak Arabic exactly like natives. They believe that speaking like natives will increase job opportunities and facilitate social participation. Moreover, they know that they will be praised by their employers, especially when they communicate with old people, as they know that old people do not alter their language and simplify it for them because the Arabic language is part of their traditions and customs. Therefore, it is not surprising that they were interested in improving their communicative skills and acquiring a high level of the local language, GA. Field (2004) has written that social environments and purposes shape language acquisition. Next, Section 8.4 provides the conclusions of this chapter.

### 8.4. Conclusion to Chapter 8

The data of this study provides valuable insight into the emergence of GPA. Theories explaining the emergence of P/Cs (i.e. substratal influence, the universalist theory and imperfect SLA) have been examined in the three subsections above. Each of these theories can explain specific aspects of GPA and its emergence. For instance, substrate influence visibly appears, as it significantly affected the speech of female GPA speakers in two features (i.e. conjunction and nominal agreement). This substrate influence might have resulted from a comparison of typologically dissimilar substrate languages, as the six languages tested in this study come from three distinct language groups: Malayalam is a Dravidian language; Punjabi, Sinhala, and Bengali are Indo-Aryan languages; and the other language groups, such as Indonesian and Tagalog, are Austronesian languages (see Chapter 4). Therefore, it is not surprising to find substrate influence where several substrate languages are involved. This result is in line with Almoaily's assumption (2012: 181-184) that '[i]f one group had been speakers of a typologically dissimilar language to the Indian languages above such as Tagalog or Indonesian, there might have been more visible effects', and thus, 'more substratal

differences might appear when comparing linguistically dissimilar substrate languages to the Indo-Aryan or the Dravidian languages'.

Likewise, the universalist theory is supported by the data of this study, as demonstrated in Sections 8.3.2 and 8.3.3 (e.g. semantic transparency, simplification, and reduction). This indicates that both factors – substrate and universal – may contribute to the emergence of contact languages. This conclusion is in line with Almoaily's statement (2012: 181) that '[t]his indeed calls for a theory of pidgin and creole genesis which is "tolerant" enough to allow for the possibility of the contribution of both substrate and universal principles in the emergence of pidgins and creoles'.

In fact, the influence of competing theories (substratist, universalist, and imperfect language acquisition) cannot be totally ruled out, as they all seem to contribute to explaining different aspects of the emergence of P/Cs, and they are not separate from one another and can be complementary. Mufwene's (1993) complementary theory of genesis, which states that factors such as the universal and the substratal can contribute to the emergence of contact languages, seems to be an adequate explanation supporting both my findings and those of Almoaily. This theory 'is the only theory that is able to account for the attested evidence that both substratal, universal, and language acquisition factors seem to be involved in the genesis of the contact language under investigation' (Almoaily, 2012: 168).

# Chapter 9: Conclusion

This thesis provides original contributions to and addresses an important gap in a) the knowledge of contact languages based on non-Indo-European languages and b) the methods to compile and analyse a transcribed spoken GPA corpus, which makes it a valuable resource and contribution to both fields of (variational) linguistic inquiries. This chapter summarises the key findings of the research presented in this thesis and its original contributions. Section 9.1 revisits the hypotheses and the main goals of this thesis and discusses how they were achieved. Then, Section 9.2 restates the aims set out in Chapter 1 and summarises the original contributions of this thesis, and sections 9.3 and 9.4 describ the strengths and limitations of the study. Thereafter, Section 9.5 reviews the contents of the thesis chapter by chapter, and finally, Section 9.6 presents suggestions for teaching Arabic as an L2 and possible directions for future work.

## 9.1. Hypotheses and Thesis Goals Revisited

This section revisits the hypotheses and goals set out in Chapter 1 and provides a description of how each goal was achieved.

**Goal 1** was to provide a concise morphosyntactic description for the cross-linguistic comparison of GA, GPA, and the major substrate languages of GPA (Bengali, Punjabi, Malayalam, Sinhala, Tagalog, and Sunda) based on an illustration of the five targeted linguistic features (i.e. definiteness, conjunction, copulas, object and possessive pronouns, and verbal and nominal agreement). This GPA description was based on qualitative descriptive analysis of the interviews. The goal was achieved in Chapter 4 and Chapter 5 by conducting a qualitative

descriptive analysis of the GPA female corpus and reviewing all available descriptive accounts known to the author for the contributing languages in the emergence of GPA. It was a challenging task, as I do not speak any of these substrate languages. For example, the word order is free in Malayalam, as explained by Müller-Gotama (1994), while others state that it is mainly a verb-final language (see Asher and Kumari, 1993). Hence, I had to verify some of this data with Almoaily's content for Bengali, Malayalam, and Punjabi data and with some linguists who speak the rest of the substrate languages[1]. Contrasts between the morphosyntactic systems of GA and GPA were used to test the first research hypothesis of this study, while contrasts between the morphosyntactic systems of the substrate languages were used to support the second hypothesis. Under this goal, the first research hypothesis was tested:

**Hypothesis 1**: There are differences between standard GA and GPA.

This hypothesis was qualitatively tested by conducting a cross-linguistic comparison of the morphosyntax of GPA and GA in an attempt to investigate the differences in the six morphosyntactic variables: (1) presence or absence of the GA definiteness marker; (2) free, bound, or dropped object and possessive pronouns; (3) use of coordinating conjunction or juxtaposition; (4) use or dropping of the copula in the present tense; (5) presence or absence of nominal agreement; and finally, (6) verb dropping, or presence or absence of verbal agreement of GPA, which was the pidgin under investigation in this project, and its lexifier, GA.

The contrasts marked in the substrate languages constitute a significant novel research contribution that is summarised in Table 5.2, which resulted from examining and analysing selected morphosyntactic features of GA, GPA, and the L1 mother tongues of female speakers in Chapters 4 and 5. These contrasts support Hypothesis 1. For example, the use of the copula

[1] Many thanks to Ramdan Sukmawan (Sunda), Rufina Maderas (Tagalog), and Tharindu Ranasinghe (Sinhala).

*fi* is optional in GPA, while in GA it is used overtly only in past and future sentences, but is covert in the present tense.

**Goal 2** was to design and compile my own corpus from transcribed spoken interviews of female GPA speakers, collected to achieve Goals 1, 3, 4, and 5. The goal was achieved completely in Chapter 6 and involved conducting interviews with 72 female GPA speakers who met certain criteria (i.e. had spent either five years or less, or more than 10 years in the Gulf and spoke Bengali, Punjabi, Malayalam, Sinhala, Tagalog, or Sunda as their L1). Compiling the corpus for this investigation was challenging, and when finding GPA speakers who met these requirements, I had to overcome an even larger obstacle: convincing them to participate in the interview. Many simply refused to be interviewed, and many others could not make it because they were too busy. Transcription and extraction of tokens was also not an easy task. It took me three and a half to four hours to transcribe only 10 minutes of speech, as the transcription was done in three stages: listening to the whole interview, transcribing it, and revising my own transcription. Once transcription was complete, I compiled a corpus of 72,000 words out of these interviews (see Figure 6.6 in Chapter 6). To make the tokens easier to access and retrieve, I devised a list of glosses for every variant, for example الفعل الرابط+ (for copula used) and الفعل الرابط- (for copula not used). Refer to Section 4.4.4 for a full list of these glosses.

**Goal 3** was to examine any differences in the use of the five selected linguistic features in the speech of GPA speakers with different L1s (Bengali, Punjabi, Malayalam, Sinhala, Tagalog, and Sunda). The goal was achieved in Chapters 6, 7, and 8 by reviewing and examining the informants' choice between the GPA linguistic variants of the five selected linguistic features in the corpus of GPA. Under this goal, the second hypothesis was tested:

> **Hypothesis 2:** There is a difference between the GPA spoken by speakers with different L1s.

The quantitative analysis of the corpus reflected the effect of the informants' L1 on their GPA output, and a significant relation was found for two features: *conjunction* (p-value = 0.001) and *nominal agreement* (*p*-value = 0.0008). The other features, namely, verbal agreement, copula, and object and possessive pronouns, did not display substratal effects.

**Goal 4** was to examine the half of the data produced by informants who had spent five years or less in the Gulf and the other half from informants who had spent 10 years or more in the Gulf by the time they were interviewed. The goal was achieved in Chapters 6, 7, and 8 by comparing the data of newcomers to the Gulf area (e.g. GPA Tagalog speakers who had spent five years or less at the time of interview) with that of long-term residents in the Gulf (e.g. Tagalog speakers who had spent 10 years or more in the area at the time of interview). Under this goal, the third hypothesis was tested:

> **Hypothesis 3**: Length of stay in the Gulf produces accommodation to standard GA.

The assumption was that if the number of years in the Gulf has a positive impact on female GPA speakers, then they actually shift towards GA after spending more than 10 years in the Gulf. The quantitative analysis of the corpus showed that GPA speakers shift to GA in three linguistic features: definiteness, conjunction, and nominal agreement. However, this shift is significant for just two features: *definiteness* (*p*-value = 0.002) and *conjunction* (*p*-value = 0.001).

**Goal 5** was to examine the effect of length of exposure to GA and whether it is greater in male or female GPA speakers. The goal was achieved in Chapters 6, 7, and 8 by conducting a comparison between my female corpus and Almoaily's male corpus, with a specific focus on the five selected linguistic features. Under this goal, the fourth hypothesis was tested:

> **Hypothesis 4:** Length of stay in the Gulf produces more accommodation to standard GA in women than men.

The assumption was that if the number of years in the Gulf has more of a positive impact on female GPA speakers than on their male counterparts, then gender variation based on the length of stay in the Gulf is achievable by demonstrating that female GPA speakers accommodate standard GA more than men. The analysis of the two corpora revealed that the length of stay in the Gulf resulted in more accommodation to standard GA in women than men for all of the linguistic features except for the use of the copula, which showed no differences. However, this shift was significant for only one feature, namely, conjunction (*p*-value = 0. 0.005). On the basis of this analysis, a list of directions for future work was produced and is presented in Section 9.5 of this chapter.

**Goal 6** was to examine the competing theories (i.e. the substratist and the universalist theories, including the imperfect SLA theory) proposed to account for the genesis of P/C languages based on the evidence found in this study. The goal was achieved partially in Chapter 3 and completed in Chapter 8. These theories can help to explain different aspects of the emergence of GPA. However, the analysis of the data revealed rather complex results. For example, substrate influence had a significant effect on the speech of GPA speakers in two features (i.e. conjunction and nominal agreement). Yet, more evidence can be found in favour of universalist theories of P/C genesis, such as reduplication, reduction, semantic transparency, and simplification. The solution consisted of adopting a theory of P/C genesis which allows for the possibility of the contribution of both substrate and universal theories in the evolution of P/Cs, such as Mufwene's (1993) complementary theory of genesis. In his theory, he claims that universal, in addition to substratal factors, can contribute to explaining the origin of contact languages.

**Goal 7** was to identify any weaknesses and limitations of the methodologies proposed in Chapter 6 and to identify directions for improvement and future research. This goal was achieved in this chapter (Chapter 9) by identifying the limitations and explaining them in detail.

On the basis of this analysis, a list of directions for future work was produced and is also presented in this chapter (Section 9.5).

## 9.2. Original Contributions of the Thesis

Achieving the goals described in Section 9.1 led to several original contributions to the literature of less-described non-Indo-European P/Cs in general and GPA specifically, and several individual contributions offer valuable insights into corpus linguistics methodologies and approaches of GPA. The main contributions are presented in their order of appearance in the chapters of the thesis.

**Contribution 1: New insight into how language and culture have an intertwined relationship in a deeply conservative community such as in Saudi Arabia.**

This examination aimed to answer the question of why GPA women show more accommodation towards standard GA than men in their rate of choices among all six linguistic variables. The analysis of Saudi society focused on one of the most prominent social restriction norms: gender segregation. This constraint – separation between the two genders (male and female) – applies to all men and women who live in this community, including foreigners. Alqasem (2017) suggests that this restriction has a great impact on the context in which the GPA language is acquired when GPA speakers come into contact with one another in a particular community, such as the Saudi one. GPA male speakers mostly deal with men due to the nature of their work. Moreover, due to social distance, when it comes to dealing with woman, these males are less confident and more cautious, and they use only a few words, which they repeat in nearly all cases. In contrast, GPA female speakers, most of whom are domestic helpers (or housemaids), live with the family that employs them, and they are accustomed to daily interactions with the whole family, which makes them feel more confident and less worried. In addition, it enables them to freely communicate with native speakers, rapidly learn the language of the host community, and effortlessly adapt to the system of GA.

**Contribution 2: An academic attempt to offer a more globalised view of contact languages by introducing GPA to the field and exploring it from both sociolinguistic and linguistic points of view.**

This investigation fills a gap in research based on non- Indo-European (Arabic) input, since most previous studies of contact languages have been based on Indo-European languages. The study contributes new knowledge to the literature regarding the two dimensions (linguistic and sociolinguistic) of GPA emergence, and therefore it is of significance. Furthermore, in this thesis, I aimed to contribute to the literature by investigating the extent to which a less-studied Arabic-lexifier pidgin, namely, GPA, complies with the proposed typological features of P/Cs. As demonstrated through data analysis, both competing theories of the genesis of contact languages (i.e. the substratist and the universalist, including the imperfect SLA theory) are supported by the data of this study and can help us explain different aspects of the emergence of GPA.

**Contribution 3: The first annotated corpus of a pidgin- or creole-Arabic-based language built from female speakers.**

This new linguistic resource fills the gap in linguistic resources for an Arabic lexified pidgin and can be used for the development of NLP applications for the GPA domain, for the evolution of modern Arabic, and for linguistic studies focused on non-Indo-European contact language studies. The size of the corpus is 72,000 words from female GPA speakers, drawn from six linguistic backgrounds (Malayalam, Punjabi, Bengali, Tagalog, Sinhala, and Sunda).

**Contribution 4: A concise morphosyntactic description for cross-linguistic comparison of GA, GPA, and the major substrate languages of GPA based on an illustration of the five targeted linguistic features.**

This comparison was made through the description of the above-mentioned languages, GA, and the L1 mother tongues of the female speakers recorded in the corpus, with the aim of testing

the research hypothesis that differences exist between standard GA and GPA. The analysis also employed a discussion of how the similarities and differences between the substrate languages may play a possible role in the variation encountered between the speakers of GPA.

**Contribution 5: The first variationist quantitative investigation of the influence of the selected linguistic features of substrate languages (L1) among female GPA speakers on the production of GPA.**

This analysis aimed to test Hypothesis 2 regarding whether a difference exists between the GPA spoken by speakers with different L1s. The testing was achieved by investigating the participating informants use morphosyntactic features similar to those found in their L1. The results are statistically significant for two features: conjunction markers and nominal agreement. Furthermore, the sampled informants produced linguistic features of contact languages that link to their L1s, which have a significant effect on their production of GPA. Moreover, divergent properties of the universal features of contact languages support the *substrate theory* of genesis. This is a valuable contribution to the field of (socio)-linguistic variation and change in contact languages.

**Contribution 6: The first variationist quantitative investigation of the effect of number of years of stay in the Gulf on female GPA speakers' choices as regards the studied morphosyntactic features.**

This analysis aimed to test Hypothesis 3 that length of stay in the Gulf produces accommodation to standard GA. The testing was done by investigating the participating informants' use of morphosyntactic features that accommodate to standard GA. The results are statistically significant for two features: conjunction markers and definiteness. The results also revealed that the sampled informants produced linguistic features of contact languages which cannot be traced to their L1s, and furthermore that divergent properties of their L1s did not have a significant effect on their production of GPA, thus supporting *universalist theories* of

genesis. This is a valuable contribution to the field of (socio)-linguistic variation and change in contact languages.

**Contribution 7: The first quantitative variationist analysis study of the influence of length of stay in the Gulf on male and female GPA speakers.**

This analysis aimed to test Hypothesis 4 that length of stay in the Gulf produces more accommodation to standard GA in women than men. The testing was done by conducting a comparative quantitative analysis of my female data with Almoaily's (2012) male data. the results indicate that length of stay in the Gulf produces more accommodation to standard GA in women than men. Statistical significance was observed for one feature, namely, conjunction markers. The analysis also revealed an influence of gender-related issues, such as GPA speaker's desire and ability to learn GA and the quality of the input that they receive from locals, and social factors on the learning of the GA language for female GPA speakers in SA. This contribution is important for social and gender variation studies, as well as for linguistic variation and change.

**Contribution 8: A set of suggestions for teaching the Arabic language as a second language.**

The findings obtained from the morphosyntax used by migrant workers and the reasons for this, will help teachers to understand how the language of migrant workers should best be adapted to Standard Arabic. They provide interesting information about the specific areas of difficulty for L2 speakers, enabling teachers to pay more attention to these linguistic phenomena both in their teaching practice and when preparing learning materials for their students, as they will have an enhanced understanding of what their students tend to struggle with. These findings bear important contributions to the educational fields, especially to the corpus-based study of language and teacher education.

### 9.3. Strengths of the Study

The strength of this research lies in its discovery of the potential factors conditioning language and gender variation in GPA morphosyntax and its examination of supporting evidence for the competing theories of P/C genesis, and thus its contribution to the literature of less-described non-Indo-European P/Cs. This section also addresses two additional major strengths: its new perspective and its methodology.

#### 9.3.1. Strengths – New Perspective

As discussed in Chapter 3, GPA has rarely been mentioned in the field of linguistics or been listed as one of the P/C languages around the world even in more recent work, such as McWhorter (2019), Mufwene (2019, 2020), Parkvall (2019), and Velupillai (2015). Despite the major advances over the middle of the last century in the field of P/Cs, more research is still needed, especially on non-European lexifier contact languages, as the number of European-language-based P/Cs is high compared to the number of non-Indo-European P/C languages. Velupillai (2015: 99) stated that:

> While there are a number of contact languages that are not related to any European languages, the vast majority of the languages that we today call pidgins, creoles and mixed languages emerged as a result of the European exploration and exploitation of the world

This limitation in the literature of non-Indo-European-language-based P/Cs is mainly due to the fact that Western scholars are hindered by cultural biases and geographical barriers, as it is too difficult for them to investigate non-European-language-based P/Cs.

Apart from this, most previously studied P/C languages are based on structurally similar languages (i.e. either European superstrate languages or African substrate languages). All the knowledge today about P/C languages – whether theories of P/C genesis or general features of P/Cs – is the result of continued investigation and documentation of European lexifier contact languages such as English, French, Portuguese, Dutch, and Spanish (cf. Holm, 1988, 2000;

Todd, 2003; Arends et al., 1995; and many others, with a few exceptions, such as Bakker, 2003; Versteegh, 2008).

Therefore, conducting more extensive investigations into lesser-described P/C languages (non-European input) and comparing them with those of Indo-European lexifier P/Cs might a) expand the current understanding of the nature of P/Cs, b) aid in the construction of more accurate theories of their geneses, and c) lead to the establishment of more accurate inventories of the typological features of contact languages. Although researchers such as Bakker (2003) and Bakker, Daval-Markussen, and Parkvall (2011) have made considerable efforts to compare Indo-European with non-Indo-European-based P/Cs, more comparative work is required on non-Indo-European-language-based P/Cs, and this is where the importance of this research lies. The new perspective found in this research contributes to the literature on GPA, and it further provides useful insights for researchers in the fields of contact languages and linguistic variation and change in general, and the development of modern Arabic specifically.

The study of linguistic and gender variation in contact languages can make a valuable contribution to the field of sociolinguistic variation and change. Yet, studies on the linguistic variation in non-Indo-European languages, such as Arabic (see Skousen, 1989; Wahba, 1996) and Korean (see Hong, 1991), are limited. In previous research on GPA, such as Smart (1990), Wiswall (2002), Holes (2008), Bakir (2010), Næss (2008), Albakrawi (2012), Salem (2013), Alghamdi (2014), Avram (2014), Al-Ageel (2015), Albaqawi (2016), and Lowi (2017), linguistic variability has hardly ever been investigated, with the exception of a small number of studies, namely, Albaqawi (2019), Almoaily (2012), and Alshammari (2019).

To date, knowledge about linguistically and socially conditioned variability in non-European contact languages (e.g. GPA) is lacking. Researchers have merely described the existence of linguistic variation, but avoided leading quantitative investigations of the components that

condition the heterogeneity (see Samarin, 1986; Mesthrie, 2002; and Avram, 2010). This lack of quantitative investigations on non-Indo-European pidgins calls for further research on linguistic variation in atypical contact languages. Hence, this study offers a new perspective that attempts to provide a quantitative and a qualitative analysis, with the aim of discovering the potential effect of three factors (female speakers' L1 and the number of years spent in the Gulf on language and gender variability in GPA morphosyntax), with a specific focus on the social context of the language (e.g. how and where it is used, and community norms and attitudes towards it). Importantly, examining the speech community in which contact languages emerge leads to an understanding of their evolution, especially as a segregated society such as SA provides the widest social context for language development and gender variation. Such a unique investigative approach to GPA supports recent research such as Ladd, Roberts, and Dediu's (2015) suggestion that linguistic structures adapt to the sociocultural environment in which languages are spoken.

This study is an attempt to introduce GPA to the academic linguistic community with the hope that this research will be a positive contribution to the field of P/C studies. Furthermore, this new insight requires both linguistics and sociolinguistics research to be incorporated into studies of contact languages.

### 9.3.2. Strengths – Methodology

This study combined two phases of analysis, qualitative and quantitative, within a single study in an effort to provide objective and accurate data. These two approaches consequently offer well-established guidelines for reasoning back and forth among the research questions, research design, and research methods. The rationale for this approach is that the qualitative items provide the researcher with interesting quotes that can be used to validate and embellish the quantitative findings (Webb, Sweet, and Pretty, 2002; Gelo, Braakmann, and Benetka, 2008; Creswell and Clark, 2017).

Additionally, the study has provided methodological information about how to build and analyse a transcribed spoken corpus written in Arabic script. The study is also ethnically diverse, given that the female Asian migrant workers who were interviewed were from the different nationalities amongst the six countries with the highest number of migrant labourers in SA. They differed in terms of their demographics (i.e. their regional backgrounds, employment, age, and educational level, as well as the period that they had spent in the Gulf countries). Such diversity in the demographics of the participants means that the collected data is, as much as possible, representative of the whole population living and working in the country, thus broadening its impact and applicability to studies of linguistic trends in the Gulf region.

## 9.4. Limitations of the Study

Although the findings from this study are significant and insightful for the field under investigation in this project, the results must be interpreted with caution, and a number of limitations should be kept in mind. There are two main categories of limitations: those resulting from the methodology and those resulting from issues with the researcher.

### 9.4.1. Possible Methodological Limitations

#### *9.4.1.1. Issues with Data Collection*

In this research study, I relied on gathering the data from two sources: (1) the primary source, namely, GPA female data, which I gathered myself, and (2) the secondary source, namely, GPA male data, which was collected by Almoaily (2012). Each data source has its benefits and limitations, which are discussed in this section.

Bailey, Wikle, and Tillery (1997:57) stated that ‘self-reports might be more valid and reliable measures of linguistic behaviour than linguists have supposed’; this kind of technique in gathering data is limited by the fact that it can rarely be independently verified. Several

potential problems can be associated with self-reports, such as the social desirability response bias, where one must take what people say – whether in interviews, in focus groups, or on questionnaires – at face value. Holtgraves (2004) defines social desirability bias as a response bias that refers to respondents' tendency to present themselves in a way that is untrue and to present a positive impression to the researcher. Social pressure is one of the reasons for the increased reliance on self-reports. Hallgren (2012: 190) mentioned 'social pressures that may promote socially desirable responses that do not reflect actual practices'. Some examples of such social pressures include fear of inadequacy, malpractice litigation, and an overarching fear of losing one's job. Unfortunately, this appeared to be the case with some of the interviewed Asian migrant workers, which had a potential impact on the reliability of the data collected from those participants. Some of them were aware of such pressures, and they thus issued socially desirable answers either by adapting the researcher's speech or pretending not to know how to answer the question in order to avoid the possible punishments associated with providing an accurate account of their linguistic behaviour. Researchers must therefore pay attention to such factors to attain more accurate data from their respondents.

Abernethy (2015) suggested one of the ways to quantify this bias is the *multisource method*, which makes use of information attained through more than one source. He added that the use of this method has been proven effective at controlling for this bias. Moreover, Blackman, Ostrander, and Herman (2005) stated that researchers can use the *multisource method* to increase the accuracy of a study and decrease the rate of false positives. In the case of GPA, researchers will have superior results if they draw upon information attained not only through self-reports but also from peers, friends, or co-workers who speak the same L1 as the individual being interviewed. Huber (1999) and Almoaily (ibid) conducted dyadic interviews with two speakers of the contact variety. Huber (1999) found that this method was helpful to him as a foreigner observing Ghanaian Pidgin English, while Almoaily (2012: 125) found it difficult 'to

determine “who said what” when transcribing the data, especially when the informants participating in one interview are not easily distinguished by their voice (e.g. similar pitch and speed) – this made me decide not to use this method extensively’.

Secondary data offers several advantages as it is time-saving and cost-efficient for the researcher, helps to improve the understanding of the problem, aids in more insightful interpretation of primary data, and provides a basis for comparison of the data that is collected by the researcher. However, there are some disadvantages associated with this. For the GPA male corpus, the sample used to generate the data was small, and it did not cover the parts of the population I wished to examine. For example, my GPA female data was much more ethnically diverse than Almoaily’s male-related data. His sample included only three linguistic backgrounds (i.e. Malayalam, Punjabi, and Bengali), while my sample, in addition to Almoaily’s three, included another three linguistic backgrounds (i.e. Tagalog, Sinhala, and Sunda). As Almoaily’s data was gathered for purposes other than the problem in mind (i.e. gender variability), the total limit of the data restricts the comparison sort to certain linguistic backgrounds.

#### *9.4.1.2. Lack of Previous Research Studies on the Topic*

Despite the increasing labour market demand for female migrants in the Gulf –a demand that is often more stable than that for men – no studies to date have investigated the gender variation in Gulf countries. This lack of evidence regarding gender variation in the GPA-related literature presents a number of challenges in this area of research, particularly in terms of interpreting or testing the accuracy of the GPA data. The difficulty might be even more critical if a given text is the only available one for that particular contact language, as in the case of the GPA text. Although this study compared the GPA production of female speakers with that of male speakers based on the number of years they spent in the Gulf, this study does not provide a complete picture of this variation analysis, as it targeted only five linguistic features (i.e.

definiteness, conjunction, copulas object and possessive pronouns, and verbal and nominal agreement) in GPA. More extensive documentation and analysis of GPA is thus required, particularly with regard to gender variation, which may enhance the generalisability of my findings.

### 9.4.2. Possible Limitations of the Researcher

Researcher-based limitations arise from situations relating to the researcher/s (regardless of whether they are the direct fault of the individual/s). In this study, I acknowledge one researcher limitation: access to data.

#### *9.4.2.1. Limited Access to Data*

Since my research involved interviewing certain people in certain places or organisations, I faced problems of limited access to these respondents. Finding people who were willing to participate in interviews was not as difficult as gaining access to some environments or communities. In some private facilities, obtaining approval from their bosses or employees to take part in the interviews was limited or sometimes even denied. Thus, to save time and efforts, I tried to avoid opting for these places (e.g. addiction and mental health treatment facilities, orphanages). In fact, gaining insider assistants can help by establishing credibility for the study and thereby encouraging honesty and commitment on the part of interviewees (Salmons, 2012).

## 9.5. Review of the Thesis

This section provides a brief review of the first eight chapters of this thesis.

**Chapter 1** presented the context of and motivations for the research presented in this thesis. The main aims of the thesis were introduced, as well as the research hypotheses on which the research is based. A list of goals to be achieved in order to fulfil the aims of the thesis and the contributions which they would generate followed. Finally, the chapter introduced the structure of the thesis and the contents of each chapter.

**Chapter 2** provided an overview of the sociolinguistic and socioeconomic situation of the Arabian Gulf region where GPA is spoken, with particular focus placed on Riyadh City, Saudi Arabia as the main setting of the research study. Then it discussed the most relevant research literature on the influence of gender role and gender segregation in second language acquisition to provide a broader perspective on the conditions and characteristics which affect language use generally and specifically in a non-Western society like Saudi Arabia and its relationship with Asian Migrant Workers

**Chapter 3** discussed some common theories on the genesis of pidgin and creole languages. It then provided a historical perspective on their development in the field of contact languages. It presented general linguistic features of pidgins and creoles and reviewed the literature of GPA and other Arabic-based contact languages. In addition, this chapter discussed the importance of the corpus-based approach in contact language studies with a specific focus on the pros and cons of using corpus evidence in sociolinguistic studies.

**Chapter 4** provided a description of the target features of the languages in contact (i.e. definiteness, conjunction, copulas object and possessive pronouns, and verbal and nominal agreement) in the pidgin under investigation in this project, and its lexifier, GA which I considered in my quantitative investigation.

Chapter 5 described six of the main substrate languages of GPA, namely Malayalam, Punjabi, Bengali, Tagalog, Sinhala, and Sunda based on an illustration of the morphosyntactic target features relevant to this project (i.e. agreement, pronouns, definiteness/ indefiniteness, coordination, and copula).

**Chapter 6** provided a detailed description of the pilot and main stages of the current study. It described the methods of collection and processing of the GPA corpus. Contrasts between the

morphosyntactic systems of the substrate languages were used to formulate the hypothesis based on substrate influence on the linguistic production of informants participating in this study. As for the influence of the length of stay on language and gender variation in GPA, GA tokens produced by long-term speakers were compared with those of the newly-settled GPA speakers, and between those in my female corpus and Almoaily's male corpus, along with some factors which could have had an effect on the informants' choice between the selected features' variants, such as exposure to GA, learners' willingness and abilities to learn a new language, and other social factors such as segregation and social distance between locals and foreign workers.

**Chapter 7** presented the findings of my fieldwork. The analysis applied on the significance of the effect of the informants' L1 and of years of residency in the Gulf on language and gender variation in GPA.

**Chapter 8** discussed the key findings from the GPA speakers' data, besides a discussion of the potential influence of the universal and substratal factors which have led to the emergence of GPA and also discussed the other socio-cultural norms that might have influenced the GPA speakers' acquisition of the GA language.

### 9.6. Suggestions for Teaching Arabic as a Second language

Arabic is a diglossic language, as Al-Sobh, Abu-Melhim, and Bani-Hani (2015) report in the words of Ferguson (1959: 274):

> Diglossia in Arabic refers to the phenomenon of co-existence of two distinct language varieties in the same speech community each of which is used for specific linguistic and communicative purposes by its speakers. In the case of Arabic, the standard variety (classical Arabic) is used in formal speeches, university lectures and news media. In contrast, the colloquial variety is used in everyday speech in informal conversational situations by ordinary educated and uneducated Arabs alike.

It is important that Arabic learners should distinguish between these two varieties of Arabic. I asked some participants if they know that there are sets of varieties: (a) CA, the formal dialect or the Qur'anic language; (b) MSA, the official language in the Arabic-speaking world, used in its oral and written form on all official occasions (one of the main differences between MSA and CA lies in the vocabulary items used in each variety); and (c) Colloquial Arabic, the informal spoken variety of Arabic used by Arabs in their daily lives. The variety is based on regional and geographical variation not only between countries, but also among areas in the same country (Al-Sobh et al., 2015). Indeed, only participants from Muslim backgrounds knew these varieties of Arabic. Some of them said that learning Arabic has been a family tradition and that they were required by their parents to study Arabic, while others said that they are interested in learning Arabic to understand the Quran and to be able to speak with Arabic speakers. However, the majority of them did not know about these varieties; they learn Arabic because they believe it would positively impact their future career and employability in the Arabic-speaking world. This could suggest that Arabic teachers need to engage more with learners' needs, and since all the informants were FWs or guest workers, their optimal need is to improve their work experience.

- My results from analysing selected features of GA, carried out from interviews with 72 female GPA speakers, revealed that the number of years of residency in their location in the Gulf was significantly related to their adaptation to the system of GA (the lexifier language). This suggests that FWs are motivated by length of stay in the region, and this makes them ascend their 'linguistic ladder in the sense of moving closer to Gulf spoken Arabic' (Bakir, 2010: 204). In addition, length of stay in the Gulf produces more accommodation to standard GA in women than men as a result of constant interaction with locals. In this sense, it is necessary to introduce a goal-oriented structure for FWs in SA to be taught and to learn the Arabic language. It is thus also necessary to offer study courses which are in demand and

oriented to the needs of the labour market, as well as to shorten the actual learning periods spent in the country learning Arabic.

Moreover, ROTA (Reach Out To Asia) carried out a survey as part of its Adult Arabic Literacy (RAAL) program, which suggested that 55% of unskilled or semi-skilled labourers did not possess Arabic literacy skills, while 45% of workers displayed, to some extent, a fair degree of spoken Arabic with no reading or writing skills. Therefore, this section expands on a limited part of this research in relation to Arabic teaching and learning, and it focuses on how Arabic is taught. This includes a set of suggestions for teaching Arabic to migrant workers as a foreign language in some Saudi independent companies and language institutes, making this research of vital importance to teachers of Arabic.

- The Saudi government should establish some training courses, such as the Arabic Literacy Program, which is designed to equip volunteer tutors with the necessary skills to effectively facilitate the adult Arabic literacy initiative.
- All young people in SA who are willing to take part in an experiential service-learning opportunity should be encouraged to develop their skills and knowledge as adult Arabic literacy trainers.
- Volunteer tutors should be aware of why expatriates choose to learn Arabic, understand their circumstances and their needs (Jolly and Bolitho, 2011), be equipped with the relevant skills to teach adults, and ensure that foreign adult learners are benefiting from their learning.
- One of the goals of such a program is to promote mutual respect between different cultures and offer people from different countries with diverse cultures a greater understanding of Arab culture and Saudi society.
- It is most important that migrant learners learn the correct variety of Arabic (the day-to-day language and the language Arabic speakers use with their parents, families, and friends) to fulfil their purpose of learning the language to benefit from their learning in practice.

- I strongly recommend that FWs try the learn some basic Arabic prior to arrival for work in Saudi-Arabia or other Arab countries.
- Native Arab speakers should be aware of the language they use when talking to FWs. I believe that if Arab speakers use their native tongue when communicating with foreigners, they will contribute to making GA the lingua franca in the future.

### 9.7. Suggestions for Future Work

This research can serve as a foundation for future research relating to GPA so as to explore various related areas, some of which are described in this section. A list of possible directions for future work relating to GPA is provided and discussed. These directions either emerged during the study or are motivated by weaknesses in the described methodology:

- As explained in Section 1.1, this study is the first attempt to investigate gender variation in GPA based on length of stay in the Gulf. Future research on GPA gender variation is essential, especially after *The Saudi Vision 2030*[2] (this is part of Crown Prince Mohammed bin Salman's programme to modernise some aspects of Saudi society; see Chapter 2 for more information about The Saudi Vision 2030), which aims to increase women's participation in the labour market from 22% to 30%. Now, the rules have been changed, and the new law covers employment regulations. All citizens now have the right to work without facing any discrimination based on gender, disability, or age group. Hospitals, banks, shops, and restaurants are examples of workplaces where women will be allowed to mix with men and work together without separation walls. Due to changes in regulations, Saudi society will witness a new era of rapid changes. Society has already begun to accept young men and women working together, which in turn will be applied to non-Saudis as

[2] A full text of the Saudi Vision 2030 (two versions, Arabic and English) is available at https://vision2030.gov.sa/en
https://www.alarabiya.net/ar/aswaq/economy/2016/04/25/2030-تفاصيل-رؤية-المملكة-العربية-السعودية.html

well. One of the reasons that GPA women show more accommodation towards standard GA then men in their rate of choices among six linguistic variables was that Saudi society is a conservative community controlled by some social norms and restrictions, such as gender segregation, which could play a major role in the process of acquiring a language. In the light of these changes, it will be interesting to see if these changes can be linked to gender variation within GPA. Moreover, gaining access to female workplaces will become much more possible for male researchers, rather than being restricted only to female researchers. Furthermore, since this corpus will be freely available to all researchers who are interested in studying Arabic contact languages, it is possible to track the development of some specific features in GPA, especially after all the changes that the *Saudi Vision 2030* implements in the Saudi society.

- The size of the corpus or sample sizes should be increased (especially when studying less frequently occurring features or features that require a fine-grained subcategorisation, e.g. pronominal forms) – this is to combat data sparsity and to ensure that any conclusions regarding statistical significance are reliable.
- Future work on GPA gender variation based on phonological analysis would be helpful, as it might reveal additional differences between its speakers than the morphosyntactic analysis revealed, especially among typologically dissimilar languages.
- Further diversifying the participants in the investigation is recommended. As seen in Section 6.3.3.3, all the participants in the study were from the Asian continent, and more substratal differences might appear when comparing linguistically dissimilar substrate languages to South or East African expats' languages, such as languages of Kenya, Ghana, Tanzania, or Uganda.
- Future analysis of possible sets of lexical features of GPA is advised. Examining this set of features is needed to examine supporting evidence for the competing theories of P/C genesis.

### 9.8. Thesis Final Remarks

This research has resulted in the creation of a new linguistically annotated dataset (corpus) for a severely under-resourced and under-documented language variety, and it could potentially be used by other researchers in the future to foster other types of linguistic analyses, leading to academic benefits (this annotated corpus of female speech will be made freely available to all researchers who are interested in studying Arabic contact languages). This thesis makes several significant contributions to the study of less-described contact languages in the literature of non-Indo-European P/C languages. These contributions include the first quantitative study of female speech in GPA, which is complementary to Almoaily's study of male speech in the GPA spoken in Gulf countries. This is an original contribution to studies on the relationship between gender and language, presenting the first female corpus of GPA, as well as the use of software such as AntConc to quantitatively analyse corpora. It also provides supporting evidence for the competing theories of P/C genesis, along with interesting sociolinguistics findings. With these results, methodologies, and resources, the thesis aims to provide useful insights for researchers interested in gender and language variation and change in general, and the evolution of modern Arabic specifically.

## Appendix A. Interview Schedule

**Questions**

**I. Interviewee demography:**

Questions for part I of the interview are formulated to determine the participants' demographic backgrounds.

1. Where are you from?
2. From which city/town are you?
3. What is your job in Saudi Arabia?
4. Please can you briefly outline your job role and responsibilities?
5. How long have you been in your current job?
6. Have you had any training course for your current job? If yes, was it in your home country or in Saudi Arabia?
7. In your view, how effective was the training course for your job?
8. Do you recommend your colleagues to be trained before coming to Saudi Arabia? (Almoaily, 2012)
9. Did you go to school? If yes, till which level?
10. Are you married? If yes, how many kids do you have? Are they with you? Or have you left them in your home country? With whom?
11. How old are you?
12. Do you have any relatives living in Saudi Arabia? Do you meet them? Where?
13. For how many years have you been living in KSA, the Gulf? (Almoaily, 2012)
14. Did you work/ live in any other Arabic speaking country before you come to Saudi Arabia? Where?
15. Do you contact your family regularly? How often? by phone? Post? Internet? Other?

**II. Linguistic background:**

Questions for part II of the interview are formulated to determine the participants' Linguistic backgrounds.

16. What is your first language?
17. Do you have a variety of this language? Do you speak it?
18. Do you speak Arabic frequently, how often? With whom you speak it and where? Do you speak it in your home country also?
19. Do you understand your Madam language quickly when she asks you to do something? If not, why?
20. Did you find Arabic difficult to use? Why? Do you think you still need to learn it more or what you have is enough for you?
21. Did you take Arabic courses before coming to Saudi Arabia? If yes, were these courses in Classical or in Modern Arabic? Did you find them helpful? If no, are there any? Why did not you consider taking one?
22. Are there any linguistic similarities between your first language and Arabic? How?

23. Do you watch Arabic TV channels? How many hours a day/ a week? What kinds of programs do you watch? Why? Can you fully understand their language? Do you feel that they may help you in learning Arabic?
24. Do you have TV channels in your first language? or any other languages rather than Arabic? Do you watch them here in Saudi Arabia?
25. Do you listen to radio? If yes, in which language?

**III. Other:**

Questions for part III of the interview are formulated to stimulate informants to produce long turns in the interviews.

26. Do you live in *kafeel*'s[3] house? If not, where?
27. Being working and living in a foreign country, do you feel that you are alone and need your family? (Almoaily, 2012)

28. How did you find Saudi Arabia, can you tell me things you like and things you don't like?
29. What do you do when you are free?
30. Do you have traditional foods in your home country? How do you prepare them?
31. Can you cook the same kinds of food here?
32. Can you tell about your experience of working abroad?
33. When you were a child, did you play games? What kinds of games?
34. Do you have any excitement story from your life? Can you remember it?
35. Have you ever been in a situation where you thought that you are about to die?
36. Is there anything here that makes you worried? Why?
37. Have you experienced traumatic life-threatening events?
38. Do you have future plans? If yes, what are they? Do you have any concerns about them? How can you achieve them?
39. Have you heard that SA lifted its ban on female drivers? If yes, how did you know that?
40. How do you feel about allowing women to drive? Are you happy or worried, and why?
41. Do you plan to drive here in SA, or do you want to work as a driver?

**IV. Direct elicitation (PowerPoint Presentation)**

These kinds of questions in the PowerPoint presentation are prepared to ask informants to reflect on objects they see in the presentation. They are designed to elicit tokens of linguistic phenomena such as prepositions and gender and number distinctions in demonstrative pronouns (Almoaily, 2012).

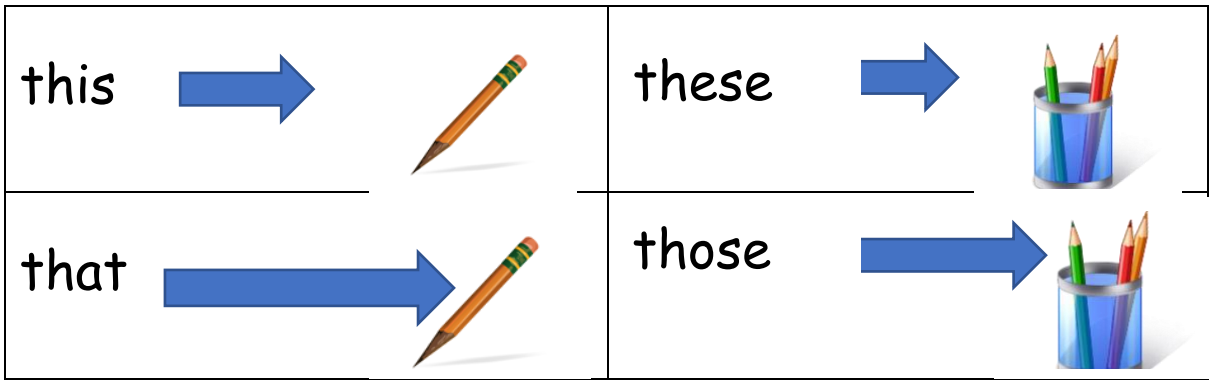
**A. Activity 1**

In this activity[4], subjects will be asked to name objects of different quantities and genders located at various distances to be described using a demonstrative pronoun.

[3] *Kafeel(kafil)* is the immigrant's sponsor in Saudi Arabia. He/she can be individual or a company.
[4] Retrieved 16 Jan 2017 from: https://en.islcollective.com/resources/printables/worksheets_doc_docx/this_that_these_those/demonstratives-classroom-this/45044

The purpose of this task is to check the use of GA demonstrative pronouns by GPA speakers. They will be asked to refer to the object they see using a full sentence.

1) Complete the sentences with this, that, these or those.
Below is an illustration of this activity:




1) ___ _____ is a computer game.

2) _____ are lorries.

3) _____ is a camera.

4) _____ is a kite.

5) _____ are balls.

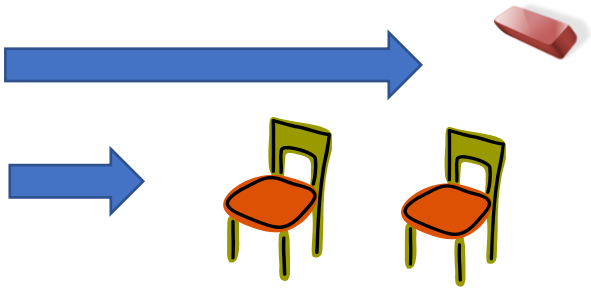
6) _____ are watches.

7) _____ is a _____

_____

8) _____ is a _____

9) _____ are _____



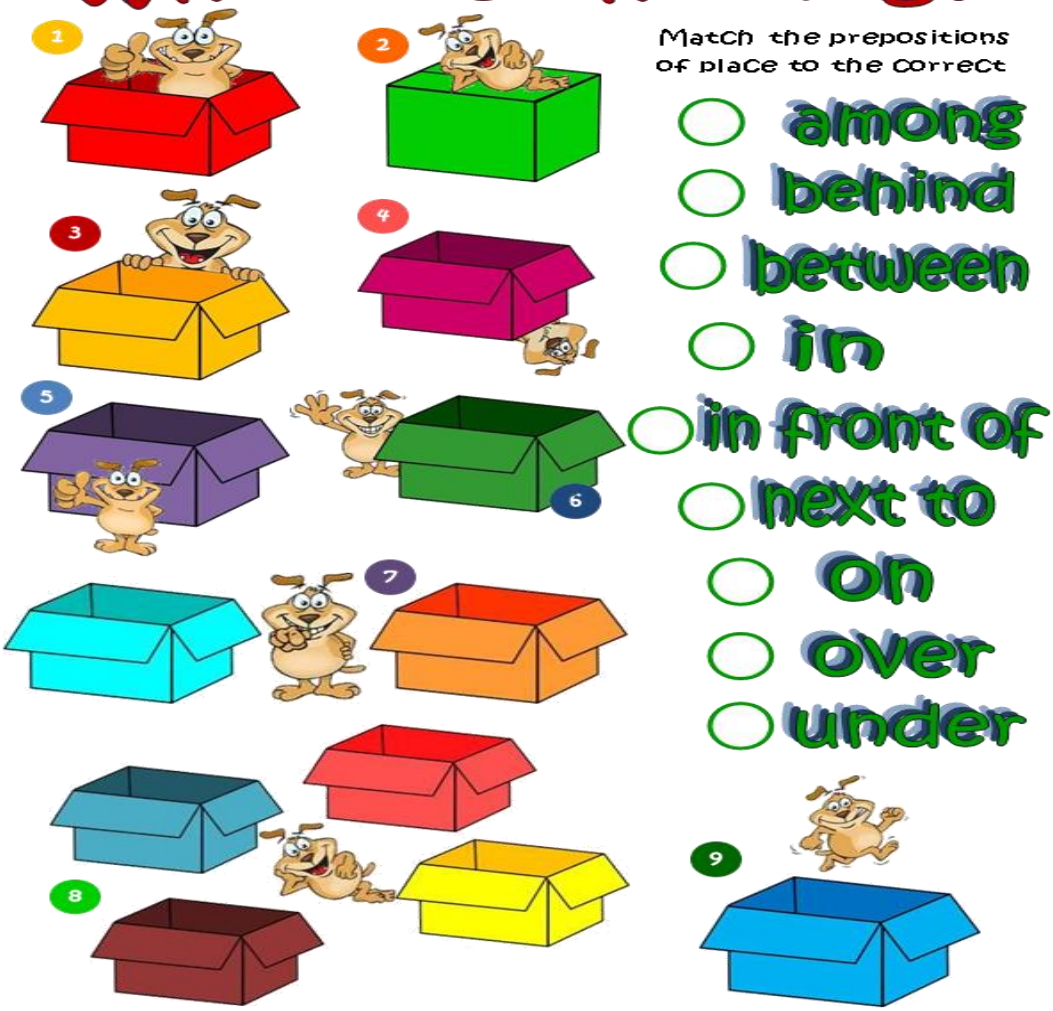
**A. Activity 2**

In this activity[5], informants will be asked to mention the location of a subject which is positioned in various places in each slide. The purpose of this task is to investigate the use of prepositions by GPA speakers.

2) Where is the dog located in each of the following pictures?

[5] Retrieved 16 Jan 2017 from: https://en.islcollective.com/resources/printables/worksheets_doc_docx/this_that_these_those/demonstratives-classroom-this/45044

# Where is the dog?




Match the prepositions of place to the correct

- among
- behind
- between
- in
- in front of
- next to
- on
- over
- under

## Appendix B. Interviewee Metadata in GPA corpus

| Interviewee Code | L1 | L12 | Education | Age | Years in SA | Length of interview | Place of interview |
|---|---|---|---|---|---|---|---|
| T2 | Tagalog | English | College | 27 | 2 | 25 m 48s | Hospital |
| P2 | Tagalog | English | College | 25 | 1.2 | 19m 7s | Hospital |
| T3 | Tagalog | English | College | 22 | 1.5 | 18m 36s | Hospital |
| T4 | Tagalog | English | College | 27 | 2.8 | 27 m 46s | Hospital |
| T5 | Tagalog | English Filipino | Secondary | 23 | 3 | 20 m 25s | Housemaid |
| T6 | Tagalog | English | College | 28 | 1 | 26 m 46s | Hospital |
| T7 | Tagalog | English | College | 39 | 12 | 16m | Hospital |
| T8 | Tagalog | English | College | 31 | 10 | 22M 41s | Hospital |
| T9 | Tagalog | English | Primary | 47 | 14 | 23m 20s | Hospital |
| T20 | Tagalog | English Filipino | College | 40 | 15 | 20m 52s | Hospital |
| T21 | Tagalog | English | Secondary | 52 | 20 | 22m 27s | Beauty centre |
| T22 | Tagalog | English Filipino | Secondary | 46 | 17 | 20m 56s | Beauty centre |
| P1 | Punjabi | Urdu | College | 41 | 4 | 22 m 18s | Hospital |
| P2 | Punjabi | Urdu | College | 40 | 5 | 20 m 47s | Hospital |
| P3 | Punjabi | Urdu English | Primary | 34 | 3.6 | 22m 42s | Housemaid |
| P4 | Punjabi | Urdu | Secondary | 28 | 1.9 | 22m 59s | Housemaid |
| P5 | Punjabi | Urdu | Secondary | 33 | 2 | 18m 58s | Housemaid |
| P6 | Punjabi | Urdu | Primary | 22 | 1.2 | 20m 51s | Beauty centre |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| P7 | Punjabi | Urdu | College | 35 | 11 | 20m 52s | Hospital |
| P8 | Punjabi | Urdu | Primary | 40 | 15 | 20m 52s | Beauty centre |
| P9 | Punjabi | Urdu | None | 49 | 18 | 22m 27s | Housemaid |
| P10 | Punjabi | Urdu | Primary | 55 | 25 | 19m 33s | Housemaid |
| P11 | Punjabi | Urdu | College | 43 | 17 | 20m 17s | Hospital |
| P12 | Punjabi | Urdu | Primary | 43 | 14 | 18m 58s | Housemaid |
| S1 | Sinhala | English Tamil | Secondary | 33 | 1.4 | 24 m | Housemaid |
| S1 | Sinhala | Tamil | Secondary | 37 | 2.5 | 12 m | Housemaid |
| S3 | Sinhala | Tamil Urdu | Primary | 42 | 4 | 22m 44s | Housemaid |
| S4 | Sinhala | Tamil English | Primary | 29 | 2.7 | 23 m 58s | Hospital Cleaner |
| S5 | Sinhala | Tamil | Primary | 33 | 3 | 22m 25s | Housemaid |
| S6 | Sinhala | Urdu Tamil | Secondary | 43 | 5 | 20m 55s | Housemaid |
| S7 | Sinhala | Tamil | Primary | 61 | 27 | 24m 33s | Housemaid |
| S8 | Sinhala | Tamil | Primary | 50 | 20 | 23m 20s | Housemaid |
| S9 | Sinhala | Tamil | Primary | 45 | 14 | 22m 36s | Housemaid |
| S10 | Sinhala | Tamil Urdu | Primary | 49 | 12 | 19m 55s | Hospital Cleaner |
| S11 | Sinhala | Tamil | Primary | 46 | 11 | 20m 17s | Housemaid |
| S12 | Malayalam | Urdu | Primary | 52 | 17 | 23m 17s | Housemaid |
| M1 | Malayalam | Urdu Tamil | Secondary | 35 | 2.6 | 27 m 55s | Housemaid |
| M2 | Malayalam | Urdu Tamil | College | 37 | 1.9 | 19m 47s | Hospital |
| M3 | Malayalam | Urdu | Secondary | 32 | 3 | 23 m | Housemaid |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| M4 | Malayalam | Urdu | Primary | 42 | 3.4 | 19m 44s | Hospital Cleaner |
| M5 | Malayalam | Urdu English | Primary | 44 | 4 | 21m 42s | Hospital Cleaner |
| M6 | Malayalam | Urdu | Secondary | 41 | 4.8 | 18m 55s | Housemaid |
| M7 | Malayalam | Urdu | Primary | 52 | 22 | 25m 12s | Housemaid |
| M8 | Malayalam | Urdu | Primary | 38 | 11 | 20m 11s | Housemaid |
| M9 | Malayalam | Urdu Hindi | College | 57 | 18 | 26m 22s | Hospital |
| M10 | Malayalam | Urdu | Primary | 46 | 12 | 23m 34s | Housemaid |
| M11 | Malayalam | Urdu | Primary | 62 | 24 | 27m 12s | Housemaid |
| M12 | Malayalam | Tamil Urdu | Primary | 56 | 16 | 24m 18s | Hospital |
| B1 | Bengali | None | Secondary | 26 | 3.5 | 27 m 22s | Housemaid |
| B2 | Bengali | Urdu | Secondary | 28 | 2 | 21m 42s | Housemaid |
| B3 | Bengali | None | Secondary | 35 | 4 | 28 m 44s | Housemaid |
| B4 | Bengali | Urdu | Secondary | 23 | 2.8 | 17m 54s | Hospital Cleaner |
| B5 | Bengali | None | College | 32 | 3 | 22m 20s | Hospital |
| B6 | Bengali | None | Primary | 35 | 3.8 | 23 m 40s | Hospital Cleaner |
| B7 | Bengali | None | Primary | 39 | 11 | 20m 54s | Hospital Cleaner |
| B8 | Bengali | Urdu | Secondary | 41 | 14 | 22m | Housemaid |
| B9 | Bengali | Urdu | Primary | 38 | 10 | 19m 57s | Housemaid |
| B10 | Bengali | None | Primary | 57 | 17 | 23m 44s | Housemaid |
| B11 | Bengali | None | Primary | 41 | 13 | 22m 50s | Housemaid |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| B12 | Bengali | None | College | 45 | 15 | 22m | Hospital |
| SU1 | Sunda | None | Secondary | 32 | 2 | 19m 57s | Housemaid |
| SU2 | Sunda | None | Secondary | 28 | 1.5 | 26 m 28s | Housemaid |
| SU3 | Sunda | Malay | Secondary | 35 | 2 | 22m 31s | Housemaid |
| SU4 | Sunda | None | Primary | 37 | 3.4 | 27 m 55s | Housemaid |
| SU5 | Sunda | None | Primary | 32 | 1.2 | 22 m 27s | Beauty centre |
| SU6 | Sunda | None | Secondary | 41 | 3 | 28 m 24s | Housemaid |
| SU7 | Sunda | None | Secondary | 47 | 12 | 22m 17s | Housemaid |
| SU8 | Sunda | None | primary | 53 | 19 | 23m 33s | Beauty centre |
| SU9 | Sunda | None | primary | 42 | 11 | 22m 32s | Housemaid |
| SU10 | Sunda | Malay | primary | 48 | 17 | 20m 38s | Housemaid |
| SU11 | Sunda | None | primary | 52 | 24 | 25m 11s | Housemaid |
| SU12 | Sunda | None | primary | 51 | 20 | 22m 54s | Beauty centre |

## Appendix C. Transcription of Pilot Interviews

Transcription codes:

| | |
|---|---|
| **Lengthening** | **:** |
| **Non-Arabic words** | () |
| **Pause** | - |
| **Raised intonation** | ? |
| **Falling intonation** | . |
| **Laughter** | @ |

**Interviewee: M1**

*Malayalam newly-settled*

Najah: **اهلين.. انتي قلتي انتي من وين؟**

M1: **انا هندي**

Najah: **ايش اللغه الاول؟**

M1: **Malalayumماالاليوم**

Najah: **ايش لغه ثاني .. لغه مدرسه؟**

M1: **اوردو**

Najah: **انتي في الهند من وين؟**

M1: **كيرلا**

Najah: **انتي ايش شغل في السعوديه؟**

M1: **انا هنا كوافيره**

Najah: **كيف كوافيره تكلمي؟**

M1: **يعني كويس كوافيره..سسوار فيه كويس .. انا قبل مافيه معلوم بس الحين ماشاءالله اربعه سنه هنا اجلس هذا ماشاءالله كل شي معلوم.. عشان قبل انا فيه نقاشه هنا .. بس شويه شويه .. انا كلو شوف شوف شوف انا ماشاءالله كلو معلوم.. تسريحه..شسوار .. شمع .. يعني ايش شغل كلو معلوم الحمدلله**

Najah: **حلو ماشاءالله .. انا اسالك عشان انا يبغى كلام انتي كثير عشان انا يبغى يسمع عربي .. مو لازم كلام صح خطاء مافيه مشكله ... طيب كم سنه انتي في السعوديه؟ واول مايجي من الهند على طول سعودية ولا روح مكان ثاني مثل دبي.. كويت**

M1: **دبي انا سته شهر بس .. بس ما فيه معلوم .. هنا معلوم كلو .. انا سوا سوا عربي .. مو واجد كلام .. هندي كلامك عشان فيه بالبيت هندي معلوم .. سته شهر بس اجلس الحين . يجي مره ثاني ..دمام قبل سته شهر .. بس الحين فيه هنا على طول يجلس اربعه سنه**

Najah: **طيب قبل يجي سعوديه درستي كورس عربي؟**

M1: **لا.. لا.. مافيه ابدا ولا شي مافيه معلوم انا .. انا دبي كمان موجود مافيه معلوم سته شهر عاشن انا كلو هندي انا سوا سوا كلام .. مافيه معلوم .. هنا سته شهر دمام كمان موجود .. مافيه معلوم ولا شي .. بس شويه .. واحد واحد ممكن حرف معلوم بس مافيه معلوم على طول يعني .. كذا سته شهر**

Najah: **طيب انتي فكر لو يجي صديق انتي جديد من الهند يبي سعوديه . . انتي لازم قول حق صديق قبل يجي هنا سعوديه هو لازم يدرس عربي قبل في الهند ولا عادي ممكن يجي هنا ويتعلم العربي بسرعه يعني مافيه مشكله**

M1: **اه.. لازم در كويس يعني عربي .. عشان الحين هذا كمان واحد واحد مدرسه فيه واحد فيه صديق فيه موجود واحد عربي الحين .. شويه شويه من بيبي صغير .. الحين معلوم يعني بيبي شويه هناك كمان معلوم**

Najah: **لما كنتي بالهند ايش مدرسه خلاص؟ابتدائي ولا ثانوي**

M1: **انا.. مدرسه يعني انا فيه عشره سنه انا مدرسه يعني 10 class مدرسه .. اوردو بعدين عربي كلاس عربي .. يعني اقراء قران**

Najah: **طيب في الهند خلاص college ؟**

M1: **لا .. مافيه college بس عشره school**

Najah: **انتي متزوجه؟**

M1: **ايوه جوزه انا**

Najah: **وفيه اطفال ؟ كم؟**

M1: **ايوه .. فيه . بيبي .. انا فيه جده الحين**

Najah: **you look too youngماشاءالله**

M1: **ايه انا فيه جده الحين ..بننتي بنتي .. انا الحمدلله**

Najah: **كم عمر بنتك؟**

M1: **بنتك ممكن الحين ثلاثه عشرين سنه**

Najah: **وانتي كم عمرك؟**

M1: **انا ممكن سته واربعين .. لا سبعه واربعين ممكن**

Najah: **فيه اولاد ؟**

M1: **اولاد.. ولد فيه واحده.. خمسه وعشرين .. ايوه**

Najah: ماشاءالله .. هذا اولاد موجود هنا في السعوديه؟
M1: لا.. مافيه هنا كلو ف الهند
Najah: طيب هنا في صديق في السعوديه؟
M1: مافيه.. مو مره .. بس كل واحد زي كذا .. صديقه كلو انا
Najah: بس انتي مايروح سوي زياره حق صديق
M1: لا ...لا مو مره فيه بنقالي فيه موجود .. بس هنا قريب . فيه انا فيه زبونه ... بنقالي فيه معلوم شويه معلوم هندي
Najah: طيب انتي كيف يتكلم مع عائله؟ كيف يتواصل؟
M1: هندي بس هندي سوا سوا كلام في البيت
Najah: كيف تتكلمي مع بنتك ؟ يستخدم جوال؟
M1: Emo كلام انترنت
Najah: طيب لما يطلع برا مشغل ممكن كلام مع ناس عربي عادي ؟ ولا بس صديق هندي؟
M1: لا.. لا.ز فيه ممكن واحد زي كذا يطلع برا ممكن قريب ممكن روح مول سوق مول ممكن فيه عربي موجود فيه شويه شويه عربي معلوم هذا زي كذا كلام معلوم .. وبعد هندي ممكن هندي سوا سوا كلام كذا
Najah: صحيح.. طيب الحين مالايالم لغه فيه شي زي عربي ولا مافيه ؟ فيه كلمات زي عربي ولا مختلف؟
M1: مختلف . يعني هذا مختلف عربي
Najah: طيب انتي يفهم عربي لما يجي زبونه شنو يبي مافيه مشكله
M1: لا.. لا.. فيه على طول زبون يجي على طول لازم يشتغل
Najah: يفهم كلام كلو زبونه
M1: ايه.. واحد واحد زبون مافيه يفهم انا ايش كلام عشان هذا لغه شويه difficult مافيه لغه شويه كلام .. عربي مايفهم ممكن . هي يبي يسوي شعر ولا شي معلوم انا يعني
Najah: يعني انتي الحين شوف مو كلو عربي زي بعض
M1: لا مو زي بعض ايه
Najah: ليش كلو مو زي بعض؟
M1: ممكن عشري ولا ثلاثين different لغه .. هذا فيه لغه زي كذا هند كمان different different هذا لغه هذا.. مو هند بس واحد لغه بس different لغه هذا فيه كلام
Najah: صحيح..انتي الحين فكر الي يجي من دمام مثلا هذا لغه والي يجي من رياض لغه ثاني والي من جده بد هذا لغه ثاني
M1: ايوه ممكن ...انا روح دمام قبل لغه هذا .. لغه هذا شويه غير .. انا قبل في دبي .. هنا كلام "رجال" " ريال " هنا كلام " ميه" هناك كلام " مويه" هنا كلام " مويه " ولا "ميه" زي كذا different انا فكر ليش كذا كلام ... لاحد غير غير لغه
Najah: صحيح ... هذا كلو عربي ب اختلاف شويه
M1: ايوه.. مختلف
Najah: طيب .. انتي فكر الحين انا خلاص فيه معلوم عربي مره كويس خلاص مو لازم اتعلم ولا ادرس عربي كذا انا كويس
M1: لا.. لا.. لازم يعني فيه يبغى كويس زياده يعني كويس..يعني انا ادرس معلوم .. يعني انا سوي درس معلوم.. يعني كويس كلام .. يعني انا يبغى زياده عشان مره كويس يجي
Najah: انتي يتفرج علىtv ؟
M1: لا.. لا.. مافيه.. هنا مافيه
Najah: حتى لو يشوف مسلسل عربي ايش يفهم كلام هذا شنو
M1: لا.. لا.. مافيه فهم .. مافهم ..مو مره فهم .. يعني فيه تركي .. هذا انقلش =
Najah: لا.. لا انا اتكلم عن عربي
M1: عربي مافيه معلوم .. انا مافيه هنا...هنا فيه يعني مسلسل شوف مافيه يفهم .. يعني ب كلام لغه سوا سوا هذا شويه شويه فهم بس مسلسل مايفهم انا.. فلم شوف مافيه فهم ولا news فهم
Najah: يعني هذا فلم يتكلم عربي مره كويس انتي مافيه فهم لازم كلام بسيط عشان انتي يفهم
M1: زي كذا انتي سوا سوا كلام... انا فهم ممكن .. مسلسل شوف مافيه فهم
Najah: طيب حتى مسيقى
M1: ميوزك كمان مافهم
Najah: صعب؟
M1: لا.. مافيه افهم

Najah: طيب انا لما اتكلم عربي لغه فصحى مثل ما اسمك ؟ اين ذهبتي اليوم؟ ماذا تاكلين؟ هذا لغه عربيه فصحى لغه كويس ميه ميه نفس لغه حق تلفزيون اخبار.. اذا احد فييه كلام زي كذا انتي تعرفين. تفهمين؟

M1: لا.. لا... بس اذا زي كذا

Najah: يعني لغه زي كذا بسيط

M1: ايوه

Najah: طيب انتي وين ساكن الحين؟

M1: ساكن هنا فيه قريب المونسيه

Najah: شركه يعني؟ ولا سكن مع اصدقاء ؟

M1: لا كلو سوا سوا بنات مع بعض

Najah: طيب انتي كلام ممكن خمسه ولا اربعه سنه في السعوديه .. انتي يفكر دايما باهل زوج ولا اولاد؟

M1: ايوه.ز ايوه .. بس فكر بس كلام كل يوم كلام بس كل يوم انا شوف كلام بس مافيه فكر زياده يعني مافيه كلام اثنين يوم ثلاثه لازم فكر.. عشان انترنت الحين موجود يعني مافيه فكر هذا بعيد .. يعني على طول كلام .. كلام فكر هذا قريب كذا

Najah: طيب ايش انتي حبي هنا في سعوديه وايش مافيه حبي؟

M1: لا.. حب سعوديه ماشاءالله .. كلو زين .. لا .. لا مافيه مشكله .. انا يجي هنا ليش .. لازم شغل انا حب .. لازم حب..انا حب هنا في مكان .. لازم انا فيه شغل ...انا مافيه حب كيف انا شغل هنا.. صح ولا

Najah: صح.. طيب كيف عن حر هنا

M1: ايه حر واجد .. حر مره واجد... برد مره واجد... مشكله @ يعني@ كذا يعني .. مكان .. بس جو مافيه .. هذا فيه مره different هذا فيه هند مافيه زي كذا . هنا فيه مشكله .. بس الحمدلله اربعه سنه لازم سوا سوا

Najah: طيب فيه وقت فاضي انتي ايش يسوي

M1: على طول جوال @ على طول فيه شوف فلم هندي ... ولا شوف سوا سوا كلام .. فيه بيت .. يعني زي كذا

Najah: تحبين انتي اكل سعودي ؟

M1: مو مره .. ب انا على طول طبخ هندي @عشان هندي الاكل حار .. صح.. زي كذا يعني@

Najah: يعني ممكن اكل هندي موجود في سوق ؟ ولا مافيه موجود

M1: موجود هنا ..بس كلو فلوس غالي مشكله انا..مايقدر يشتري دايما انا اكل ..طبخ يعني .. ماقدر برا انا يشتري يعني

Najah: وشو اكل سعودي انتي حبي مثل ايش

M1: هنا في سعودي بس هذا .. ايش كلام هذا..

Najah: كبسه

M1: ايوه.. كبسه يعني زياده بس كبسه

Najah: تعرفين تسوين ؟

M1: مافيه تععرفين بس سوي كبسه يعني كان ممكن انا فيه واحد صديق هنا .. قريب .. بس كبسه كذا

Najah: طيب انتي ايش فكر يجلس هنا في السعوديه ولا روح

M1: فيه شغل لازم يجلس..فيه يبغى خلاص سنتين لازم روح كمان في البيت لازم .. في البيت كمان شوف بيبي كل شي موجود الحمدلله

Najah: فيه انتي خوف هنا ؟ خوف سوي حادث .. خوف موت

M1: لا..لا..لا..لا الحمدلله مافيه خوف...عشان انا فيه هند كلو يطلع برا ... هناك كثير زحمه .. انا مافيه خوف هناك .. هنا مافيه ولا شي .. بس هنا سياره .. بس مشكله هناك هند trainوbus كل شي موجود .. هناك على طول accident ... هنا لا... الحمدلله ..بس هنا طريق يعني سياره بس كثير

Najah: طيب مستقبل فكر انتي شنو سوي اذا روح للهند؟

M1: انا فيه مستقبل حق بيبي انا @ الحين اولاد بس .. انا مستقبل خلاص .. الحين بيبي اولاد كلو فيه فكر انا مستقبل حق كويس.. زي كذا فكر

Najah: طيب الحين انا ابي اوريك صور وانتي قولي ايش هذا صوره مين موجود في الصوره سمي شكل موجود داخل صوره .. هذا ايش وهذا ايش وكيف نستخدم administrative pronouns اسماء الاشاره بالعربي.. يعني I want you to name the objects that you see in the pictures وكيف قول هذا شي اذا كان بعيد او قريب.. اوك بس شوف صوره وتكلمي

M1: 1- هذه بنت
2- هذا سياره اثنين
3- هذا كاميرا
4- هذا ولد
5- هذا فيه كوره

**6- هذا ولد ثلاثه**
**7- هذا قلم**
**8- هذا ساعه**
**9- هذا اثنين عجوزه**
**10-هذه بنت**
**11- هذا طباخ**
**12- هذا ولد**
**13- هذا اثنين في بيت**
**14- هذه بنت**
**15-هذه عجوزه اثنين**

Najah: **طيب الحين انتي شوفي هذا صوره وقولي وين كلب موجود داخل او برا صندوق او وين .. هذا اسمه كلب . وهذا اسمه صندوق.. انتي كلام وين موجود كلب وكيف نستخدم** propositions **حروف الجر بالعربي**

M1: ok

**1-هذا كلب جوا صندوق..داخل**
**2-هذا كلب فوق صندوق**
**3- هذا صندوق برا ..هذا كلب**
**4-هذا صندوق حق تحت ..كلب ..تحت صندوق كلبه**
**5- هذا صندوق برا كلبه .. كلب برا صندوق**
**6- هذا على جنب كلبه .. على جنب صندوق**
**7- هذا صندوق فيه وسط واحد كلبه**
**8- هذا اربعه صندوق في الوسط كلبه**
**9- هذا صندوق فوق هذا واحد كلبه**

Najah: **طيب شكرا شكر** thank you so much for your help and cooperation **ومع السلامه**

**Interviewee: M2**

*Malayalam Long-term resident*

**Najah:** السلام عليكم...نتي من وين؟ انا من سعوديه انتي من وين ؟
**M2:** في الهند
**Najah:** من وين من الهند؟
**M2:** كيرلا
**Najah:** ايش هذا لغه في الهند ؟
**M2:** Malalayumمالاليوم
**Najah:** هذا لغه اول .. ولما كبير شويه كلام ايش .. ايش لغه ثاني؟
**M2:** هندي
**Najah:** انتي كم سنه هنا في السعوديه؟ او في الخليج مثل كويت ولا دبي
**M2:** هنا في Saudi Arabia خمسه وعشرين سنه
**Najah:** ماشاءالله .. طيب فيه شغل قبل يجي سعوديه مث شغل كويت ولا بحرين
**M2:** لا.. لا.. اول مره في سعوديه بس
**Najah:** طيب شغلك في السعوديه ايش؟
**M2:** اه .. ممرضه عادي .. الحين فيه head nurse
**Najah:** اوكي.. تكلمي اكثر عن الشغل كيف يسوي
**M2:** اول انا شغل في العياده(name of the clinic ) لين ستعشر سنه في العياده مع الدكتور ( name of the doctor) بعدين يجي مستشفى هنا .. فيه الحين ..تم تسعه سنه مستشفى (name of the hospital )
**Najah:** طيب كم الساعه يجي الصباح ؟ يروح الساعه كم؟
**M2:** اه .. يعني فيه اجي عشان شغل ساعه سبعه صباح لين ساعه ثلاثه عصر.. لكن انا مافيه روح كل يوم ثلاثه .. فيه روح اربعه ممكن خمسه .. كذا
**Najah:** طيب لك كم سنه وانتي يشتغل نفس هذا شغل؟
**M2:** هذا مستشفى تسعه سنه
**Najah:** طيب انتي لما جيتي هنا اول مره اخذتي training course بالعربي قبل يجي سعوديه؟
**M2:** لا .. مافيه على طول =
**Najah:** يعني اول مره يجي قبل خمسه وعشرين سنه كيف كانت اللغه العربيه؟
**M2:** لا.. مافيه مره مره يعرف .. لين اسمع بعدين اكتب
**Najah:** طيب لما صديق انتي يبي يجي السعوديه .. انتي تقولين لها احسن خوذي كورس بالعربي ولا عادي ممكن تجي وتتعلم العربي هنا بسرعه؟
**M2:** فيه training course احسن لكن فيه يجي.. يعني فيه ناس فيه سرعه learn فيه ناس شويه فيه training course احسن قبل يجي
**Najah:** درستي اخر مستوى ايش؟
**M2:** Diploma nursing
**Najah:** هذا بعد high school صح؟
**M2:** لا high school فيه college بعدين فيه Diploma nursing
**Najah:** طيب كم عمرك؟
**M2:** الحين ؟
**Najah:** ايوه
**M2:** تسعه واربعين سنه
**Najah:** ماشاءالله .. طيب هنا في السعوديه فيه اقارب لك؟
**M2:** فيه اممم..my husband فيه ثنين ولدين فيه
**Najah:** كم عمرهم؟
**M2:** ولد.. اول ولد ثلاثه عشرين سنه هو فيه شغل هنا هو مهندس حق mechanical مهندس .. فيه شغل هنا فيه واحد ثاني في مدرسه في secondary school

Najah: كم عمره؟
M2: ستطعش سنه
Najah: وكلهم موجودين هنا؟
M2: ايوه
Najah: يعني انتي ماتشعري بالغربه@
M2: لا مافيه@
Najah: طيب فيه اهل بالهند؟
M2: فيه ماما
Najah: كيف تتكلمين معاها؟
M2: فيه جوال.. كلام في جوال
Najah: ايش اللغه الي تتكلمين معاهم؟
M2: فيه مايلاليم
Najah: دائما تتكلمين ملاللاليم مع بعض حتى في البيت؟
M2: ايه دائما... في البيت بعد
Najah: الحين لغه مالاليالم فيه شبه مع لغه عربي ولا مختلف في grammar تركيب الجمله؟ يعني لما تقول جمله بالعربي فيه شبه زيو في مالاليلام؟
M2: فيه مشكله عربي هذا grammar فيه مشكله مو سيم سيم مالاليالم
Najah: مافيه شبه مره ؟
M2: لا.. مافيه مره
Najah: طيب والكلمات vocabulary فيه نفسو ممكن موجود في مالايالم؟
M2: موجود في الهند في المدرسه مثل مدرسه فيه كلاس عربي .. لكن مافيه كلام بس اكتب فيه exam مافيه احج كلام عربي
Najah: اطيب لما يجي مراجع يكلم معه عربي بكل سهوله؟
M2: ايوه.. انا اعرفه
Najah: كم جلستي هنا لين صرتي كويسه بالعربي؟
M2: يعني ممكن سنه كامل فيه مشكله شويه .. بعدين شويه شويه انا فيه واحد صديق مغربيه مع سوا سوا فيه شغل هي علم انا عربي
Najah: طيب لما يجي ناس مختلف ناس من دمام مثلا ولا جده انتي يشوف اختلاف عربي ولا كلو عربي زي واحد؟
M2: لا .. مافيه عربي كلو زي واحد .. فيه ..فيه من دمام فيه ثاني .. في السعوديه في جده.. فيه ثاني
Najah: طيب في الهند نفس الشي ؟
M2: ايوه فيه نفس الشي.. مالايالم فيه مكان مكان يختلف
Najah: طيب تتفرجين على tv ؟
M2: اول فيه يجي اول مره فيه يجي هنا سعوديه فيه عذا بس chanel اثنين بس..واحد عربي .. واحج انقلش .. بس. من زمان هذا.. مافيه هذا cable .. مافيه هذا دش مافيه.. يعني فيه كل يوم ..كل يوم فيه بس شوف عربي فلمات فيه كرتونات فيه كلو فيه شوف
Najah: وتفهمين اذا فيه مسلسل عربي؟
M2: اممم.. اممم. افهم
Najah: وتسمعين مسيقى عربي وتفهمين؟
M2: ايوه .. تفهمينه
Najah: حلو.. طيب انتي قلتي انتي بساكنه مع زوج صح؟
M2: ايوه بيت مع بعض
Najah: انتي الحين ممكن خمسه وعشرين سنه في السعوديه؟ صح طيب تكلمي كيف حياه هنا في السعوديه شي انتي حبي ولا شي انتي مايحبي هنا في السعوديه
M2: انا مره حبي هنا ... لكن هنا شغل كويس . فيه راتب كويس .. فيه ناس بعد كويسين .. فيه سيم food اكل هذا كلو فيه كويس هنا .. كلو هذا هنا كويس
Najah: طيب فيه شي ماحبيتيه ؟
M2: اه.. هذا مو سيم سيم هذا في الهند مافيه freedom.. freedom مافيه ..هذا بس مافيه ثاني بس مافيه freedom هنا سيم سسيم هند ولا فيه ثاني .. بس هذا .. لكن انا مافيه feel هذا مشكله انا كويس .. مافيه مشكله ::
Najah: طيب و الجو و الحر؟

M2: لا هذا كويسس .. عادي .. مافيه مشكله
Najah: طيب و الاكل هنا ... الاكل السعودي ؟
M2: مره..مره كويس@@
Najah: تعرفين تتطبخين؟
M2: ايوه
Najah: وش تطبخين؟
M2: فيه انا سوي هذا مخلوطه.. هذا رز مع كوسى هذا..هذا ايش اسمو هذا؟
Najah: كبسه
M2: ايوه.. سيم سيم هذا كبسه .. انا سوي في البيت .. فيه مندي .. فيه سوي بالبيت.. فيه ثاني بعد ..كوسى .. جزر..فيه بطاطا كلو مع دجاج فيه فوق فيه رز كذا سوي بالبيت .. انا كلو اطفال . .. زوج .. انا بعد كلو حبو هذا
Najah: ماشاءالله والاكل الهندي ممكن يشتري من السوق ؟ موجود
M2: ايه موجود فيه اشتري
Najah: طيب تكلمي اكثر عن خبره.. انتي جلستي خمسه و عشرين سنه
M2: يعني فيه.. كل خبره فيه كويس يعني مافيه مشكله يعني كلو كويس
Najah: طيب كل خمسسه وعشرين سنه هنا في السعوديه؟ ولا تروحين وترجعين ؟
M2: لا. لا.. فيه تروح وترجعين .. اول كلو سنتين فيه روح .. الحين كلو سنه فيه روح
Najah: طيب الحين بسالك سؤال ثاني.. انتي لما كنتي صغيره .. ايش الالعاب او هوايه hobbies
M2: Deawing ايش اسمو هذا drawing بالعربي؟
Najah: رسم
M2: اه رسم
Najah: كيف ترسمين ؟ وايش يستخدم اللوان؟
M2: اللوان على ورق
Najah: فيه لعبه ثاني غير الرسم ؟
M2: لا
Najah: بس رسم
M2: بس
Najah: انتي الحين رسام كويس
M2: الحين مافيه وقت@@
Najah: طيب فيه مره صار عليك حادث؟
M2: لا الحمدلله
Najah: ولا اهل ولا زوج؟
M2: لا.. الحمدلله
Najah: فيه مشاكل تفكرين فيها مثل خوف من شي من مرض .ز من موت... أي شي
M2: لا ولله .. هذا من الله . الحمدلله مافيه خوف
Najah: الحمدلله .. طيب انتي كيف تفكرين بالمستقبل؟
M2: مسستقبل يعني ايش؟
Najah: يعني future
M2: مستقبل.. انا فكر بعد سنه ولا سنتين لازم فيه روح هند؟؟ بعدين اجلس ماما.. ماما فيه يعني عائلي كبير بس اجلس هناك سوا سوا مافيه هنا في السعوديه .. خلي هذا ولد ممكن زوج هنا.. خلي هنا .. انا ابغى فيه روح عشان ثاني ولد بعد ممكن بعد شهر .. شهرين فيه روح هند عشان college
Najah: يعني انتي فتره شويه وخلاص بعدين روح الهند... يجي بعدين ولا خلاص؟
M2: ان شاءالله .. انا شوف ان شاءالله
Najah: طيب الحين انا ابي اوريك صور وانتي قولي ايش هذا صوره مين موجود في الصوره سمي شكل موجود داخل صوره .. هذا ايش وهذا ايش وكيف نستخدم administrative pronouns اسماء الاشاره بالعربي.. يعني I want you to name the objects that you see in the pictures وكيف قول هذا شي اذا كان بعيد او قريب.. اوك بس شوف صوره وتكلمي
M2: 1- هذا بنت قريب
2- اثنين سياره بعيد
3- هذا واحد كاميرا بعيد
4- هذا ولد قريب

5- هذا اثنين كوره قريب
6- هذا ثلاثه نفر بعيد
7- هذا قلم قريب
8- هذا ساعه بعيد
9- هذا اثنين حرمه قريب
10-هذا واحد بنت بعيد
11- هذا واحد طباخ بعيد
12- هذا ولد قريب
13- هذا اثنين بيت قريب
14- هذا واحد بنت بعيد
15-هذا اثنين حرمه قريب

**Najah:** طيب الحين انتي شوفي هذا صوره وقولي وين كلب موجود داخل او برا صندوق او وين .. هذا اسمه كلب . وهذا اسمه صندوق.. انتي كلام وين موجود كلب وكيف نستخدم propositions حروف الجر بالعربي

**M2:** ok

1- كلب فيه جوا في صندوق
2- كلب فيه فوق صندوق
3- كلب فيه برا صندوق
4- كلب تحت صندوق
5- كلب فيه قدام صندوق
6- كلب من ورا صندوق
7- كلب وسط اثنين صندوق
8- كلب من وسط اربعه صندوق
9- كلب من بعيد صندوق

**Najah:** طيب شكرا شكر thank you so much for your help and cooperation ومع السلامه

***Interviewee: P1***

***Punjabi newly-settled***

Najah: انتي من وين؟ انا من سعوديه انتي من وين ؟
P1: باكستان .. انا فيه باكستان
Najah: انا سعيده يشوف ناس من باكستان .. من وين في باكستان ؟
P1: اسلام اباد
Najah: ايش هذا لغه في باكستان ؟
P1: بنجابي
Najah: ولما كبير شويه كلام ايش .. ايش لغه ثاني؟
P1: اوردو لكن فيه مدرسه فيه انقلش
Najah: حلو كم عمرك؟
P1: سبعه عشرين
Najah: سبعه عشرين ايش؟ يعني شهر ولا سنه؟
P1: سنه !
Najah: انتي هنا وشو شغل في مسنشفى؟ ايش سوي؟ متى دوام ؟ متى خلاص؟ مين يشوفي كل يوم
P1: اممم عمليات ممرضه
Najah: تكلمي اكثر عن شغل؟ يعني انتي شغل كم ساعه؟ كيف سوي؟
P1: انا هنا ممرضه .. كلو عمليه انا assistant بعدين شوف patient غرفه افاقه .. بعدين غرفه فوق
Najah: طيب كم الساعه يجي الصباح ؟ يروح الساعه كم؟
P1: اممم..اثنينطعش ساعه..اثنينطعش اول .. سبعه يوم شهر بعدين ليل it's too difficult مافيه interaction مع patient مافيه كلام عربي كلو انجلش
Najah: بس انتي كويس الحين ماشاءالله لما يجي مريض تعرفين تتكلمي عربي .. ممكمن تقولين له روح فوق .. روح غرفه ثاني
P1: ايوه .. شويه شويه بس
Najah: طيب انتي خلاص مدرسه ابتدائي وبعدين وشو؟
P1: مدرسه اول primary بعدين college
Najah: طيب انتي لما يجي اول مره للسعوديه درستي Arabic training course
P1: لا.. مافيه
Najah: ولا في المدرسه درستي لغه عربي
P1: لا.. مافيه
Najah: طيب نفس السؤال انا سالت صديق انتي قبل شويه اذا صديق يجي سعوديه=
P1: لا انا واحد هنا من family
Najah: طيب الحين انتي كلام صديق في باكستان هو يبي يجي سعوديه عشان فيه شغل هنا.. انتي كلام هو لا يجي هنا اول لازم يتعلم عربي قبل يجي ولا مافيه مشكله يجي اول هنا سعوديه بعدين يتعلم عربي مع ناس؟ انتي كيف فكر احسن حق صديق؟
P1: if فيه هنا شغل كويس كلام عربك سرعه سرعه if in عمليات لازم course ايوه مشكله انا writte و read كلو عربي لكن مافيه معلوم ترجمه
Najah: كيف؟
P1: قران
Najah: اه.. ماشاءالله تقرائين قران عربي من قران ماشاءالله . طيب انتي قلتي انتي ساكن هنا
P1: ايوه.. اسكان ..واحد صديق roommate
Najah: فيه اصدقاء اقارب هنا في السعوديه؟
P1: لا.. مافيه
Najah: كيف تتكلمين مع صديقات هنا من الهند او فلبين؟
P1: انجليش
Najah: طيب كيف تتكلمين مع ماما
P1: في بنجابي
Najah: طيب ايش الطريقه الي تسخدمينها عشان تكلمي ماما يعني تلفون؟
P1: ايوه whatsup تلفون
Najah: كم مره كلم ماما؟
P1: ساعه ثنين ساعه ثلاثه كل يوم
Najah: تشعرين انتي يبي ماما

P1: ايوه كثير
Najah: طيب بنجدابي لغه فيه تشايه مع عربي ؟ ولا فيه اختلاف
P1: مافيه بنجابي بس شويه شويه مع اوردو
Najah: كيف كلمات و grammar
P1: حروف
Najah: بس كلمات يختلف ولا فيه=
P1: كلمات فيه .. بس شويه شويه .. حروف كلو حروف سيم سيم
Najah: صحيح حروف اردو زي عربي طيب و تركيب جمله grammar فيه مثل عربي؟
P1: لا .. يختلف
Najah: مثل هنا في عربي الجمله يبداء ب فعل بعدين subject اسم بعدين object لكن بنجابي كيف؟
P1: اه with اسم noun بعدين object بعدين verb
Najah: اه يختلف عن عربي
P1: ايوه
Najah: اه طيب مريض مايعرف انقليزي شنو سوي انتي
P1: شويه شويه انا معلوم
Najah: حلو.. . انتي تتفرجين tv
P1: لا
Najah: اممم طيب و المسيقى انتي تسمعين ميوزيك عربي ولا لا؟
P1: أي... ممكن
Najah: تفهمين ولا صعب؟
P1: شويه .. شويه @ " حبيبتي"
Najah: حلو . طيب تعجبك المسيقى العربيه؟
P1: ايوه .. @ كويس
Najah: انتي الحين فيه سنه ونص تقريبا في السعوديه.. كيف تشوفي الحياه هنا في السعوديه؟
P1: سعودي.. اممم .. كويس.. هنا بس فيه مشكله ... بعدين كلو كويس.. اممم .. برا كثير حراره ...داخل كلو كويس عشان air condition
Najah: طيب والاكل؟ كيف هنا
P1: اكل مافيه كويس@@
Najah: ماحبيتيه؟
P1: انا مافيه كويس اكل .. هنا مافيه اممم... كثيرrice مافيه vegetable taste مافيه كويس meet ..انا باكستان احسن
Najah: ممكن تروحين مطعم يشوف اكل باكستاني ؟ ولا مافيه
P1: شويه..شويه.. بس انا مافيه روح برا but شويه شويه انا شوف كويس
Najah: طيب .. لما كنتي صغير .. ايش الالعاب الي تلعبينها ؟
P1: اممم .. بس كلام ماما... كلو writing dairy انا شي انا كلام .. انا صديقه..كلو ..كل يوم.. انا اليوم شغل ... ايش.. ايش .. انا مذكرات
Najah: بس مافيه اللعاب ثاني؟
P1: اللاب توب .. فيه شغل لاب توب
Najah: طيب انتي للحين يتذكر المذكرات هذه؟
P1: ايوه.. فيه accommodation السكن .. ايوه .. عشره سنه @@
Najah: ماشاءالله.. طيب انتي فيه اشياء تخافين منها ؟
P1: انا.. روح برا بعدين.. فيه خوف
Najah: من ايش؟ ايش الي يخوف برا؟
P1: انا مافيه معلوم كثير عربي.. كلام خوف.. انا روح مول مشكله كلام if كلو عربي مشكله... telling for so خوف .. بعدين انا روح for sometime مشكله barging with people مشكله .. انا مافيه خوف كلو .. بس شويه شويه ... بيئة مختلف
Najah: طيب صار لك هنا حادث؟
P1: مافيه
Najah: الحمدلله طيب انتي مازلتي صغيره ...تكلمي عن المستقبل؟ ايش انتي فكر في المستقبل؟
P1: انا فكر مستقبل .. انا روح مدينه قريب قريب مسجد نبوي ...

**Najah:** اه يعني انتي تتمنى تروحي مدينه

**P1:** فيه اوردو كلام "تمنى" فيه عربي "اتمنى" امممم it's difficult قريب مسجد نبوي ..بعدين شغل موجود واحد مستشفى inside مسجد نبوي انا روح there ان شاءالله

**Najah:** ان شاءالله.. طيب وش فيه كمان

**P1:** انا call ماما .. هنا في مدينه .. ان شاءالله

**Najah:** ان شاءالله.... طيب الحين انا ابي اوريك صور وانتي قولي ايش هذا صوره مين موجود في الصوره سمي شكل موجود داخل صوره .. هذا ايش وهذا ايش وكيف نستخدم administrative pronouns اسماء الاشاره بالعربي.. يعني I want you to name the objects that you see in the pictures وكيف قول هذا شي اذا كان بعيد او قريب.. اوك بس شوف صوره وتكلمي

1- هذا بنت قريب
2- هذا اثنين سياره بعيد
3- هذا كاميرا بعيد
4- هذا ولد قريب
5- هذا اثنين كوره قريب
6- ثلاثه مسافر بعيد
7- هذا قلم
8- هذا ساعه
9- هذا اثنين ماما
10-هذا بنت
11- هذا طباخ
12- هذا واحد ولد
13- هذا اثنين مكان
14- هذا واحد بنات
15-هذا اثنين ماما

**Najah:** طيب الحين انتي شوفي هذا صوره وقولي وين كلب موجود داخل او برا صندوق او وين .. هذا اسمه كلب . وهذا اسمه صندوق.. انتي كلام وين موجود كلب وكيف نستخدم propositions حروف الجر بالعربي

**P1:** ok

1- كلب داخل صندوق
2- كلب فوق صندوق
3- كلب بعدين صندوق
4- كلب تحت صندوق
5- كلب يمين صندوق
6- كلب يسار صندوق
7- كلب سوا سوا اثنين صندوق
8- كلب سوا سوا اربعه صندوق
9- كلب بعيد صندوق

**Najah:** طيب شكرا شكر thank you so much for your help and cooperation ومع السلامه

**Najah** انتي من وين؟ انا من سعوديه انتي من وين ؟

**P2** باكستان .. انا فيه باكستان

**Najah** انا سعيده يشوف ناس من باكستان .. من وين في باكستان ؟

**P2** اسلام اباد

**Najah** ايش هذا لغه في باكستان ؟

**P2** بنجابي

**Najah** ولما كبير شويه كلام ايش .. ايش لغه ثاني؟

**P2** اوردو لكن فيه مدرسه فيه انقلش

**Najah** حلو كم عمرك؟

**P2** سبعه عشرين

**Najah** سبعه عشرين ايش؟ يعني شهر ولا سنه؟

**P2** سنه !

**Najah** انتي هنا وشو شغل في مسنشفى؟ ايش سوي؟ متى دوام ؟ متى خلاص؟ مين يشوفي كل يوم

**P2** اممم عمليات ممرضه

**Najah** تكلمي اكثر عن شغل؟ يعني انتي شغل كم ساعه؟ كيف سوي؟

**P2** انا هنا ممرضه .. كلو عمليه انا assistant بعدين شوف patient غرفه افاقه .. بعدين غرفه فوق

**Najah** طيب كم الساعه يجي الصباح ؟ يروح الساعه كم؟

**P2** اممم..اثنينطعش ساعه..اثنينطعش اول .. سبعه يوم شهر بعدين ليل it's too difficult مافيه interaction مع patient مافيه كلام عربي كلو انجلش

**Najah** بس انتي كويس الحين ماشاءالله لما يجي مريض تعرفين تتكلمي عربي .. ممكمن تقولين له روح فوق .. روح غرفه ثاني

**P2** ايوه .. شويه شويه بس

**Najah** طيب انتي خلاص مدرسه ابتدائي وبعدين وشو؟

**P2** مدرسه اول primary بعدين college

**Najah** طيب انتي لما يجي اول مره للسعوديه درستي Arabic training course

**P2** لا.. مافيه

**Najah** ولا في المدرسه درستي لغه عربي

لا.. مافيه

**Najah** طيب نفس السؤال انا سالت صديق انتي قبل شويه اذا صديق يجي سعوديه=

لا انا واحد هنا من family

طيب الحين انتي كلام صديق في باكستان هو يبي يجي سعوديه عشان فيه شغل هنا.. انتي كلام هو لا يجي هنا اول لازم يتعلم عربي قبل يجي ولا مافيه مشكله يجي اول هنا سعوديه بعدين يتعلم عربي مع ناس؟ انتي كيف فكر احسن حق صديق؟

if فيه هنا شغل كويس كلام عربك سرعه سرعه if in عمليات لازم course ايوه مشكله انا writte و read كلو عربي لكن مافيه معلوم ترجمه

كيف؟

قران

اه.. ماشاءالله تقرائين قران عربي من قران ماشاءالله . طيب انتي قلتي انتي ساكن هنا

ايوه.. اسكان ..واحد صديق roommate

**Najah** فيه اصدقاء اقارب هنا في السعوديه؟

**P2** لا.. مافيه

**V** كيف تتكلمين مع صديقات هنا من الهند او فلبين؟

**P2** انجليش

**Najah** طيب كيف تتكلمين مع ماما

P2 في بنجابي

Najah طيب ايش الطريقه الي تسخدمينها عشان تكلمي ماما يعني تلفون؟

P2 ايوه whatsup تلفون

Najah كم مره كلم ماما؟

P2 ساعه ثنين ساعه ثلاثه كل يوم

Najah تشعرين انتي يبي ماما

P2 ايوه كثير

Najah طيب بنجدابي لغه فيه تشايه مع عربي ؟ ولا فيه اختلاف

P2 مافيه بنجابي بس شويه شويه مع اوردو

Najah كيف كلمات و grammar

P2 حروف

Najah بس كلمات يختلف ولا فيه=

P2 كلمات فيه .. بس شويه شويه .. حروف كلو حروف سيم سيم

Najah صحيح حروف اردو زي عربي طيب و تركيب جمله grammar فيه مثل عربي؟

P2 لا .. يختلف

Najah مثل هنا في عربي الجمله يبداء ب فعل بعدين subject اسم بعدين object لكن بنجابي كيف؟

P2 اه with اسم noun بعدين object بعدين verb

Najah اه يختلف عن عربي

P2 ايوه

Najah اه طيب مريض مايعرف انقليزي شنو سوي انتي

P2 شويه شويه انا معلوم

Najah نو .. لا .. مافيه

P2 طيب راديو ولا ميوزيك

Najah لا . مافيه

P2 حتى ميوزك مافيه؟

Najah ميوزيك فيه بس انجلش وتقالوق

P2 يعني لو يسمع مسيقا عربي فيه معلوم؟

Najah لا مافيه معلوم

P2 طيب انتي يبغى يشوف ماما كثير؟

Najah ايوه... ايوه

P2 كيف كلم اهل..ماما بابا في فلبين؟ كيف يسوي اتصال مع عائله؟

Najah ايوه فيديو كول .. جوال .. موبايل

P2 طيب انتي فيه هنا سعوديه سنه و ثمانيه شهر.. كيف يشوف سعوديه يعني برا مستشفى؟

Najah شويه اوكي.. انا فيميل مافيه روح اني وير

P2 اه .. طيب تكلمي عربي

Najah مشكله مافيه معلوم

P2 شويه شويه يتكلم عربي

Najah اه..معليش تو ديفيكلت

P2 طيب يتكلم الحين عن اكل.. ايش اكل هنا؟

Najah رايس

P2 لا .. بالعربي

Najah اوه مافيه معلوم اربك ايش هذا.. مافيه معلوم اربك رايس اربك تشيكن

P2 طيب انتي الحين ياكل اكل فلبيني او سعودي اكل؟

Najah فلبيني

P2 كلو فلبيني حتى في كمبني؟

Najah ايوه كلو فلبيني

P2 طيب لما انتي صغير في فلبين ايش الالعاب الي انتي يلعب دايما؟

Najah اه ... فلبين.... مافيه... لا.... مافيه معلوم عربك

P2 طيب انتي فيه فكر كثير في مستقبل فيوتشر؟

Najah ايش فكر..هذا

P2 يعني انتي ايش فكر حق فيوتشر؟

Najah اه... اربيك توو ديفيكلت .. مافيه معلوم explain Arabic معليش .. معليش

P2 فيه شي انتي worried about مثل حادث accident

Najah فيه بس.. مافيه معلوم ان اربك سوررري

P2 **********

Najah طيب لما يدرس عربي.. درستي عربي standard Arabic العربي الفصحه؟

P2 ايش يعني فصحه؟

Najah فصحه يعني .. انا الحين مافيه كلام عربي كويس ميه ميه. لكن اذا يبي يتكلم عربي فصحه عربي ممتاز يقول من انتي؟ ما اسمك؟ اين ذهبتي اليوم؟ ومع من؟ هل تناولتي وجبه الافطار قبل ذهابك؟

P2 اه

Najah هذا standard Arabic

P2 ايوه.. انا تعلم standard Arabic

Najah الحين قرامر عربي مثل قرامر تقالوق؟ يعني مثل العربي او فيه اختلاف؟ يعني قرمر عربي جمله يبدا ب verb اول وبعدين sbject بتقلوق نفس الشي؟

P2 امممممم

Najah يعني مثلا ذهبت ندى للمدرسه؟ او مدرسه ذهبت ندى؟

P2 لا.. لا.. يختلف

Najah كيف؟

P2 الحين مايعرف قول كيف ..لكن يختلف

Najah طيب انتي تشوفين TV هنا؟ شوفي برامج عربي ولا ايش؟

P2 شوف انجليزي بس

Najah يعني مافيه قناه فيها عربي؟ او ميوزك عربي؟

P2 لا.. كلو انجليزي

Najah طيب تكلمي عن السعوديه.. ايش اشياء كويس يعني انتي حبي واشياء مافيه حبي؟

P2 السعوديه كويس عشان غيه شغل.. يعني straight سيدا مافيه كذا او كذا.. عشان انتي معلوم حكومه..صح؟ يعني كويس احسن من فلبين

Najah ماشاءالله كويس لكن مو لازم كلام عن شغل .. ممكن كلام عن حياه هنا

P2 مش كله.. هنا مو زي فلبين.. فلبين. يعني عادي .. يعني لازم اللبس عبايه

Najah اه .. طيب.. كيف اكل هنا كويس؟

P2 هنا كويس عشان فيه بعد مطعم .. وفيه مطعم فلبين.. وفيه بعد يعني سوي زي طبخ فلبين

Najah حلو.. يعني تطبخين في البيت؟

**P2** ايوه

**Najah** حتى لما يروح برى ممكن يشتري اكل فلبين؟

**P2** ايوه.. ايوه..صح..فيه موجود كل شي

**Najah** انتي وقت فراغ يعني وقت مافيه شغل ..ايش سوي؟

**P2** يعني.. موجود بيت.. فيه سوي علم بيبي انا .. سوي نظيف.. سوي كذا@

**Najah** فيه طبخ هنا انتي يحب يسويه هنا؟

**P2** طبخ عربي؟

**Najah** ايوه

**P2** بس كبسه

**Najah** تعرفين تسوين كبسه؟

**P2** ايوه

**Najah** كيف تعلمتي؟ من وين؟

**P2** مال يوتيوب..انا معلوم حق يوتيوب.. انا شوف حق يوتيوب عربي.. بعدين كذا@ معلوم سوي

**Najah** حلو ماشاءالله..طيب انتي لما كنتي صغير ايش اللعاب سوي؟

**P2** انا صغير؟

**Najah** ايوه

**P2** اللعب حق.. بس انا معلوم فلبين بس..عادي .. عادي فلبين.. يعني .. اممم..حق اللعب ..حقق معلوم هذا نفس ..فيه نفس... لازم اثنين نفر.. يعني سوي كذا ..كذا @@

**Najah** مافيه مشكله .. تكلمي عادي انا بس يبي يشوف كيف عربي انتي مو مهم صح او غلط

**P2** اه .. طيب

**Najah** انتي الحين يتذكر هذه اللعبه؟ مافيه انسي؟

**P2** ايوه.. لازم مافيه انسي@

**Najah** طيب انتي يعلم اطفالك كيف يلعبو هذه اللعبه؟ ولا خلاص؟

**P2** لا.. خلاص..الحين. قبل كذا..بس الحين كلو حق high technology يعني يوتيوب.. games صح؟

**Najah** ايوه.. صحيح مافيه زي اول activities

**Najah** طيب انتي نفكرين بالمستقبل future كثير؟

**P2** ايوه .. لازم .. لازم فيه كثير@

**Najah** ايش تفكرين؟

**P2** يعني ان شاءالله بعد. عشره سنه انا خلاص .. انا بسوي بزننس يعني مال بلاد انا .. وماسوي شغل شي... بس سوي بزنس .. وان شاءالله .. فيه فلوس..فيه دخل ..يعني كذا

*Interviewee: T1*

*Tagalog newly-settled*

**Najah:** انتي من وين؟ انا من سعوديه انتي من وين ؟
**T1:** من فلبين
**Najah:** انا كلام عربي انتي كلام ايش في فلبين؟
**T1:** تاقلاوق من فلبين
**Najah:** اانتي لما صغير كلام ايش؟ كلام ايش اول شي؟
**T1:** تاقلاوق
**Najah:** ولما كبير شويه كلام ايش .. ايش لغه ثاني؟
**T1:** انجلش
**Najah:** كم سنه هنا في سعوديه؟
**T1:** سنه...واحد سنه..اثنين... شهر..واحد سنه..ثنين شهر
**Najah:** انتي هنا ايش شغل في مستشفى؟
**T1:** ايش؟
**Najah:** انتي هنا وشو شغل في مسنشفى؟ ايش سوي؟
**T1:** nurse
**Najah:** كيف سوي nurse ؟ يعني انتي شغل كم ساعه؟ كيف سوي؟
**T1:** هنا one year...واحد ... ايش؟
**Najah:** لا.. لا...اني يجي الصباح الساعه كم؟ يروح الساعه كم؟
**T1:** خمسه ساعه...عشره..
**Najah:** عشره ايش؟ الصباح؟ ولا متى؟
**T1:** صباح ساعه ثنين.. ساعه... ثنعش.. بعدين ليل ساعه خمسه..ساعه عشره
**Najah:** طيب .. قبل يجي سعوديه فيه شغل ثاني قبل؟ يعني فيه شغل قبل في كويت؟ او دبي؟
**T1:** لا.. لا..بس.. اول فلبين .. هنا.. رياض
**Najah:** في فلبين كيف يتكلم عربي اول ؟training course اخذتي
**T1:** لا.. مافيه
**Najah :** طيب لما يجي هنا سعوديه اخذتي أي course training بالعربي؟
**T1:** لا
**T1:** طيب لما يشوف صديق جديد.. انتي كلام احسن اخذ course training عربي قبل يجي ؟ ولا مافيه مشكله ممكن يتتعلم عربي؟
**T1:** ايش؟
**Najah:** يعني الحين انتي كلام صديق في فلبين هو يبي يجي سعوديه عشان فيه شغل هنا.. انتي كلام هو لا يجي هنا اول لازم سويcourse training بالعربي قبل يجي ولا مافيه مشكله يجي اول هنا سعوديه بعدين يتعلم عربي مع ناس؟ انتي كيف فكر احسن حق صديق؟
**T1:** Wait I will translate in English…it, ok?
**Najah:** لا..لا انا يبي يتكلم انتي بالعربي مابي انجلش
**T1:** As what I understand … it's hard to explain in Arabic
**Najah:** Ok, I will skip this question
**T1:** Yes, please..skip …skip…pass, I promise
**Najah:** طيب لما درستي في فلبين.. ايش المستوى الي درستي فيه؟ يعني خلاص هاي سكول وبعدين؟
**T1:** Elementary, high school, college
**Najah:** طيب.. كم عمرك؟

T1: سبعه عشرين

Najah: انتي فيه زواج؟

T1: مافيه

Najah: انتي هنا وين ساكنه؟ يعني مع company؟ ولا friend؟ هنا في سعوديه انتي وين عايش؟ وين نوم؟ مع سكن شركه ولا مع صديق؟

T1: مع company..

Najah: طيب هنا في رياض انتي فيه صديق من فلبين؟

T1: ايوه

Najah: شوف صديق هنا ؟فيه سوي زياره؟

T1: ايوه شويه

Najah: متى؟ انتي فيه شوف؟

T1: فيه aunt يجي هنا .. بس مافيه انا.. مافيه روح

Najah: طيب انتي يسخدم عربي وين؟ ممكن يطلع برا من مستشفى ممكن يجي كلام عربي مع ناس؟ ولا مافيه؟

T1: ايش هذا؟

Najah: يعني برا مستشفى انتي كلام ايش ؟ عربي او انجلش او تاقالوق؟

T1: انجلش شويه

Najah: طيب لما يجي بايشنت مريض هو مافيه معلوم انجلش كيف انتي فيه كلام معه؟

T1: انا سوي صديق ... امممم translation @ ايوه @

Najah: انتي يبي يتعلم عربي او لا؟ ?you want to learn Arabic or not

T1: ايوه... بس صديق كويس ميه ميه هو ساعد انا

Najah: طيب هنا ناس كلو يتعلم عربي وانتي ؟

T1: كلو .. لا مافيه كلو

Najah: يعني ممكن كلو يعرف عربي .. وانتي مافيه معلوم عربي ؟ كيف فكر انتي يتعلم عربي ؟ يعني فيه dictionary فيه بيت حق عربي؟

T1: لا ... مافيه @

Najah: صح؟ TV اكيد انتي فيه

T1: no .. لا .. مافيه

Najah: طيب radio ولا music

T1: لا . مافيه

Najah: مافيه؟music حتى

T1: Music فيه بس انجلش وتقالوق

Najah: يعني لو يسمع مسيقا عربي فيه معلوم؟

T1: لا مافيه معلوم

Najah: طيب انتي يبغى يشوف ماما كثير؟

T1: ايوه... ايوه

Najah: كيف كلم اهل..ماما بابا في فلبين؟ كيف يسوي اتصال مع عائله؟

T1: ايوه video call .. جوال .. mobile

Najah: طيب انتي فيه هنا سعوديه سنه و ثمانيه شهر.. كيف يشوف سعوديه يعني برا مستشفى؟

T1: شويه اوكي.. انا female مافيه روح anywhere

N:ajah: اه .. طيب تكلمي عربي

T1: مشكله مافيه معلوم

Najah: شويه شويه يتكلم عربي

T1: اه..معليش too difficult

Najah: طيب يتكلم الحين عن اكل.. ايش اكل هنا؟

T1: rice

Najah: لا .. بالعربي

T1: اوه مافيه معلوم Arabic ايش هذا.. مافيه معلوم Arabic rice Arabic chicken

Najah: طيب انتي الحين ياكل اكل فلبيني او سعودي اكل؟

T1: فلبيني

Najah: كلو فلبيني حتى في company؟

T1: ايوه كلو فلبيني

Najah: طيب لما انتي صغير في فلبين ايش الالعاب الي انتي يلعب دايما؟

T1: اه ... فلبين.... مافيه... لا.... مافيه معلوم Arabic

Najah: طيب انتي فيه فكر كثير في مستقبل future؟

T1: ايش فكر..هذا

Najah: يعني انتي ايش فكر حق future؟

T1: اه... Arabic too difficult .. مافيه معلوم explain Arabic معليش .. معليش

Najah: فيه شي انتي worried about مثل حادث accident

T1: فيه بس.. مافيه معلوم in Arabic سوررري

Najah: طيب الحين انا ابي اوريك صور وانتي قولي ايش هذا صوره مين موجود في الصوره سمي شكل موجود داخل صوره .. هذا ايش وهذا ايش وكيف نستخدم administrative pronouns اسماء الاشاره بالعربي.. يعني I want you to name the objects that you see in the pictures وكيف قول هذا شي اذا كان بعيد او قريب.. اوك بس شوف صوره وتكلمي

T1: 1- بنت

2- سياره..اثنين سياره

3- كاميرا

4- ولد

5- كوره ..اثنين كوره

6- ثلاثه cameramen ..ثلاثه مصور

7- قلم

8- ساعه

9- اثنين عجوزه

10- بنت

11- chef .. طباخ

12- ولد

Najah: انتي ماستخدمتي الضمائر pronouns اذا كان بعيد faraway او قريب closed مثل هنا او هناك

T1: مافيه معلوم .. انا معلوم هنا..هنا

Najah: لازم يستخدم عربي كثير

T1: 13- اثنين بيت

14- بنت

15- اثنين عجوز

**Najah:** طيب الحين انتي شوفي هذا صوره وقولي وين كلب موجود داخل او برا صندوق او وين .. هذا اسمه كلب . وهذا اسمه صندوق.. انتي كلام وين موجود كلب وكيف نستخدم propositions حروف الجر بالعربي

**T1:** ok

1- كلب جوا هذا صندوق
2- كلب فوق صندوق
3- كلب مافيه معلوم..كلب ورا صندوق
4- كلب تحت صندوق
5- كلب infront .. قدام..كلب قدام صندوق
6- كلب beside ..كلب جنب
7- كلب وسط اثنين صندوق
8- كلب وسط اربعه صندوق
9- كلب فوق صندوق

**Najah:** طيب شكرا شكر thank you so much for your help and cooperation ومع السلامه

***Interviewee: T2***

*Tagalog Long-term resident*

**Najah:** هاي .. من وين انتي؟
**T2:** من فلبين
**Najah:** ايش المدينه الي انتي فيها في فلبين؟ ايش المدينه؟
**T2:** يعني مدينه؟
**Najah:** يعني انا هنا من رياض.. فيه ناس ثاني من جده.. انتي من وين في فلبين؟
**T2:** انا من داناو
**Najah:** كم سنه انتي هنا في سعوديه؟
**T2:** انا خمسه سنه
**Najah:** كلو في سعوديه ولا ممكن يشتغل في كويت؟
**T2:** اول انا في كويت
**Najah:** يعني كلو سنه كم؟ في الكويت او السعوديه.. يعني في الخليج؟
**T2:** انا خمسه سنه بعد من كويت.. وهنا بعد خمسه سنه
**Najah:** ايش يشتغل هنا؟ ايش سوي؟ وين؟
**T2:** انا شغل في مستشفى
**Najah:** كيف شغل؟
**T2:** كويس الحمدلله
**Najah:** يعني كل يوم يجي في صباح ..بعدين يروح يشوف مريض
**T2:** انا شغل ساعه سته ..يعني خلاص عشره..يعني ثمانيه عشره
**Najah:** انتي قلتي هنا في سعوديه خمسه سنه؟
**T2:** ايوه
**Najah:** نفس شغل ؟ او غير؟
**T2:** اول لا ..اول شغل مال شركه
**Najah:** ايش يسوي يعني؟
**T2:** فيه مكتب
**Najah:** طيب انتا لما يجي سعوديه.. اخذتي ترينق كورس بالعربي؟
**T2:** لا..لا..مافيه
**Najah:** انتي ايش فكر احسن لو ياخذ كورس عربي ولا مافيه مشكله؟
**T2:** ولله كويس فيه كورس عربي ..بس عادي.. يعني بس شويه شويه يعلم
**Najah:** يعني تعلمتي شويه شويه عربي عشان من مرضى؟
**T2:** ايوه..ايوه
**Najah:** انتي الان مبسوط من عربي؟مايحتاج ياخذ كورس زياده؟
**T2:** ايوه..ايوه
**Najah:** لما يجي واحد جديد مثل صديق انتي هذا..ايش يقوليله لازم ياخذ كورس عربي احسن او ممكن يتعلم؟
**T2:** ممكن يتعلم شويه شويه
**Najah:** انتي لما كنتي في فلبين.. فيه مدرسه؟
**T2:** ايوه
**Najah:** ايش خلص؟
**T2:** يعني خلاص( بالنيرسنق ) وبعدين خلاص( غير واضح)
**Najah:** كم سنه.. كم سنه؟

T2: يعني اربعه سنه

Najah: كل مرحله اربعه سنه؟ وبعدين اربعه سنه ثاني؟

T2: ثلاثه سنه واحد ثاني

Najah: انتي متزوجه؟

T2: ايوه

Najah: فيه اطفال؟

T2: ايوه

Najah: كم ؟

T2: فيه واحد هنا من الحين @ جديد

Najah: يعني شنو سمي هذا

T2: حامل @

Najah: ايه ماشاءالله ..كم شهر؟

T2: سته شهور

Najah: اولاد فيه فلبين مع مين؟

T2: مافيه..موجود هنا

Najah: كم بنت؟ كم ولد؟

T2: كلو ولد

Najah: كم عمر؟

T2: واحد سنه و اثنين شهور

Najah: فيه قريب ريلتيف بالسعوديه؟

T2: لا.. مافيه

Najah: انتي وين ساكن الحين؟

T2: مع جوز و اطفال

Najah: ايه ماشاءالله.. انا ابغاك انتي تتكلمين كثير احسن.. مابي بس ايوه ..لا...عادي مافيه مشكله هذا كلو فريندلي تشات

T2: ايوه

Najah: انتي كلام اول فيه كويت؟

T2: ايوه

Najah: وكيف كلام معع ناس ؟ نفسه عربي هنا؟

T2: لا.. غير..صح؟ كلام يعني.. فيه ثاني..يعني هنا ليش سوي كذا..الكويت ليش سوي جذي @

Najah: ايوه.. صحيح

T2: مو نفس الشي

Najah: يعني انتي شفتي الفرق؟

T2: ايوه

Najah: كنتي تفهمين عليهم ماعندك مشكله؟

T2: ايوه..عشان بعد فيه شغل مكتب

Najah: ايش اللغه الاولى في فلبين؟ first language

T2 اه .. تقالوق

Najah: كلو فلبينن يتكلم تاقالوق؟

T2: لا..لا...مو كلو..يعني ثاني مكان لوسون..هذا تاقلوق..وباساسي ..هذا بسايا..من دوناو هذا ثاني كلام..هذا مسلم مكان

**Najah:** اه.. طيب انتي الحين ينكلم عربي دائما..دائما؟ ولا متى؟ وكم ساعه ؟ و وين؟

**T2:** مو دايم..يعني ناس كلام انجليزي.. انا كلام انجليزي.. ناس كلام عربي.. كلام بس عربي.. يعني مو بيرفكت يعني مو كلو

**Najah:** طيب وبرى شغل يتكلم عربي؟

**T2:** برى؟

**Najah:** ايوه

**T2:** ايوه كلام عربي

**Najah:** ممكن يشوف ناس ..يتكلم عربي؟

**T2:** ايوه.. ايوه

**Najah:** اذا يجي واحد يتكلم عربي .. انتي تفهمين بسرعه؟ ولا لازم يعيد كلام؟

**T2:** يعني عربي.. يعني.. مافيه .. ايش هذا.. يعني عربي يعني بايزك انا فيه افهم

**Najah:** طيب.. لما يجي من مدينه ثاني..مثلا دمام.. او أي مكان ثاني عادي مافيه فرق عن ساكن رياض؟

**T2:** ايوه.. معلوم

**Najah:** انتي اول ماتكلكتي عربي زمان .. كان صعب ؟ او بسيط؟

**T2:** يعني.. يعني.. انا معلوم .. عشان انا اول فيه كويت

**Najah:** قبل كويت؟

**T2:** قبل كويت.. فيه حق فلبين. فيه يعني .. حق حكومه. فيه ورق حق ربك بيزيك يعني خلي قبل اطلع مال روح سفر يعني .. روح اركب.. فيه هذا كتاب

**Najah:** حلو .. يعني انتي درستي بالفلبين؟

**T2:** ايوه

**Najah:** طيب لما يدرس عربي.. درستي عربي standard Arabic العربي الفصحه؟

**T2:** ايش يعني فصحه؟

**Najah:** فصحه يعني .. انا الحين مافيه كلام عربي كويس ميه ميه. لكن اذا يبي يتكلم عربي فصحه عربي ممتاز يقول من انتي؟ ما اسمك؟ اين ذهبتي اليوم؟ ومع من؟ هل تناولتي وجبه الافطار قبل ذهابك؟

**T2:** اه

**Najah:** هذا standard Arabic

**T2:** ايوه.. انا تعلم standard Arabic

**Najah:** الحين قرامر عربي مثل قرامر تقالوق؟ يعني مثل العربي او فيه اختلاف؟ يعني قرمر عربي جمله يبدا ب verb اول وبعدين sbject بتقلوق نفس الشي؟

**T2:** امممممم

**Najah:** يعني مثلا ذهبت ندى للمدرسه؟ او مدرسه ذهبت ندى؟

**T2:** لا.. لا.. يختلف

**Najah:** كيف؟

**T2:** الحين مايعرف قول كيف ..لكن يختلف

**Najah:** طيب انتي تشوفين TV هنا؟ شوفي برامج عربي ولا ايش؟

**T2:** شوف انجليزي بس

**Najah:** يعني مافيه قناه فيها عربي؟ او ميوزك عربي؟

**T2:** لا.. كلو انجليزي

**Najah:** طيب تكلمي عن السعوديه.. ايش اشياء كويس يعني انتي حبي واشياء مافيه حبي؟

**T2:** السعوديه كويس عشان غيه شغل.. يعني straight سيدا مافيه كذا او كذا.. عشان انتي معلوم حكومه..صح؟ يعني كويس احسن من فلبين

**Najah:** ماشاءالله كويس لكن مو لازم كلام عن شغل .. ممكن كلام عن حياه هنا

T2: مش كله.. هنا مو زي فلبين.. فلبين. يعني عادي .. يعني لازم اللبس عبايه

Najah: اه .. طيب.. كيف اكل هنا كويس؟

T2: هنا كويس عشان فيه بعد مطعم .. وفيه مطعم فلبين.. وفيه بعد يعني سوي زي طبخ فلبين

Najah: حلو.. يعني تطبخين في البيت؟

T2: ايوه

Najah: حتى لما يروح برى ممكن يشتري اكل فلبين؟

T2: ايوه.. ايوه..صح..فيه موجود كل شي

Najah: انتي وقت فراغ يعني وقت مافيه شغل ..ايش سوي؟

T2: يعني.. موجود بيت.. فيه سوي علم بيبي انا .. سوي نظيف.. سوي كذا@

Najah: فيه طبخ هنا انتي يحب يسويه هنا؟

T2: طبخ عربي؟

Najah: ايوه

T2: بس كبسه

Najah: تعرفين تسوين كبسه؟

T2: ايوه

Najah: كيف تعلمتي؟ من وين؟

T2: مال يوتيوب..انا معلوم حق يوتيوب.. انا شوف حق يوتيوب عربي.. بعدين كذا@ معلوم سوي

Najah: حلو ماشاءالله..طيب انتي لما كنتي صغير ايش اللعاب سوي؟

T2: انا صغير؟

Najah: ايوه

T2: اللعب حق.. بس انا معلوم فلبين بس..عادي .. عادي فلبين.. يعني .. اممم..حق اللعب ..حقق معلوم هذا نفس ..فيه نفس... لازم اثنين نفر.. يعني سوي كذا ..كذا @@

Najah: مافيه مشكله .. تكلمي عادي انا بس يبي يشوف كيف عربي انتي مو مهم صح او غلط

T2: اه .. طيب

Najah: انتي الحين يتذكر هذه اللعبه؟ مافيه انسي؟

T2: ايوه.. لازم مافيه انسي@

Najah: طيب انتي يعلم اطفالك كيف يلعبو هذه اللعبه؟ ولا خلاص؟

T2: لا.. خلاص..الحين. قبل كذا..بس الحين كلو حق high technology يعني يوتيوب.. games صح؟

Najah: ايوه.. صحيح مافيه زي اول activities طيب انتي نفكرين بالمستقبل future كثير؟

T2: ايوه .. لازم .. لازم فيه كثير@

Najah: ايش تفكرين؟

T2: يعني ان شاءالله بعد. عشره سنه انا خلاص .. انا بسوي بزننس يعني مال بلاد انا .. وماسوي شغل شي... بس سوي بزنس .. وان شاءالله .. فيه فلوس..فيه دخل ..يعني كذا

Najah: ايوه.. صح.... طيب الحين انا ابي اوريك صور وانتي قولي ايش هذا صوره مين موجود في الصوره سمي شكل موجود داخل صوره .. هذا ايش وهذا ايش وكيف نستخدم administrative pronouns اسماء الاشاره بالعربي.. يعني I want you to name the objects that you see in the pictures وكيف قول هذا شي اذا كان بعيد او قريب.. اوك بس شوف صوره وتكلمي:

T2 1- هذا بنت

2- هذا اثنين سياره

3- هذا كاميرا

4- هذا ولد

5- هذا اثنين كوره

6- هذا ثلاثه مصور

7- هذا قلم

8- هذا ساعه

9- هذا اثنين عجوزه

10-هذا بنت

11- هذا طباخ

12- هذا ولد

13- هذا اثنين بيت

14- هذا واحد بنت

15-هذا اثنين عجوزه

**Najah:** طيب الحين انتي شوفي هذا صوره وقولي وين كلب موجود داخل او برا صندوق او وين .. هذا اسمه كلب . وهذا اسمه صندوق.. انتي كلام وين موجود كلب وكيف نستخدم propositions حروف الجر بالعربي

**T2:** ok

1- كلب جوا هذا صندوق

2- كلب برا هذا صندوق

3- كلب برا صندوق

4-- كلب تحت صندوق

5- كلب قدام صندوق

6- كلب ورا صندوق

7- هذا كلب وسط صندوق

8- هذا كلب نص صندوق اربعه

9- كلب jump صندوق

**Najah:** طيب شكرا شكر thank you so much for your help and cooperation ومع السلامه

**Interviewee: S1**

***Sinhalese newly-settled***

**Najah:** انتي من وين؟ **from where**

**S1:** سيريلنكا

**Najah:** ايش لغه اول؟ **first language**

**S1:** سينهالا **Sinhala**

**Najah:** طيب ثاني لغه **second language**

**S1:** انجلش **English**

**Najah:** طيب فيه ثالث لغه **third**

**S1:** تاميل

**Najah:** كم يجلس انتي هنا في سعوديه؟

**S1:** امممم **8mounth Saudi**

**Najah:** يعني كم بالعربي؟؟ ثمانيه

**S1:** ثمانيه شهر

**Najah:** انتي من وين في سيرلينكا ؟ انا هنا من رياض صديق من دمام .. انتي سرلينكا من وين مدينه؟

**S1:** سيرلنكا ...نقمبو

**Najah:** انتي وشو شغل هنا ؟

**S1:** خدامه

**Najah:** طيب انتي كيف شغل هنا ؟ متى يقوم من النوم ؟ كلمي عربي

**S1:** بيت **good**

**Najah:** كلام عربي مافيه مشكله .. متى يسوي تنظيف ؟ متى يقوم

**S1:** ببيي **good** ... مدام **good** @@انا **good** ..مافيه مشكله ببيي كويس .. مدام كويس ...بيت كويس .. انا كويس @@ **no Arabi** ... انا ببيي مدرسه **morning go** ببيي اكل ...اكل خبز ...حليب ..ببيي **rice** شويه اكل ببيي **play ball**

**Najah:** طيب انتي لما كنتي في سيرلنكا فيه **Arabic training course** ولا مافيه

**S1;** مافيه

**Najah:** فيه زوج ؟ فيه **married**

**S1:** ايوه

**Najah:** كم زوج؟

**S1:** واحد

**Najah:** فيه اطفال؟ **babies**

**S1:** فيه **baby**

**Najah:** كم؟ عادي تكلمي عربي

**S1:** ثنين بنت .. ثتين ولد

**Najah:** كم عمر اولاد ؟ يعني **age**

**S1:** صغير ولد ثنين..كبير بنت ثنين

**Najah:** فيه مدرسه اولاد بنات؟

**S1:** ايوه.. كلو بنت مدرسه..

**Najah:** فيه ماما هناك؟

**S1:** انا ماما؟

**Najah:** ايوه

**S1:** فيه..فيه

**Najah:** طيب... انتي هنا شنو يسوي؟ قصص بصل..سوي مويه حار..كلام انتي

**S1:** امممم... بصل..سوي..**rice**... امممم ..دجاج سوي ...قهوه ..شاي..سوي

**Najah:** شلون سوي رز؟

S1: امممم..قدر مويه...@@

Najah: طيب كيف سوي اكل في سيرلنكا؟ كلام

S1: Raice ..طماطم ..جزر...بقدونس..بطاطا...كلو فيه اكل ..سيرلنكا اكل .. لحم no .. دجاج

Najah; ليش مافيه لحم؟

S1: No like

Najah: عشان فلوس كثير.. ولا ليش مافيه لحم

S1: لا..لا.. my family no like

Najah: انتي هنا في سعوديه فكر فيه روحي سيرلنكا بعدين؟

S1: روحي؟ انا سعودي روحي سيرلنكا؟

Najah: ايه يعني كم سنه في سعوديه بعدين روحي سيرلنكا؟

S1: اه two years روح سيرلنكا

Najah: ليش؟

S1: انا روحي سعوديه two years خلاص سيرلنكا .. خلاص

Najah: انتي يحب اللعاب play games لما كنتي بيبي صغير

S1: انا... play…hobbies..football..كوره

Najah: يلعب واجد كوره؟ كيف

S1: Netball كويس ... .like امممم

Najah: طيب فيه اخت او اخ.. في عربي كلام اختي يعني sister واخوي يعني brother

S1: اه .. واحد ولد ...بنتي

Najah: لا .. مافيه كلام بنتي هذا اختي

S1: اختي… seven كلو my family nine

Najah: ماشاءالله كلو اخ واخت في سيرلنكا

S1: كلو سيرلنكا ... no اثنين كويت ..كلو سيرلنكا.. واحد موت صغير اختي

Najah: طيب انتي هنا في سعوديه ممكن انتي فيه خوف ؟

S1: خوف؟

Najah: ايوه يعني فكر واجد .. واجد فيه بنت ولد؟

S1: كلو good كلو كويس

Najah ماشاءالله كويس..طيب الحين سؤال ثاني نفس سؤال صديق ..الحين انا ابي اوريك صور وانتي قولي ايش هذا صوره مين موجود في الصوره سمي شكل موجود داخل صوره .. هذا ايش وهذا ايش وكيف نستخدم administrative pronouns اسماء الاشاره بالعربي.. يعني I want you to name the objects that you see in the pictures وكيف قول هذا شي اذا كان بعيد او قريب.. اوك بس شوف صوره وتكلمي

S1: 1- بنت واحد

2- سياره اثنين

3- كاميرا واحد

4- واحد ولد

5- كوره اثنين

6- ولد ثلاثه

7- قلم واحد

8- واحد ساعه

9- ثنين بنت

**10-ثنين بنت**

**11- واحد ولد**

**12- واحد ولد**

**13- ثنين بيت**

**14- واحد بنت**

**15-ثنين كبير بنت**

**Najah: طيب الحين انتي شوفي هذا صوره وقولي وين كلب موجود داخل او برا صندوق او وين .. هذا اسمه كلب . وهذا اسمه صندوق.. انتي كلام وين موجود كلب وكيف نستخدم propositions حروف الجر بالعربي**

**S1: ok**

**1- كلب داخل صندوق**

**2- كلب مافيه داخل..فوق صندوق**

**3- كلب صندوق برا**

**4- كلب صندوق تحت**

**5- كلب صندوق برا..قدام**

**6- كلب صندوق @@مافيه معلوم**

**7- كلب صندوق ثنين مافيه معلوم**

**8- كلب صندوق..مافيه معلوم..كلب صندوق اربع وسط**

**9- كلب صندوق فوق**

**Najah: طيب انتي قول كلب صندوق فوق هنا عربي كلام كلب فوق صندوق ..في سيرلنكا كيف كلام زي كذا**

**S1: ايوه سيرلنكا كلب බල්ලා - ballā صندوق peTTiya පෙට්ටිය**
**فوق uDin උඩින්**
**طيب قولي كلو هذا كلام كيف بسيرلنكا ballā; peTTiya uDin**
**පෙට්ටිය බල්ලාඋඩින්**

**Najah: ايوه يعني باعربي قول ballā uDin peTTiya**

**طيب شكرا شكر thank you so much for your help and cooperation ومع السلامه**

**Interviewee: S2**
*Sinhalese Long-term resident*

**Najah** السلام عليكم

**S2** وعليكم سلام

**Najah** كيف حال .. كويس

**S2** الحمدلله

**Najah** طيب انتي من وين؟

**S2** سرلينكا

**Najah** من وين من سرلينكا؟ يعني ايش مدينه؟

**S2** كيرلا

**Najah** طيب لغه في سرلنكا؟ انا هنا سعوديه كلم عربي انتي في سرلنكا ايش كلام؟

**S2** سنهلا Sinhala

**Najah** طيب هذا لغه اول يعني لغه واحد فيه لغه ثاني انتي كلام غير سنهالا ولا مافيه

**S2** لا.. مافيه

**Najah** طيب انتي ايش يشتغل هنا؟

**S2** نظافه

**Najah** نظافه .. ايش يسوي ؟

**S2** امممم .. حمام .. غسيل..دكتور..عياده..عياده غسيل..كنسي..امممم.. شاهي.. زباله شيلي

**Najah** طيب كم سنه انتي هنا في سعوديه؟

**S2** امم .. اربعه سنه خلاص..امممم..من مستشفى اربعه نه خلاص خلاص ثلاثه شهر .. بيت شغل ثمانيه سنه

**Najah** طيب كلو كم سنه في سعوديه؟

**S2** كلو ثمانيه سنه بعدين..خمسه سنه .. ثلاثطعش سنه

**Najah** طيب انتي اول مره يجي سعوديه فيه شغل قبل مثل كويت ولا دبي؟

**S2** كويت .. ثنين سنه

**Najah** ايوه وبعدين وين شغل؟

**S2** كويت اثنين سنه .. هنا سعودي ثمانيه سنه ..خدامه

**Najah** ماشاءالله وبعدين وين شغل؟

**S2** خمسه سنه مستشفى

**Najah** طيب انتي اول مره يروح كويت فيه معلوم عربي قبل يروح ولا مافيه معلوم؟

**S2** معلوم..معلوم

**Najah** كيف معلوم؟

**S2** اول يجي سعوديه..اثنين سنه خلاص ..روحي كويت

**Najah** اه .. طيب لما يجي سعوديه اول مره هنا كيف تعلم عربي؟

**S2** عربي .. واجد .. مافيه معلوم..بيت..روحي بيت .. ماما .. بابا..ولد .. بنت ..كلام ..اسمع بعدين اجي كويس

**Najah** ايوه..يعني انتي مافيه درس عربي اول في سرلينكا

**S2** لا.. مافيه.. اكل معلوم..شراب مافيه معلوم..ماما .. بابا.. كلام اسمع بعدين كلو معلوم @@

**Najah** ايوه..طيب انتي الحينن كويس عربي.. ممتاز ماشاءالله لما يجي صديق جديد من سرلينكا يبغى شغل عهنا في سعوديه اول مره .. انتي كلام حق صديق انتى لا يجي هنا سعوديه عاى طول لازم وي عربي تعلم عربي اول بعدين يجي ولا مافيه مشكله هو يجي هنا بعدين ممكن يتعلم عربي بسرعه؟

**S2** اه .. مافيه مشكله يجي.. قرقر.. بعدين معلوم ..يتعلم عربي

**Najah** يعني مافيه مشكله حتى لو مافيه عربي هو يجي عنا بعدين يتعلم سوا صديق كيف يتكلم عربي هنا

**S2** ايه .. كذا احسن

**Najah** طيب انتي لما فيه سرلنكا فيه study دراسه داخل مدرسه ولا مافيه؟

**S2** ايه مدرسه

**Najah** وشو ادرس؟

**S2** اه .. عشره سنه

**Najah** وبعدين؟

**S2** بعدين خلاص..خلي بيت..بعدين جواز خلاص.. يجي هنا

**Najah** كم عمرك انتي؟

**S2** انا اثنين خمسين سنه

**Najah** ماشاءالله .. طيب انتي فيه زوج واولاد؟

**S2** ايه فيه بنت واحد

**Najah** كم عمرو

**S2** بنتي خمسه عشرين

**Najah** ماشاءالله فيه زوج بنت؟ ولا مافيه؟

**S2** بنت جواز

**Najah** عندها اطفال؟

**S2** مافيه.. مافيه.. اول سنه جواز يجي ..باقي شويه واحد سنه يجي مافيه بنتي

**Najah** ان شاءالله بعدين فيه ببيي

**S2** ايوه

**Najah** طيب انتي سعودي وين بيت؟

**S2** اول سعودي؟

**Najah** لا .. الحين

**S2** هنا مستشفى ورا ..company

**Najah** طيب مافيه صديق هنا؟

**S2** فيه .. فيه .. بنقلادش

**Najah** لا.. مو مستشفى برا مستشفى

**S2** لا .. مافيه .. برا روحي

**Najah** طيب انتي كلم بنت ؟

**S2** ايوه

**Najah** كيف

**S2** ايوه جوال كلم

**Najah** كم مره كلم .. كثير ولا شويه ؟

**S2** هيا فيه شغل .. مستشفى سرلنكا هي nurse خلاص.. انا كلم .. اسبوع واحد مره كلام

**Najah** حلو .. كويس.. طيب انتي لما يجي يتكلم مع ناس برا مستشفى .. يتكلم عربي ولا بس سينهالا لغه

**S2** برا

**Najah** ايوه

**S2** امممم... سرلنكا نفرات شوفي .. سرلنكا كلام ..زثاني نفرات سوا سوا كلام عربي

**Najah** ايوه.. طيب انتي يتكلم عربي كثير ولا شويه؟

**S2** انا كثير معلوم .. هذا عشره سنه بيت...كويت فيه اثنين سنه .. سعوديه ثمانيه سنه.. كثير معلوم

**Najah** طيب الحين انتي ماشاءالله بعد عشره سنه . انتي يشوف عربي صعب ؟

**S2** لا.. لا.. مافيه صعب ..سرعه سرعه يجي

**Najah** طيب انتي شفتي nurse صديق هذا من فلبين قبل شويه هي مافيه معلوم عربي مره

**S2** هذا جديد

**Najah** ايوه بس هو هنا موجود واحد سنه

**S2** واحد سنه بس مستشفى .. مستشفى مافيه كثير كلام عربي .. كلو انقلشش

**Najah** طيب وين كلام عربي واجد؟

**S2** كلام عربي.. فيه شغل بيت .. كلو عربي .. ماما .. بابا .. بنت

**Najah** طيب الحين سينهالا كلام نفس عربي كلام ولا يختلف؟

**S2** لا .. ثاني.. عربي كلام .. انقليزي كلام .. كلام يجي سرعه.. سرعه .. سرلينكا كلام مايجي سرعه سرعه كلام

**Najah** ايوه يعني كلام عربي صعب

**S2** ايه صعب

**Najah** طيب انتي لما يتكلم عربي يسوي mix مع سرلينكا

**S2** لا

**Najah** طيب فيه كلام عربي نفسو كلام موجود في سرلينكا؟

**S2** فيه شويه مثل صابون نفسو عربي

**Najah** ايوه.. طيب انتي يشوف تلفزيون؟

**S2** الحين لا

**Najah** طيب اول فيه بيت شغل اتني يشوف تلفزيون؟

**S2** ايوه فيه... مسلسل... معلوم

**Najah** كويس مافيه مشكله كلو معلوم

**S2** ايوه

**Najah** يتذكر مسلسل عربي انتي حبو؟

**S2** انا حبي ذهب كثير @@ بس غالي @@

**Najah** @@ طيب يسمع مسيقى عربي؟

**S2** بيت لا.. مافيه.. سعودي بيت مافيه جوال .. على طول شغل شغل كثير.. بس هنا فيه اشتري جوال .. هنا يسمع

**Najah** اه طيب يسمع مسيقى عربي ولا ايش؟

**S2** هندي .. انا كثير حبي هندي

**Najah** طيب انتي فيه فكر واجد؟

**S2** ايه واجد فكر.. بابا موت .. ماما موت..جوز مافيه سوا سوا انا.. بنتي حبي .. انا يجي هيا فيه روح مدرسه .. الحين فيه كويس

**Najah** طيب الحمدلله.. اذا انتي مافيه شغل هنا شنو يسوي ؟

**S2** مافيه شغل هنا مسكينه.. كثير مسكينه.. سرلنكا مافيه فلوس .. فلوس مافيه يجي هنا

**Najah** لا.. لا.. انتي لما يروح بيت عشان نوم ومافيه سوي شي مافيه شغل يعني انتي free فيه وقت فاضي شنو يسوي هذا وقت سوي كتابه..سوي اتصال ..وي نوم

**S2** لا.. فيه كثير شغل ..company ساعه عشره خلاص شغل .. روحي بيت..واحد مره ملابس خياطه .. غسيل..سوي اكل .. كذا شويه شويه شغل مافيه شغل .. مافيه نوم .. سوي نظيف بيت .. حمام نظيف ..برا كنسي نظيف.. كذا شغل فيه

**Najah** كويس كويس..طيب انتي لما كنتي صغير في سرلنكا وشو اللعاب تلعبيها وانتي صغير ؟

**S2** امممم فيه ثاني هذا سرلنكا game مافيه تلفزيون..امم ايش سعودي فيه بيت .. كذا @@

**Najah** اه يعني سوي حجره ونطي @@

**S2** ايوه.. ايوه..@@

**Najah** طيب ايش فيه بعد كمان؟

**S2** بس.. امممم... برا روحي سوا سوا صديق اللعب .. اللعب مويه .. بيت هذا مويه مويه .. روحي تحت اللعب شويه

**Najah** طيب هنا في السعوديه كيف حياه ؟ وكيف حياه في الكويت؟

**S2** كويت كويس.. كويت ..سعودي عبايه هذا سكر على طول @ سعودي ..كويت مافيه .. كذا روحي برا مافيه مشكله

**Najah** طيب .. وشو شي انتي مافيه حبي هنا؟

**S2** هنا.. امم .. مافيه مشكله .. انا سيم سيم شغاله نظافه .. حرمه .. مافيه مشكله

**Najah** طيب الحين يجي عند اكل=

**S2** اكل انا سوي اكل سرلينكا اكل سوي .. سعودي نفرات اعطي.. سعودي اكل انا حبو

**Najah** وشواكل انتي سعودي يحبي مثل ايش؟

**S2** هنا .. اكل دجاج .. سمك...اكل بيتزا .. فطيره..سمبوسه@@

**Najah** وانتي تعرفين تسوين اكل سعودي؟

**S2** اممم .. معلوم هذا سعودي بيت .. رمضان انا سوي سمبوسه .. فطيره سوي

**Najah** طيب وشو يختلف اكل هنا عن اكل كويت ؟ فيه اختلاف ولا مافيه

**S2** كويت.. سعودي.. سيم سيم اكل .. كويس

**Najah** طيب .. والناس سيم ناس سعودي ولا فيه اختلاف؟

**S2** كلو مكان كويس نفرات فيه.. مافيه كويس نفرات فيه

**Najah** صحيح صحيح

**S2** كويت هذا..كويت فيه شغل مكان.. مكان كويس بس ناس زياده مافيه كويس .. شويه كويس ..شويه ..شويه مافيه كويس.. كذا فيه .. كذا مشكله كثير

**Najah** طيب .. انتي شنو فكر سوي في سعوديه عشان بس خلاص روحي سرلنكا

**S2** ايوه.. انا يبغى روحي..الحين انا شويه هذا يمكن سبعه شهر انا خلاص كلو...ثمانيه شهر خلاص

**Najah** اه ماشاءالله ...خلاص مافيه يرجع؟

**S2** لا.. خلاص.. بعدين انا خلاص عجوزه... عمر بنتي ممكن حامل.. مافيه انا شوف@

**Najah** فيه شي انتي فيه خوف هنا؟

**S2** مافيه خوف .. ايوه

**Najah** مافيه سوي حادث مافيه مريض؟

**S2** ايوه

**Najah** فيه اول؟

**S2** لا.. بس هذا برد .. كحه يجي بس

**Najah** طيب انتي فكر خوف بيت مشكله سرلنكا؟

**S2** لا ماغيه خوف فكر

**Najah** مافيه خوف كيف انا يجيب فلوس؟

**S2** اول انا ودي فلوس..الحين شغل فلوس خلي انا سوا سوا روحي

**Najah** اه يعنيي معاكي خبي فلوس بعدين روحي هناك .. طيب ايش يسوي فيه فلوس اذا رحتي سرلنكا.

**S2** اممم.. مافيه بيت .. سوي بيت ابغى .. انا جوز سوا سوا مافيه

**Najah** ايش ؟

**S2** مافيه جوز انا .. طلق خلاص

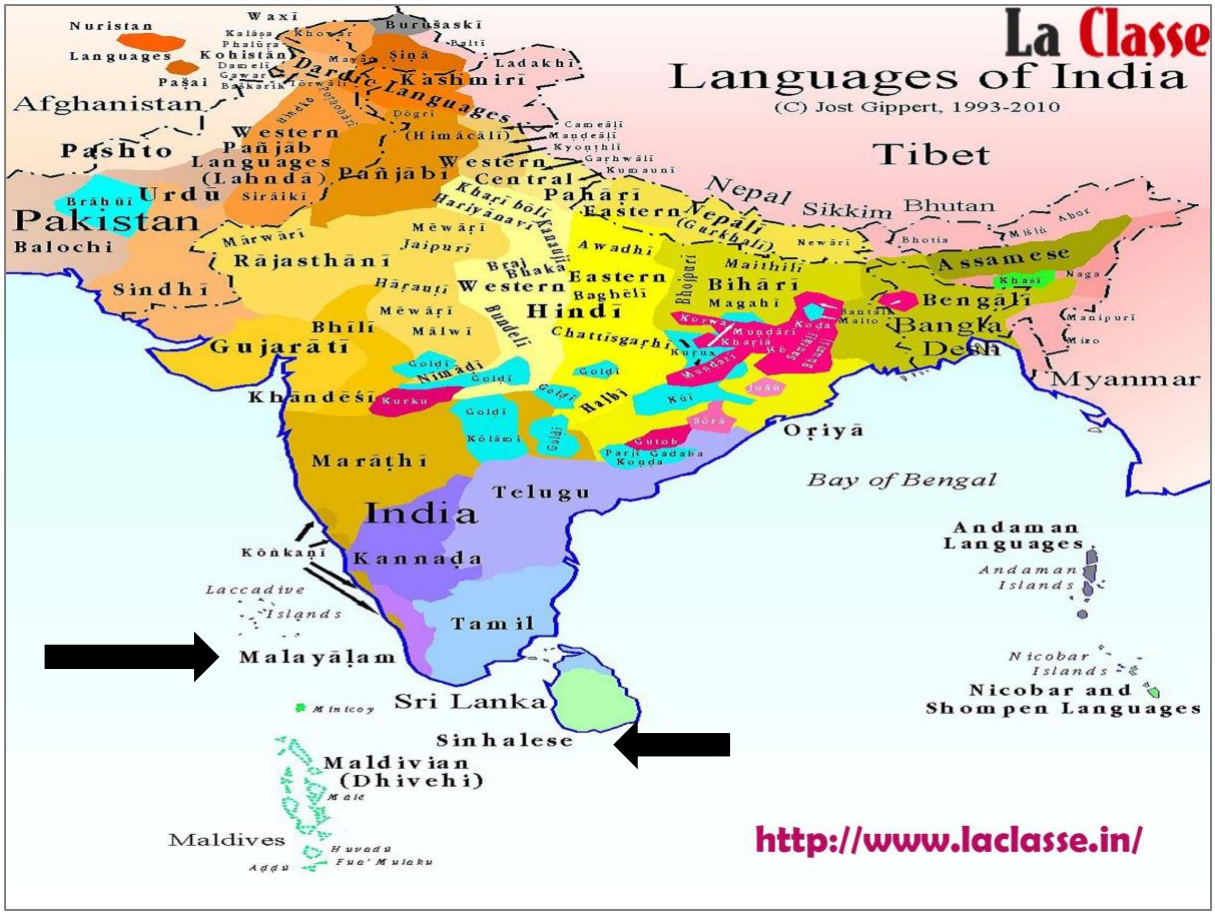
**Najah** اه مافيه زوج

**S2** ايوه.. بعدين فيه شويه فلوس سوي بيت صغير . بعدين يقعدي هنا...مافيه كبير ..شويه صفير بيت سوي

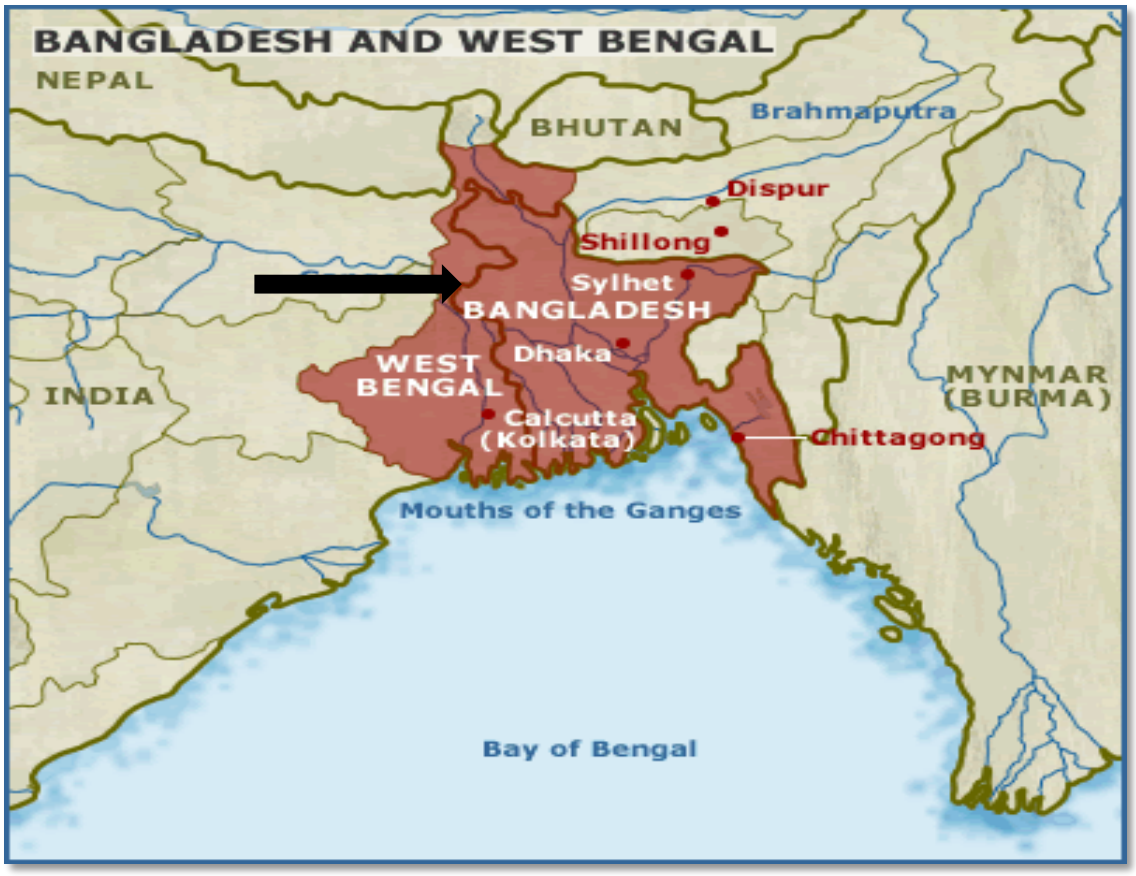
**Najah** ماشاءالله كويس..طيب الحين سؤال ثاني نفس سؤال صديق . وشو هذا صوره

## Appendix D. Maps


Map 1: Punjabi, Pakistan Language.




Map 2: Malayalam, Kerala Language and Sinhala, Sri Lanka Lanaguage.




Map 3: Bengali, Bangladesh Language


Map 4: Sunda, Indonesia Language.


Map 5: Tagalog, Philippines Language.

## Appendix E. Ethics Approval for the Research Project

Faculty of Arts: Ethics Committee
George Wallis Building
Wulfruna Street
Wolverhampton
WV1 1DT

20 February 2019

**Ethics Approval Application Number.** 2018/19: 09

**Researcher.** Najah Albaqwi

**Level of Research.** PhD
**Director of Studies.** Dr. Michael Oakes

**Title of Research.** Gender Variation in Gulf Pidgin Arabic

**Decision.** Approved

Dear Najah
Thank you for your very full response to the issues raised by the Faculty Ethics Committee. We are pleased to inform that we can now approve your Ethical Approval Application.

Please ensure that you are conversant with the latest guidelines on recruiting research participants and data security. See the Ethics Guidance web pages
https://www.wlv.ac.uk/research/about-our-research/policies-and-ethics/ethics-guidance

If you make any substantial changes to your research, you will have to complete a new request for ethical approval. This letter only relates to ethical issues, and has no bearing on other aspects of your research, such as methodology and theoretical framework.

Please do not hesitate to contact the Faculty Ethics Committee if you have any questions. We wish you the very best with your research

*Dr Stephen Jacobs*
Chair of the Faculty of Arts Ethics Committee
MK507, George Wallis Building
[E-mail address redacted]

## References


Aarts, B. (2000). Corpus linguistics, Chomsky and fuzzy tree fragments. *Language and Computers*, 33, 5-14.

Abdel-Jawad, H. R. E. (1981). *Lexical and phonological variation in spoken Arabic in Amman.* (Doctoral thesis). University of Pennsylvania.

Abdeljawad, I., Mat-Nor, F., Ibrahim, I., & Abdul-Rahim, R. (2013). Dynamic capital structure trade-off theory: Evidence from Malaysia. *International Review of Business Research Papers*, 9(6), 102-110.

Abed, L. G. (2018). *Gulf Pidgin Arabic: An Empirical Investigation into its Emergence and Social Attitude* (Doctoral dissertation). Trinity College Dublin.

Abernethy, M. (2015). Self-reports and observer reports as data generation methods: An assessment of issues of both methods. *Universal Journal of Psychology*, 3(1), 22-27.

Abu-Haidar, F. (1924). Male/female linguistic variation in a Baghdad community. *A Miscellany of Middle Eastern Aricles. In Memoriam Thomas Muir Johnstone, 83*.

Abu-Haidar, F. (1989). Are Iraqi women more prestige conscious than men? Sex differentiation in Baghdadi Arabic. *Language in Society, 18*(4), 471-481.

Adolphs, S., & Carter, R. (2013). *Spoken corpus linguistics: From monomodal to multimodal*. Routledge.

Ahmed, L. (1992). *Women and gender in Islam: Historical roots of a modern debate*. Yale University Press.

Al Lily, A. E. (2011). On line and under veil: Technology-facilitated communication and Saudi female experience within academia. *Technology in Society*, 33(1-2), 119-127.

Al-Ageel, H. (2015) Requests in Saudi Pidgin Arabic. *Business Management and Strategy*, 6(1), 111-137.

Al-Azraqi, M. (2010). Pidginization in the eastern region of Saudi Arabia: Media presentation. In R. Bassiouney (Ed.), Arabic and the Media. Linguistic Analyses and Applications, (159– 174). *Leiden, Brill.*

Al-Hathloul, S. (2017). Riyadh Development Plans in the Past Fifty Years (1967-2016). *Current Urban Studies*, 5(01), 97.

Al-Mubarrad, A. A. A. M. (1864). Ibn YazId, al-Kdmil, ed. *W. Wright, Leipzig*, *92*.

Al-Muhannadi, M. (1991). *A sociolinguistic study of women's speech in Qatar*, (Doctoral dissertation). University of Essex.

Al-Sobh, M. A., Abu-Melhim, A. R. H., and Bani-Hani, N. A. (2015). Diglossia as a result of language variation in Arabic: Possible solutions in light of language planning. *Journal of Language Teaching and Research,* 6(2), 274-279.

Al-Wer, E. (1991). *Phonological variation in the speech of women from three urban areas in Jordan*, (Doctoral dissertation). University of Essex, Colchester.

Al-Zubeiry, H. Y. A. (2015). Linguistic analysis of Saudi pidginized Arabic as produced by Asian foreign expatriates. *International Journal of Applied Linguistics and English Literature*, *4*(2), 47-53.

Albakrawi, H. (2012). The linguistic effect of foreign Asian workers on the Arabic Pidgin in Saudi Arabia. *Language*, 2(9).

Albaqawi, N. S. (2016). Unity and diversity within pidginized Arabic as produced by Asian migrant workers in the Arabian Gulf, *Journal of Second Language Teaching and Research*. Volume 5, Special Issue, Wolverhampton University https://pdfs.semanticscholar.org/4727/9557bd5d30fc16939172994abb65c08bc5c6.pdf

Albaqawi, N., & Oakes, M. (2019). Compiling and Analysing a Corpus of Transcribed Spoken Gulf Pidgin Arabic Based on Length of Stay in the Gulf. In *Proceedings of the 3rd Workshop on Arabic Corpus Linguistics* (pp. 7-15).

Alfaifi, A., and Atwell, E. (2013). Arabic Learner Corpus v1: A New Resource for Arabic Language Research. In proceedings of the Second Workshop on Arabic Corpus Linguistics (WACL-2). Lancaster University, UK.

Alghamdi, E. A. (2014). Gulf Pidgin Arabic: A Descriptive and Statistical Analysis of Stability. *International Journal of Linguistics*, 6(6), 110.

Alkohlani, F. A. (2016). The problematic issue of grammatical gender in Arabic as a foreign language. *Journal of Language and Cultural Education*, 4(1), 17-28.

Alleyne, M. C. (1971). Acculturation and the cultural matrix of creolization. *Pidginization and creolization of languages*, 169-186.

Alleyne, M. C. (1980). *Comparative Afro-American: an historical-comparative study of English-based Afro-American dialects of the New World* (Vol. 11). Karoma Publishers.

Almoaily, M. (2008). *A data-based description of Urdu Pidgin Arabic*, (Master thesis). Newcastle University.

Almoaily, M. (2012). *Language variation in Gulf Pidgin Arabic*, (Doctoral thesis). Newcastle University.

Alotaibi, A. S., (2018). *The Copula in Arabic: Description and Analysis*, (Doctoral dissertation). University of Essex.

Alotaiby, F., Alkharashi, I., & Foda, S. (2009). Processing large Arabic text corpora: Preliminary analysis and results. In *Proceedings of the second international conference on Arabic language resources and tools* (pp. 78-82).

AlQahtani, S. J. (2016). *The Structure and Distribution of Determiner Phrases in Arabic: Standard Arabic and Saudi Dialects*, (Doctoral dissertation). Université d'Ottawa/University of Ottawa.

Alqasem, M. F. M. (2017). Limitation of Feminine Language in a Society and its Effect on the Foreigner's Speech during Acquiring Language Aspects (Saudi Arabic Pidgin as a Model).

Alshammari, W. F. (2010). An Investigation into Morpho-syntactic Simplification in the Structure of Arabic Based Pidgin in Saudi Arabia. *Aman, Mu'tah University: MA Dissertation.*

Alshammari, W. F. B. (2018). *The Development of and Accommodation in Gulf Pidgin Arabic: Verbal and Pronominal Form Selection*, (Doctoral dissertation). Indiana University.

Andor, J. (2004). The master and his performance: An interview with Noam Chomsky. *Intercultural Pragmatics*, 1(1), 93-111.

Ansaldo, U., and Matthews, S. (2004). The Origins of Macanese Reduplication. In G. EscureandA. In Geneviève Escure and Armin Schwegler, eds. *Creoles, Contact and Language Change. Linguistic and Social Implications*. Amsterdam/ Philadelphia: John Benjamins. 1–20.

Ansaldo, U., Matthews S., and Smith, G. (2011). The Cantonese substrate in China Coast Pidgin. *Creoles, their Substrates, and Language Typology. Amsterdam: Benjamins*, 289-301.

Aoun, J. E., Benmamoun, E., and Choueiri, L. (2010). *The syntax of Arabic*. Cambridge University Press.

Aparnna, T., Raji, P. G., & Soman, K. P. (2010). Integer linear programming approach to dependency parsing for Malayalam. In *2010 International Conference on Recent Trends in Information, Telecommunication and Computing* (pp. 324-326). IEEE.

Arends, J., Muysken, P., & Smith, N. (Eds.). (1995). *Pidgins and creoles: An introduction* (Vol. 15). John Benjamins Publishing.

Asher, R. E. (1968). Existential, possessive, locative and copulative sentences in Malayalam. In *The Verb 'Be'and Its Synonyms* (pp. 88-111). Springer Netherlands.

Asher, R. E. and Kumari, T. C. (1997). *Malayalam*. Psychology Press.

Asis, M. M., and Piper, N. (2008). Researching international labor migration in Asia. *The sociological quarterly*, 49(3), 423-444.

Aslan, O. (2009). The role of gender and language learning strategies in learning English. *Unpublished Master's thesis.*

Avram, A. (2011). Pseudo-Reduplication, reduplication and repetition: The case of Arabic-lexified pidgins and creoles. *RRL, LVI*, 3, 225-256.

Avram, A. (2012). On the functions of FI in the verbal system of Arabic pidgins. *Romano-Arabica*, *12*, 35-58.

Avram, A. (2013). Multifunctionality in Gulf Pidgin Arabic: The case of Fi. *Linguistics in the Gulf, 4*.

Avram, A. (2014). Immigrant workers and language formation: Gulf Pidgin Arabic.

Avram, A. (2015). On the developmental stage of Gulf Pidgin Arabic. *In Arabic varieties: Far and wide. Proceedings of the 11th International Conference of AIDA* (pp. 87-98).

Avram, A. (2017). On the periphery of the periphery: Gulf Pidgin Arabic. *Lisan Al-Arab*, 9, 61.

Avram, A. (2017). Superdiversity in the Gulf: Gulf Pidgin Arabic and Arabic Foreigner Talk. *Philologica Jassyensia*, 13(2/26), 175-190.

Avram, A. (2018). On the relationship between Arabic Foreigner Talk and Pidgin Arabic. *Arabic in Contact*, *6*, 251.

Baayen, R.H. and Schreuder, R. (2003) *Morphological structure in language processing. Berlin* (Vol. 151). Mouton de Gruyter.

Babbie, E. R. (2015). *The practice of social research*. Nelson Education.

Bacon, S. M., and Finnemann, M. D. (1992). Sex differences in self-reported beliefs about foreign-language learning and authentic oral and written input. *Language learning*, 42(4), 471-495.

Badawi, E. S. M. (Ed.). (1996). *Understanding Arabic: essays in contemporary Arabic linguistics in honor of El-Said Badawi.* American Univ in Cairo Press.

Badran, M. (1995). *Feminist, Islam and nation. Gender and the making of Modern Egypt*. Princeton University Press.

Bae, C. H. (2004). Aramaic as a Lingua Franca during the Persian Empire (538-333 BCE). *Journal of Universal Language*, 5(1), 1-20.

Bahloul, M. (1993). The Copula in Modern Standard Arabic (SA). *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 209-209.

Bailey, B. L. (1966). *Jamaican creole syntax*. Cambridge University Press.

Bailey, G., Wikle, T., & Tillery, J. (1997). The effects of methods on results in dialectology. *English World-Wide*, 18(1), 35-63.

Baker, P. (1996). The potential for the development of Arabic-based and other contact languages along the maritime trade routes between the Middle East and China, from the start of the Christian era. de Gruyter Mouton.

Baker, P. (2010). *Sosiolinguistic and corpus linguistics.* Edinburgh University Press.

Baki, R., (2004). Gender-segregated education in Saudi Arabia: Its impact on social norms and the Saudi labor market. *Education Policy Analysis Archives*, 12(28).

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. (2003). Pidgin inflectional morphology and its implications for creole morphology. In *Yearbook of Morphology 2002* (pp. 3-33). Springer, Dordrecht.

Bakker, P. (2011). Unpublished review of the book Une grammaire de la servitude, by Bizri, F (2010) Librairie Orientaliste Paul Geuthner, p. 280. In 2. Almoaily, M. (2012) *Language Variation in Gulf Pidgin Arabic*, (doctoral thesis). Newcastle University.

Bakker, P., & Matras, Y. (Eds.). (2013). Contact languages: *A comprehensive guide* (Vol. 6). Walter de Gruyter.

Bakker, P., & Muysken, P. (1995). Mixed languages and language intertwining. *Pidgins and creoles: An introduction*, 41-52.

Barnbrook, G. (1996). Language and computers: A practical introduction to the computer analysis of language. *Edinburgh Textbooks in Empirical Linguistics*.

Bassiouney, R. (2009). Arabic sociolinguistics. Edinburgh University Press.

Basu, S. (2017). Household labor supply and intermarriage of immigrants: differences by gender. *IZA Journal of Development and Migration*, *7*(1), 8.

Bauer, L. (1926). Das palestinische Arabisch. *Leipzig: J. C. Heinrichs*.

Baumert, J. (1992). Koedukation Oder Geschlechtertrennung. *Zeitschrift für pädagogik*, 38(1), 83-110.

Bayley, R., & Preston, D. R. (Eds.). (1996). *Second language acquisition and linguistic variation* (Vol. 10). John Benjamins Publishing.

Becker, A., & Veenstra, T. (2003). Creole prototypes as basic varieties and inflectional morphology. *Information structure and the dynamics of language acquisition*, 26, 235.

Beeching, K. (2006). Synchronic and diachronic variation: the how and why of sociolinguistic corpora. In *Corpus linguistics around the world* (pp. 49-61). Brill Rodopi.

Bell, A., & Holmes, J. (Eds.). (1990). *New Zealand ways of speaking English* (Vol. 65). Multilingual Matters.

Benmamoun, E. (2008). Clause structure and the syntax of verbless sentences. *Current Studies in Linguistics Series*, *45*, 105.

Bennett, G. R. (2010). An introduction to corpus linguistics. *Using corpora in the language learning classroom: corpus linguistics for teachers*, 4-22.

Berg, B. L., Lune, H., & Lune, H. (2004). *Qualitative research methods for the social sciences* (Vol. 5). Boston, MA: Pearson.

Bhatia, T. (2013). *Punjabi*. Routledge.

Biber, D, Conrad. S. and Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Bickerton, D. (1981). Roots of Language.

Bickerton, D. (1984). The language bioprogram hypothesis. *Behavorial and Brain Sciences,* 7:173-221.

Bickerton, D. (1990). Species and Language. Chicago, IL: Chicago University Press.

Birkner, V. (2015). The advantages and disadvantages of employing corpus evidence in sociolinguistic studies. The Teacher Magazine nr. 2 (126). *Retrieved 11 September 2016* from:https://www.academia.edu/11326263/Advantages_and_Disadvantages_of_employing_corpus_evidence_in_sociolinguistic_studies

Bizri, F. (2005). The Pidgin Madam, a new Arabic pidgin. *LINGUISTIQUE*, *41*(2), 53-66.

Bizri, F. (2010). *Pidgin Madame. Une grammaire de la servitude*. Paris: Paul Geuthner

Bizri, F. (2014). Unity and diversity across Asian migrant Arabic pidgins in the Middle East. *Journal of Pidgin and Creole Languages*, 29(2), 385-409.

Blackman, G. L., Ostrander, R., & Herman, K. C. (2005). Children with ADHD and depression: a multisource, multimethod assessment of clinical, social, and academic functioning. *Journal of Attention Disorders*, 8(4), 195-207.

Blanc, H. (1970). Dual and pseudo-dual in the Arabic dialects. *Language*, 42-57.

Bloomfield, L. (1933). Language History, dir. *H. Hoijer, New York, Holt, Rinehart et Winston*.

Bolozky, S., & Haydar, A. F. (1986). Colloquial gender neutralization in the numeral systems of Modern Hebrew and Lebanese Arabic. *al-'Arabiyya*, 19-28.

Boucherit, A., & Lentin, J. (1989). Les dialectes féminins dans le monde arabe: des dialectes minoritaires et leur évolution. *Linx*, *21*(1), 17-37.

Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. Routledge.

Bresnan, J. (2000). Pidgin genesis and optimality theory. *Processes of language contact: Case studies from Australia and the Pacific*, 145-173.

British Educational Research Association. (2004). *Revised ethical guidelines for educational research.* Bera.

Brown, K., & Ogilvie, S. (2010). *Concise encyclopedia of languages of the world*. Elsevier.

Brustad, K. (2000). The syntax of spoken Arabic: A comparative study of Moroccan. *Egyptian, Syrian, and Kuwaiti Dialects*.

Bucholtz, M., & Du Bois, J. W. (2006). Transcription in action: Resources for the representation of linguistic interaction. *Retrieved October*, *12*, 2016 from: http://www.linguistics.ucsb.edu/projects/transcription/

Buchstaller, I. (1999). Causality, Conditionalit and Concessivity in Hawaiian Creole. *Unpublished MA Thesis*. Konstanz University.

Buchstaller, I., & Khattab, G. (2013). Population samples. *Research methods in linguistics*, 74-95.

Buchstaller, I., & Khattab, G. (2014). Population Samples. In R. J. Podesva, & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 74–95). Cambridge: Cambridge University Press.

Bush-Caver, H., & Williams, M. T. (2009). Creoles. *Retrieved January*, *12*, 2017 from: https://www.everyculture.com/multi/Bu-Dr/Creoles.html

Caldwell, R. (1875). *A comparative grammar of the Dravidian or South-Indian family of languages*. Trübner.

Chambers, J. K., Trudgill, P., & Schilling-Estes, N. (2004). The Handbook of Language Variation and Change (2nd). *Victoria: Blackwell Publishing*.

Cheng, L. L. S., and Corver, N. (Eds.), (2013). *Diagnosing syntax* (Vol. 46). Oxford University Press.

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

Chetrit, J. (1986). Stratégies discursives dans la langue des femmes judéoarabophones du Maroc. *Massorot, 2*, 41-66.

Chiswick, B. R., & Houseworth, C. (2011). Ethnic intermarriage among immigrants: Human capital and assortative mating. *Review of Economics of the Household*, *9*(2), 149-180.

Chomsky, N. (1965). Aspects of the theory of syntax Cambridge. *Multilingual Matters: MIT Press.*

Coelho, F. A. (1881). *Os dialectos românicos ou neo-latinos na África, Ásia e América*. Sociedade de Geographia de Lisboa. In Almoaily, M. (2012). *Language variation in Gulf Pidgin Arabic*, (doctoral thesis). Newcastle University.

Combiths, P. N., Barlow, J. A., and Sanchez, E. (2019). Quantifying phonological knowledge in children with phonological disorder. *Clinical linguistics & phonetics*, *33*(10-11), 885-898.

Computing, R. (2013). R: A language and environment for statistical computing. *Vienna: R Core Team.* http://www.R-project.org/, pge: 217.

Coon, J. (2013). Predication, tenselessness, and what it takes to be a verb. In *Proceedings of the 43rd annual meeting of the North East Linguistics Society*.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. (3nd ed.). Thousand Oak, CA: SAGE Publication.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Crowley, T. (2008). Pidgin and creole morphology. *The handbook of pidgin and creole studies*, 74-97.

Dasgupta, P. (2003). Bangla. In Cardona, G. Dhanesh Jain, eds. 2003. *The Indo-Aryan Languages*. 351–390. London: Routledge.

Dasgupta, P. (2007). Bangla//D. Jain, G. Cardona (eds.). Indo-Aryan languages. Abingdon.

DE BEL-AIR, F. (2014). *Demography, migration and labour market in Saudi Arabia*. Retrieved from https://cadmus.eui.eu/handle/1814/32151

De Granda, G. (1968). La tipología «criolla» de dos hablas del área lingüística hispánica. *Thesaurus*, *1*(2), 193-205.

De Roeck, A., & Al-Fares, W. (2000, October). A morphologically sensitive clustering algorithm for identifying Arabic roots. *In Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics* (pp. 199-206).

De Vaus, D. (2001). *Research design in social research*. Sage.

DeCamp, D. (1971). Toward a generative analysis of a post-creole speech continuum. *Pidginization and creolization of languages*, *347*, 370.

DeGraff, M. (1999). *Creolization, language change and language acquisition: An epilogue*. na. In Thurgood, G. (2005). DeGraff, ed.: Language creation and language change. *Word: Journal of the International Linguistic Association*, *56*(2), 312-316.

Dekkak, M. (1979). *Sex dialect in Tlemcen: an Algerian urban community*, (Doctoral dissertation), School of Oriental and African Studies. University of London.

DeVotta, N. (2004). *Blowback: Linguistic nationalism, institutional decay, and ethnic conflict in Sri Lanka*. Stanford University Press.

Dharmadāsa, K. E. Ō. (1992). *Language, religion, and ethnic assertiveness: the growth of Sinhalese nationalism in Sri Lanka*. University of Michigan Press.

Dörnyei, Z. (2007). *Research methods in Applied Linguistics*. Oxford, UK: Oxford University Press.

Doumato, E. A., & Posusney, M. P. (eds.). (2003). *Women and globalization in the Arab Middle East: gender, economy, and society*. Lynne Rienner Publishers.

Drechsel, E. J. (1996). Native American contact languages of the contiguous United States. *Stephen. A. Wurm, Peter Mühlhaüser & Darrell T. Tryon (eds.), Atlas of Intercultural Communication in the Pacific, Asia, and the Americas*, 2, 1213-1239.

Dresch, P. (2006). Foreign matter: the place of strangers in Gulf society. In *Globalization and the Gulf* (pp. 210-232). Routledge.

Dresser, N. (2011). *Multicultural manners: Essential rules of etiquette for the 21st century*. John Wiley & Sons.

Dryer, M. (2011) Order of object and verb. In Haspelmath, M., Dryer, M., Gil, D. and Comrie, B. (eds.). *The World Atlas of Language Structures*. Munich: Max Planck Digital Library. Available online at https://wals.info

Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual review of anthropology*, *21*(1), 461-488.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language learning*, *27*(2), 315-330.

Eisenstein, M. (1982). A study of social variation in adult second language aquisition. *Language Learning*, *32*(2), 367-391.

Eklund, R. (1996). Derek Bickerton's bioprogram: a proposal and its critics'. *The Creolist Archives: Papers On-line*. Retrieved from: https://pdfs.semanticscholar.org/f9b5/3ae40b5b5ff64603594bdd28bf21d1cf59a3.pdf.

Eubank, L. (Ed.). (1991). Point counterpoint: *Universal Grammar in the second language* (Vol. 3). John Benjamins Publishing.

Evans, D. (2007). Corpus building and investigation for the Humanities. University of Nottingham http://www. corpus. bham. ac. uk/corpus-building. Shtml.

Facey, W. (1992). *Riyadh, the old city: from its origins until the 1950s*. Immel Publishing.

Faraclas, N. (1988). Nigerian Pidgin and the languages of Southern Nigeria. *Journal of Pidgin and Creole languages*, 3(2), 177-197.

Farhady, H. (1982). Measures of language proficiency from the learner's perspective. *TESOL quarterly*, *16*(1), 43-59.

Feagin, C. (2013). Entering the Community. *The handbook of language variation and change*, 17-37.

Feery, K. (2008). Current perspectives on the role of gender in second language acquisition (SLA) research. *The ITB Journal*, *9*(1), 4.

Feghali, H. J. (2004). *Gulf Arabic: the dialects of Riyadh and eastern Saudi Arabia: grammar, dialogues, and lexicon*. Dunwoody Press.

Ferguson, C. (1959). Diglossia. *Word*, *15*(2), 325-340.

Ferguson, C. (1971). Absence of copula and the notion of simplicity. *Pidginization and creolization of languages*, 141-150.

Ferguson, C. (1991). Diglossia revisited. *Southwest Journal of Linguistics*, *10*(1), 214-234.

Ferguson, C. (1991). Individual and social in language change: Diachronic changes in politeness agreement in forms of address. *The Influence of Language on Culture and Thought: Essays in Honour of Joshua A. Fishman's Sixty-fifth Birthday. Berlin: Walter de Gruyter*, 183-197.

Ferguson, C. (1996). Epilogue: Diglossia revisited. In *Contemporary Arabic Linguistics in Honor of El-Said Badawi*. The American University in Cairo, 49-67.

Field, F. (2004). Second language acquisition in creole genesis. *CREOLE LANGUAGE LIBRARY*, *27*, 127-160.

Finch, S. (2001). Copular elements in Bengali and the stage/individual level distinction. *21st South Asian Languages Analysis Roundtable, Un t iversity of Konstanz*.

Firestone, S. (1970). The Dialectic of Sex: The Case for Feminist Revolution, New York: William Morrow & Co.

Fishman, P. (1983). Interaction: the work women do. Language gender and society, ed. by B. Thorne, C. Kramarae & N. Henly, 89-102.

Flowerdew, J. (2012). Corpora in language teaching from the perspective of English as an international language. *Principles and practices for teaching English as an international language*, 226-243.

Foulkes, P., & Docherty, G. (2007*). Phonological variation in England*. na.

Fox, J. W., Mourtada-Sabbah, N., & Al Mutawa, M. (eds.). (2006). *Globalization and the Gulf*. Routledge. Retrived from: http://countrystudies.us/saudi0arabia/21htm

Fraenkel, J. R. & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). New York: NY: McGraw-Hill.

Freed, A. F. (2003). Epilogue: Reflections on language and gender research. *The handbook of language and gender*, 699.

Freeze, R. (1992). Existentials and other locatives. *Language*, 553-595.

Fries, C. C. (1952). The structure of English. *New York: Harcourt Brace.*

Friginal, E., & Hardy, J. (2013). *Corpus-based sociolinguistics: A guide for students*. Routledge.

Froehlich, H. (2015). Corpus analysis with AntConc. *Programming Historian*.

Furtado, D. (2012). Human capital and interethnic marriage decisions. *Economic inquiry*, *50*(1), 82-93.

Gair, J. W. (1968). Sinhalese diglossia. *Anthropological Linguistics*, 1-15.

Gair, J. W., & Karunatilaka, W. S. (1974). Literary Sinhala. Ithaca, New York: South Asia Program, Cornell University.

Gandin, S. (2014). Investigating loan words and expressions in tourism discourse: A corpus driven analysis on the BBC travel corpus. *European Scientific Journal*, *10*(2).

Gass, S. M., & Selinker, L. (2008). Second Language Acquisition: An Introductory Course, Routledge Taylor and Francis Group.

Gast, V., & Diessel, H. (Eds.). (2012). *Clause linkage in cross-linguistic perspective: data-driven approaches to cross-clausal syntax* (Vol. 249). Walter de Gruyter.

Geiger, W. (1995). *A grammar of the Sinhalese language*. Asian educational services.

Gelo, O., Braakmann, D., & Benetka, G. (2008). Quantitative and qualitative research: Beyond the debate. *Integrative psychological and behavioral science*, 42(3), 266-290.

Ghosh, S. (2009). Typology of Copular Clauses in Eastern Indic Languages. *Indian Linguistics*, 70(1-4), 123.

Girgis, M. (2002). Will nationals and Asians replace Arab workers in the GCC? *Egyptian Center for Economic Studies (ECES) Working Paper* No. 74. Le Caire, ECES.

Goertz, G., & Mahoney, J. (2012). *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.

Goodman, J. S. (1967). The development of a dialect of English-Japanese Pidgin. *Anthropological Linguistics*, 43-55.

Gordon, D. (2004). "I'm tired. You clean and cook." Shifting gender identities and second language socialization. *Tesol Quarterly*, *38*(3), 437-457.

Gramley, S. (2012). *The History of English: an Introduction*. Abingdon, England: Routledge.

Granger, S. (1997). Automated retrieval of passives from native and learner corpora: precision and recall. *Journal of English Linguistics*, *25*(4), 365-374.

Greenfield, William (1830). *A defence of the Surinam Negro-English version of the New Testament.* London: Samuel Bagster.

Grey, D. E. (2014). *Doing research in the real world* (3rd edition). Los Angeles: Sage.

Gripaldo, R. M. (Ed.). (2005). *Filipino cultural traits: Claro R. Ceniza lectures* (Vol. 4). CRVP.

Gunasekara, A. M. (1891). *A comprehensive grammar of the Sinhalese language: adapted for the use of English readers and prescribed for the Civil Service examinations*. GJA Skeen.

Guntman, A., & Avanzati, B. (2013). The Language Gulper: An Insatiable Appetite for Ancient and Modern Tongues.

Hachimi, A. (2001). Shifting sands: Language and gender in Moroccan Arabic. *Gender across languages: the linguistic representation of women and men*, *1*, 27-51.

Haeri, N. (1987). Male/female differences in speech: An alternative interpretation. *Variation in Language*, 173-182.

Haeri, N. (1991). *Sociolinguistic variation in Cairene Arabic: Palatalization and the qaf in the speech of men and women*, (doctoral dissertation). University of Pennsylvania, Philadelphia.

Haeri, N. (1996). *The Sociolinguistic Market of Cairo. Gender, Class and Education*. London-New York: Kegan Paul International.

Haeri, N. (2000). Form and ideology: Arabic sociolinguistics and beyond. *Annual Review of Anthropology*, *29*(1), 61-87.

Hall Jr, R. A. (1962). The life cycle of pidgin languages. *Lingua*, 11, 151-156.

Hall, R. A. (1966). *Pidgin and creole languages* (Vol. 7). Ithaca: Cornell University Press.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1), 23.

Hamdan, S. S. (1990). Social change in the Saudi family. Iowa State University. Retrieved from https://lib.dr.iastate.edu/rtd/11187/

Hammarström, H., Forkel, R., & Haspelmath, M. (2017). Glottolog 3.0. *Max Planck Institute for the Science of Human History*.

Hancock, I. (1986). The domestic hypothesis, diffusion and componentiality: An account of Atlantic Anglophone creole origins. *Substrata versus universals in creole genesis*, 71-102.

Hancock, I. F. (1977). Recovering pidgin genesis: Approaches and problems. *Albert Valdman (ed.)*, 277-294.

Hardjadibrata, R. (1981). *Anticipatory verbal intensifiers in Sundanese*. Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan= National Center for Language Development, Ministry of Education and Culture.

Hardjadibrata, R. (1985). *Sundanese: a syntactical analysis*. Dept. of Linguistics, Research School of Pacific Studies, The Australian National University.

Harrington, K., Litosseliti, L., Sauntson, H., & Sunderland, J. (2008). *Gender and language research methodologies*. Palgrave Macmillan.

Hassan, A. J. (1987). *A text-based model for the disambiguation of the temporal inerpretation of the verb in modern standard Arabic*, (doctoral dissertation). University of Salford.

Heath, J. (Ed.). (2008). *The veil: Women writers on its history, lore, and politics*. University of California Press.

Heine, B. (1979). 6. Some linguistic characteristics of African-based pidgins. In *Readings in creole studies* (pp. 89-98). John Benjamins.

Henry, G., & Zerwekh, R. (2002). SEAsite: Web-based interactive learning resources for Southeast Asian languages and cultures. *CALICO journal*, 499-512.

Herskovits, M. J., Herskovits, F. S., & Kolinski, M. (1936). *Suriname folk-lore* (No. 27). Ams PressInc

Hilpert, M. (2014). *Construction grammar and its application*. Edinburgh University Press.

Hmeidi, I., Kanaan, G., & Evens, M. (1997). Design and implementation of automatic indexing for information retrieval with Arabic documents. *Journal of the American Society for Information Science*, 48(10), 867-881.

Holes, C. (1990). *Gulf Arabic*. Psychology Press.

Holm, J. A. (1988). *Pidgins and creoles: Volume 1, theory and structure* (Vol. 1). Cambridge University Press.

Holm, J. A. (1989). English-Based Varieties. *Pidgins and Creoles: Reference Survey*, *2*, 405-546.

Holm, J. A. (2000). *An introduction to pidgin and creoles*. Cambridge: Cambridge University Press.

Holmes, J. (1990). Hedges and boosters in women's and men's speech. *Language & Communication*, *10*(3), 185-205.

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, 30(2), 161-172.

Hong, Y. (1991). *A sociolinguistic study of Seoul Korean: with a special section on language divergence between North and South Korea* (Vol. 12). Research Center for Peace and Unification of Korea.

Huang, L. F. (2013, November). Compiling a Corpus of Taiwanese Students' Spoken English. In *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27)* (pp. 199-205).

Huber, M. (1999). Ghanaian Pidgin English in its West African context. *A Sociohistorical and Structural Analysis. Amsterdam/Philadelphia: Benjamins*.

Hudson, R. A. (1996). *Sociolinguistics*. Cambridge university press.

Hunston, S. (2002). *Corpora in applied linguistics.*1st ed. Cambridge*:* Cambridge University Press.

Hymes, D. E. (1971). *Pidginization and creolization of languages*. Cambridge: Cambridge University Press.

Ibrahim, M. H. (1986). Standard and prestige language: A problem in Arabic sociolinguistics. *Anthropological linguistics*, *28*(1), 115-126.

Isa, B. Z., Halilu, K. A., & Ahmed, H. K. (2015). The concept of pidgin and creole. *Journal of Humanities and Social Science*, *20*(3), 14-21.

Jabeur, M. (1987). *A Sociolinguistic study in Rades: Tunisia*, (doctoral disseratation). University of Reading, UK.

Jawad, H. (1998). *The rights of women in Islam: An authentic approach*. Springer.

Jayaseelan, K. A. (2001). IP-internal topic and focus phrases. *Studia Linguistica*, *55*(1), 39-75.

Jenkins, J. (2003). *World Englishes: A resource book for students*. Psychology Press.

Jespersen, O. (1949). *A modern English grammar on historical principles. Part VII*. Munksgaard.

Jesperson, O. (1909). 1919. *A modern English grammar on historical principles*, *7*.

Johnstone, T. M. (1967). *Eastern Arabian dialect studies* (Vol. 17). Oxford UP.

Jolly, D., & Bolitho, R. (2011). A framework for materials writing. *Materials development in language teaching*, 107-134.

Joseph, B. (2004). On change in Language and change in language. *Language*, 80(3), 381-383.

Jourdan, C. (1991). Pidgins and creoles: the blurring of categories. *Annual Review of Anthropology*, *20*(1), 187-209.

Kamal, K. (1991). Speech acts in an indigenised variety: Sociocultural values and language variation. *English around the world: Sociolinguistic perspectives*, 308.

Kapiszewski, A. (2006). Arab versus Asian migrant workers in the GCC countries. *South Asian migration to Gulf countries: History, policies, development*, 46-70.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American anthropologist*, *86*(1), 65-79.

Kendall, T. (2011). Corpora from a sociolinguistic perspective. *Revista Brasileira de linguistica aplicada*, vol, no (2), PP. 361-389.

Kettemann, B., & Cossée, M. (1998). Fremdsprachen—Mädchensache? Geschlechtsspezifische Aspekte des Fremdsprachenerwerbs in der Schule. *Moderne Sprachen*, *42*, 11.

Khoury, N. F., & Moghadam, V. M. (Eds.). (1995). *Gender and development in the Arab world: women's economic participation: patterns and policies*. United Nations University Press.

Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the special issue on the web as corpus. *Computational linguistics*, *29*(3), 333-347.

Klein, W., & Perdue, C. (1997). The Basic Variety (or: Couldn't natural languages be much simpler?). *Second language research*, *13*(4), 301-347.

Knapik, A. (2012). On Selected Origins of Contact Languages. A Socio-Historical Perspective. *Styles of Communication*, *3*(1).

Kremers, J. (2003). Adjectival agreement in the Arabic noun phrase. *Proceedings of Console XI. Available online at http://www. sole. leidenuniv. nl/index. php3*.

Kremers, J. (2005). Arabic pidgins and Arabic colloquials: A syntactic comparison. *University of Nijmegen, Netherlands*.

Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. Sage.

Labov, W. (1966). The social stratification of English in New York City.

Labov, W. (1972a). Sociolinguistic patterns. Philadelphia: University of Pennsylvania Press, 43-54.

Labov, W. (1972b). The linguistic consequences of being a lame. In *Language in the Inner City*. Philadelphia: University of Pennsylvania Press, 255-97.

Labov, W. (1982). The social stratification of English in New York City, 3rd printing. *Washington: Center for Applied Linguistics*.

Labov, W. (2001). Principles of linguistic change. Volume 2: Social factors. *LANGUAGE IN SOCIETY-OXFORD-*, *29*.

Ladd, D. R., Roberts, S. G., & Dediu, D. (2015). Correlational studies in typological and historical linguistics. *Annu. Rev. Linguist.*, 1(1), 221-241.

Lakoff, R. (1973). Language and woman's place. *Language in society*, *2*(1), 45-79.

Lamel, L., Gauvain, J. L., Adda, G., Adda-Decker, M., Canseco, L., Chen, L., Galibert, O., Messaoudi, A., and Schwenk, H. (2004). Speech transcription in multiple languages. In: Proceedings of the ICASSP 2004. https://ieeexplore.ieee.org/document/1326655.

Larsen-Freeman, D., & Long, M. H. (1991). The linguistic environment for language acquisition. *An Introduction to Second Language Acquisition Research. Harlow, Essex: Longman*.

Laura, B. (2012). Women's transport: Solutions needed: *Arab News*, Retrieved 6 October 2018 https://www.arabnews.com/node/325728

Lee, R. (2006). *Globalization, language, and culture*. Infobase Publishing.

Leech, G. (1992). Corpora and theories of linguistic performance. *Directions in corpus linguistics*, *1992*, 105-122.

Leech, G. (2004). Developing linguistic corpora: a guide to good practice adding linguistic annotation. *Lancaster University*.

Leech, G. (2014). The state of the art in corpus linguistics. In *English corpus linguistics* (pp. 20-41). Routledge.

Lefebvre, C. (1998). *Creole genesis and the acquisition of grammar: The case of Haitian Creole* (Vol. 88). Cambridge University Press.

Lefebvre, C. (2004). *The genesis of pidgin and creole languages: A state of the art.* In *Issues in the study of Pidgin and Creole languages*. Amsterdam: John Benjamins Pub.

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of psychiatric research*, 45(5), 626-629.

Lewis, M. P., Simons, G. F., & Fennig, C. D. (2009). Ethnologue: languages of the world, Dallas, Texas: SIL International. *Online version: from http://www.ethnologue.com/*, *12*(12), 2010.

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh University Press.

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.

Llamzon, T. (1966). A Note on Predication in Tagalog. *Philippine Sociological Review*, 14(3), 150-155.

Long, D. E. (2005). *Culture and customs of Saudi Arabia: Culture and customs of the Middle East*. Westport, CT: Greenwood Press.

Long, D., E. (2005). *Culture and customs of Saudi Arabia*. Connecticut: Greenwood Press.

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3), 319-344.

Ludwig, J. (1983). Attitudes and expectations: A profile of female and male students of college French, German, and Spanish. *The Modern Language Journal*, 67(3), 216-227.

Macaulay, R. (2002) Discourse Variation. In: Chambers, J.K., Trudgill, P., and Schilling-Estes (Eds.). *The handbook of language Variation and change*. New York: Blackwell Publishing Co.

MacDougall, B. G., & De Abrew, K. (1979). *Sinhala: Basic Course*. US Government Printing Office.

Malik, A. N. (1995). *The phonology and morphology of Panjabi*. Munshiram Manoharlal Publishers Pvt. Limited. Department of Linguistics, Punjabi University.

Malmasi, S., Zampieri, M., Ljubešić, N., Nakov, P., Ali, A., & Tiedemann, J. (2016). Discriminating between similar languages and Arabic dialect identification: A report on the third dsl shared task. *In Proceedings of the third workshop on NLP for similar languages, varieties and dialects* (VarDial3) (pp. 1-14).

Maltz, D. N., & Borker, R. A. (1982). A cultural approach to male-female miscommunication. *A cultural approach to interpersonal communication: Essential readings*, 168-185.

Markey, T. L. (1982). Afrikaans: Creole or non-creole? *Zeitschrift für Dialektologie und Linguistik*, 169-207.

Massey, D. (1994). Space, place, and gender University of Minnesota Press. *Minneapolis* MN.

Mayring, P. (2001). Combination and integration of qualitative and quantitative analysis. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 2, No. 1).

McArthur, T. (1998). *The English languages.* Cambridge: Cambridge University Press.

McEnery, T. Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press (Edinburgh Textbooks in Empirical Linguistics), 2nd edition, 2001.

McEnery, T., & Wilson, A. (2003). Corpus linguistics. *The Oxford handbook of computational linguistics*, 448-463.

McEnery, T., and Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

McEnery, T., and Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press. In Richards, K., & Pilcher, N. (2016). An individual subjectivist critique of the use of corpus linguistics to inform pedagogical materials. *Dialogic Pedagogy: An International Online Journal*, *4*. https://files.eric.ed.gov/fulltext/EJ1148524.pdf.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.

McMahon, A. M. S. (1994). *Understanding language change*. Cambridge; New York: Cambridge University Press

McWhorter, J. (2019). Pidgins and Creoles. In *Oxford Research Encyclopedia of Linguistics*.

McWhorter, J. H. (1995). The scarcity of Spanish-based creoles explained. *Language in society*, 24: 213–44.

Mehrotra, R. R. (1997). Reduplication in Indian Pidgin English. *English today* 50, vol. 13 45-49.

Meng, X., & Meurs, D. (2009). Intermarriage, language, and economic assimilation process: A case study of France. *International Journal of Manpower*, *30*(1-2), 127-144.

Mesthrie, R. (2002). South Africa: A sociolinguistic overview. *Language in South Africa*, 11-26.

Metz, H. C. (1993). Saudi Arabia: A Country Study, Federal Research Division. In *Library of Congress, Washington, DC*. Retrived from: https://www.loc.gov/item/93028506/.

Michaelis, S. M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.). (2013). *The atlas of pidgin and creole language structures*. Oxford University Press.

Miller, C. (2003). The relevance of Arabic-based pidgins-creoles for Arabic linguistics.

Milne, W., S. (1993). *Apractical Bengali grammar*. New Delhi: Asian Educational Services.

Milroy, L., & Gordon, M. (2003). Beyond Phonology: Analyzing and Interpreting Higher Level Variation. Sociolinguistics. *Method and Interpretation*, 169-197.

Milroy, L., & Gordon, M. (2008). *Sociolinguistics: Method and interpretation* (Vol. 13). John Wiley & Sons.

Miyaoka, O., Sakiyama, O., & Krauss, M. E. (Eds.). (2007). *The vanishing languages of the Pacific Rim*. Oxford University Press.

Mohanan, T., & Mohanan, K. P. (1999). Two forms of BE in Malayalam. In *Proceedings of the LFG99 Conference. Stanford, CA: CSLI Publications. University of Manchester, Manchester, UK*.

Montagu, C. (2015). *Civil society in Saudi Arabia: The power and challenges of association*. Chatham House for the Royal Institute of International Affairs.

Montrul, S. (2004). The acquisition of Spanish. *Morphosyntactic Development in*.

Morgan, R. (1970). Know your enemy: A sampling of sexist quotes. *Sisterhood is powerful: An anthology of writings from the women's liberation movement*, 31-6.

Mufti, M. H. S. (2000). The Health Sector in Saudi Arabia. *Healthcare Development Strategies in the Kingdom of Saudi Arabia*, 9-20.

Mufwene, S. (1990). Transfer and the substrate hypothesis in creolistics. *Studies in second language acquisition*, *12*(1), 1-23.

Mufwene, S. (2001). *The ecology of language evolution*. Cambridge University Press.

Mufwene, S. (2002). Pidgin and creole languages. *International Encyclopaedia of the Social and Behavioral Sciences, Elsevier, Oxford*.

Mufwene, S. (2006). Language endangerment: An embarrassment for linguistics. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* (Vol. 42, No. 2, pp. 111-140). Chicago Linguistic Society.

Mufwene, S. S. (2019). Pidgins 17 and Creoles. *The Handbook of World Englishes*, 299.

Mufwene, S. S. (2020). Creoles and pidgins. *The Routledge Handbook of Language Contact*, 300.

Mufwene, S., & Condon, N. (eds.). (1993). *Africanisms in Afro-American language varieties*. University of Georgia Press.

Mühlhäusler, P. (1986). *Pidgin and creole linguistics*. Oxford: Blackwell.

Mühlhäusler, P. (1997). *Pidgin and creole linguistics*. London: Battlebridge Publications.

Mühlhäusler, P., & Harré, R. (1990). *Pronouns and people: The linguistic construction of social and personal identity*. Blackwell.

Müller-Gotama, F. (2001). *Sundanese* (Vol. 369). Spotlight Poets.

Müller, F., & Waibel, B. (2000). *Corpus linguistics: An introduction*. University of Freiburg Englisches Seminar. Retrieved from: https://www.anglistik.uni-freiburg.de/abteilungen/sprachwissenschaft/lsmair/corpuslinguistics

Mundy, P. J., Butchart, D., Ledger, J., & Piper, S. (2001). On the vultures of Africa. In *Wings Over Africa—Proceedings of the International Seminar on Bird Migration: Research, Conservation, Education and Flight Safety* (pp. 110-115).

Murchison, C. E. (1933). A handbook of child psychology, Vol. 1, 2nd rev.

Murtada, Y. H. (1996). *Vocational education in the Kingdom of Saudi Arabia*, (Doctoral dissertation). Institute of Education, University of London.

Muysken, P., & Smith, N. (1995). The study of pidgin and creole languages. *Pidgins and creoles: An introduction*, 3-14.

Næss, U. G. (2008). *"Gulf Pidgin Arabic": individual strategies or a structured variety?: a study of some features of the linguistic behaviour of Asian migrants in the Gulf countries* (Master's thesis). University of Oslo.

Nair, R. S. S. (2012). A grammar of Malayalam. *Language in India*, *12*, 1-135.

Naro, A. J. (1978). A study on the origins of pidginization. *Language*, 314-347.

Nasrin, M. B., & Van Der Wurff, W. A. M. (2015). *Colloquial Bengali*. Routledge.

Nesselhauf, N. (2011). Corpus linguistics: a practical introduction. *Retrived from:*http://www.as.uniheidelberg.de/personen/Nesselhauf/files/Corpus%20Linguistics%20Practial%20Introduction.pdf.

Neuman, W. L. (2007). *Basics of Social Research: Qualitative and Quantiative Approaches* (2nd ed.). New York, NY: Pearson Education.

Newman, J. (2008). Spoken corpora: Rationale and application. *Taiwan journal of linguistics*, *6*(2).

Niblock, T. (2006). *Saudi Arabia: Power, legitimacy and survival*. London: Routledge.

Nochlin, L., Gornick, V., & Moran, B. (1971). Woman in sexist society: Studies in power and powerlessness. *Edited by Gornick, Vivian, Barbara, K. Moran*, 344-366.

NOttmeyer, O. K. (2015). Intermarriage and the economic success of immigrants. *IZA world of labor*.

Owens, J. (2001). Creole Arabic: The orphan of all orphans. *Anthropological linguistics*, 348-378.

Paiva, V. L. M. O. (2011). Identity, motivation and autonomy in second language acquisition from the perspective of complex adaptive systems. *Identity, motivation and autonomy in language learning*, 57-72.

Paolillo, J. C. (2000). Formalizing formality: An analysis of register variation in Sinhala. *Journal of Linguistics*, *36*(2), 215-259.

Parkvall, M. (2019). Motivation in pidgin and creole genesis. *Language Dynamics and Change*, 9(2), 238-264.

Perdue, C. (ed.). (1993). *Adult Language Acquisition: Volume 2, the Results: Cross-Linguistic Perspectives* (Vol. 2). Cambridge University Press.

Pillai, S. K. (1965). *Malayalam lexicon: a comprehensive Malayalam--Malayalam-English dictionary on historical and philological principles*. University of Kerala.

Piper, N. (2004). Rights of foreign workers and the politics of migration in South-East and East Asia. *International migration*, *42*(5), 71-97.

Poplack, S. (1989). The care and handling of a mega-corpus: The Ottawa-Hull French project. *Language change and variation*, 4.

Porta, M. (Ed.). (2014). *A dictionary of epidemiology*. Oxford university press.

Potsdam, E., & Alanazi, M. (2014). Fi in Gulf Pidgin Arabic. *Kansas Working Papers in Linguistics*, 35.

Powers, W. R. (2005). *Transcription techniques for the spoken word*. Rowman Altamira.

Prabhakara Variar, K. E. (1979). *Studies in Malayalam grammar* (Vol. 21). University of Madras, India.

Precht, K. (2008). Sex similarities and differences in stance in informal American conversation 1. *Journal of Sociolinguistics*, 12(1), 89-111.

Prinzl, M. G. (2010). Hans Lindquist, Corpus Linguistics and the Description of English (Edinburgh University Press, 2009): a Detailed EvaluationIn Corpus Linguistics and the Description of English Hans Lindquist offers another introductory book to corpus linguistics, but aims. *Opticon1826*, (9).

Punch, K. F., & Oancea, A. (2014). *Introduction to research methods in education*. Sage.

Puri, R. K., Watson, C. W., & Newing, H. (2011). Conducting Research in Conservation. *A social science perspective*, 376.

Qafisheh, H. A. (1977). A Short Reference Grammar of Gulf Arabic.Tucson, Ariz., University of Arizona Press.

Querol-Julián, M., & Fortanet-Gómez, I. (2014). Evaluation in discussion sessions of conference presentations: theoretical foundations for a multimodal analysis. *Linguistics: Germanic & Romance Studies/Kalbotyra: Romanu ir Germanu Studijos*, 66.

Quirk, R., Greenbaum, S., Leech, S., & Svartvik, J. (1985). *A comprehensive grammar of the English language.* London: Longman.

Rajaraja V. A. R. and Roy, C. J. (1999). *Kerala Paaniniiyam: a treatise on Malayalam grammar*. Trivandrum, Kerala, India: International School of Dravidian Linguistics.

Ray, P. S. (1966). *Bengali language handbook*. Washington, DC: Center.

Raymond, K. H., & Abdullah, M. O. H. A. (1999). *Foreign labor in Asia: issues and challenges*. Nova Science.

Reinecke, J. E. (1975). A Bibliography of Pidgin and Creole Languages.

Rice, K. (2006). Ethical issues in linguistic fieldwork: An overview. *Journal of Academic Ethics*, 4(1-4), 123-155.

Richards, N., (2009). *The Tagalog copula.* Proceedings of AFLA 16.

Rickford, J. R. (1998). The creole origins of African-American Vernacular English: Evidence from copula absence. *African-American English: Structure, history and use*, 154-200.

Rickford, J. R., & McNair-Knox, F. (1994). Addressee-and topic-influenced style shift: A quantitative sociolinguistic study. *Sociolinguistic perspectives on register*, 235-276.

Rivas, M. (2017). Gender and Second Language Acquisition.

Riyadh in year 2013 Hijri. (2016). Retrieved from http://www.arriyadh.com/Eng/ Ab Arriyad/Content/getdocument.aspx?f=/openshare/Eng/Ab-Arriyad/Content/ Riyadh-in-year-2013.doc_cvt.htm.

Roberts, S. J., & Bresnan, J. (2008). Retained inflectional morphology in pidgins: A typological study.

Rodriguez, D. (2010). Language Group Specific Informational Report.

Romaine, S. (1988). *Pidgin and creole languages*. London; New York: Longman.

Rosenhouse, J. (1998). Women's speech and language variation in Arabic dialects. *al-'Arabiyya*, 123-152.

Roszak, B., & Roszak, T. (eds.). (1969). *Masculine/feminine: Readings in sexual mythology and the liberation of women* (Vol. 220). Harper & Row.

Roux, A. (1952). Quelques notes sur le langage des Musulmanes marocaines. *Orbis*, *1*(2), 376-84.

Royal, A. M. (1985). Male-Female pharyngealization patterns in Cairo Arabic: A sociolinguistic study of two neighborhoods. In *Texas Linguistic Forum* (No. 27).

Ruane, J. M. (2005). *Essentials of research methods: A guide to social science research*. Blackwell publishing.

Ryding, K. (2005). *A reference grammar of Modern Standard Arabic*. New York: Cambridge University Press.

Ryding, K. (2014). *Arabic: A linguistic introduction*. Cambridge: Cambridge University Press.

Sadiqi, F. (2003). *Women, gender and language in Morocco*. Leiden-Boston: Brill.

Sadiqi, F. (2007). *Language and gender in Arabic, Encyclopaedia of Arabic language and linguistics.* Brill: Leiden, vol. 2, 642-650.

Sakoda, K., & Siegel, J. (2003). *Pidgin Grammar: an introduction to the Creole English of Hawai'i*. Bess Press.

Salamah, S. (2018). *The quality of life, social care and family relationships of older unmarried Saudi women living in Jeddah: a qualitative study*, (Doctoral dissertation). University of Surrey.

Saldanha, G. (2009). Principles of corpus linguistics and their application to translation studies research. *Tradumàtica: traducció i tecnologies de la informació i la comunicació*, (7). Retrieved from: https://www.raco.cat/index.php/Tradumatica/article/view/154828/206722.

Salem, A. A. (2013). Linguistic Features of Pidgin Arabic in Kuwait. *English Language Teaching*, 6(5), 105-110.

Salmons, J. (2012). Designing and Conducting Research with Online Interviews, Cases in Online Interview Research. Pada books.google.co.id. *Diakses Jumat*, 17 Maret 2017.

Salt, M. (1992). Morphological Sketch of Tagalog. Retrieved 31 Jul 2019 from https://anthropology.uwo.ca/faculty/creider/9640a/MackenzieTagalog.pdf.

Samarin, W. J. (1971). Salient and substantive pidginization. *Pidginization and creolization of languages*, 117-140.

Samarin, W. J. (1982). Goals, roles, and language skills in colonizing Central Equatorial Africa. *Anthropological Linguistics*, 410-422.

Samarin, W. J. (1986). Protestant missions and the history of Lingala. *Journal of Religion in Africa*, 138-163.

Samarin, W. J. (2000). The status of Sango in fact and fiction. *Language change and language contact in pidgins and creoles*, 301-334.

Sankoff, G., & Brown, P. (1976). The origins of syntax in discourse: A case study of Tok Pisin relatives. *Language*, 631-666.

Sankoff, G., & Laberge, S. (1974). On the acquisition of native speakers by a language. Pidgins and creoles: Current trends and prospects, ed. by David DeCamp and Ian F. Hancock, 73-84. Retrived from:

http://www.langlxmelanesia.com/kivung%20vol%206%20no%201%20on%20the%20acquisition%20of%20native%20speakers%20by%20a%20language.pdf

Sapir, E. (1921). *Language: An introduction to the study of speech*. Courier Corporation.

Schachter, P., & Otanes, F. T. (1972). 1983. *Tagalog reference grammar*.

Schachter, P., & Otanes, F. T. (1983). *Tagalog reference grammar*. Univ of California Press.

Schilling-Estes, N. (2007). Variation and the Sociolinguistic Interview: a reconsideration. In 25th *International AESLA Conference* (pp. 971-986).

Schmitt, N. (Ed.). (2013). *An introduction to applied linguistics*. Routledge.

Schroeder, C. C. (1996). Enhancing undergraduate education: An imperative for student affairs. *About Campus*, *1*(4), 2-3.

Schuchardt, H. (1909). Die Lingua franca. *Zeitschrift für romanische Philologie (ZrP)*, *33*(4), 441-461.

Schulz, E. (2004). *A student grammar of modern standard Arabic*. Cambridge University Press.

Schumann, J. H. (1986). Locative and directional expressions in basilang speech. *Language Learning*, 36(3), 277-294.

Sebba, M. (1997). *Contact languages: Pidgins and creoles*. London, Macmillan.

Serrano, J. M., Garzón, F. C., & Manzanares, J. V. (2003). The Creolization of Pidgin: A Connectionist Exploration. In *Proceedings of EuroCogSci'03, the European Cognitive Science Conference 2003: Institute of Cognitive Science, Osnabrück, Germany, September 10-13, 2003* (p. 229). Lawrence Erlbaum Associates.

Seuren, P. A. (1984). The bioprogram hypothesis: facts and fancy. *Behavioral and Brain Sciences*, *7*(2), 208-209.

Seuren, P., & Wekker, H. (1986). Semantic transparency as a factor in creole genesis. *Substrata versus universals in creole genesis*, 57-70.

Shaalan, K., Monem, A. A., & Rafea, A. (2006). Arabic morphological generation from Interlingua. In *International Conference on Intelligent Information Processing* (pp. 441-451). Springer, Boston, MA.

Shackle, C. (1970). Punjabi in Lahore. *Modern Asian Studies*, *4*(3), 239-267.

Shuy, R. W., Wolfram, W., & Riley, W. K. (1968). *Linguistic correlates of social stratification in Detroit speech.* US Department of Health, Education, and Welfare, Office of Education, Bureau of Research.

Sībawayh, A. U., & Derenbourg, H. (1881). *Kitāb Sîbawayh: Le livre de Sîbawaihi: traité de grammaire arabe*. Paris: Impr. Nationale.

Siddiqui, M. A. (1998). Academic libraries in Saudi Arabia: A survey report. *The Reference Librarian*, *28*(60), 159-177.

Siegel, J. (2008a). Pidgins/creoles and second language acquisition. *The handbook of pidgin and creole studies*, 189-218.

Siegel, J. (2008b). *The emergence of pidgin and creole languages*. Oxford University Press.

Siegel, J. (2010). Pidgins and creoles. *Sociolinguistics and language education*, 232-264.

Sinaga, A. (2014). *Difference between Qualitative and Quantitative Analysis and How it Should be Applied in our Research.* Yogyakarta, Indonesia: Research Gate.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Sinclair, J. (2004). Trust the text. In *Trust the text* (pp. 19-33). Routledge.

Singh, I. (2000). Pidgins and Creoles: An Introduction (London: Arnold).

Skousen, R. (1989). *Analogical modeling of language*. Springer Science & Business Media.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". Retrived from:https://philarchive.org/archive/SLOFTA

Smart, J. R. (1990). Pidginization in Gulf Arabic: A first report. *Anthropological linguistics*, 83-119.

Smith, N. (2008). The origin of the Portuguese words in Saramaccan. *Roots of creole structures: Weighing the contribution of substrates and superstrates*, *33*, 153.

Souag, L. (2006). *The earliest recorded pidgin - Maridi Arabic*? Accessed 21 August 2016 from http://lughat.blogspot.com/2006/12/earliest-recorded-pidgin-maridi-arabic.html

Srnka, K. J., & Koeszegi, S. T. (2007). From words to numbers: how to transform qualitative data into meaningful quantitative results. *Schmalenbach Business Review*, 59(1), 29-57.

Stewart, W. A. (1962). Creole languages in the Caribbean. *Study of the role of second languages in Asia, Africa and Latin America*, 34-53.

Stewart, W. A. (1967). *Sociolinguistic factors in the history of American Negro dialects* (pp. 74-89). Peter Lang Publishing, Inc. New York, NY.

Subramoniam, V. I. (ed.). (1997). *Dravidian encyclopaedia*. International School of Dravidian Linguistics.

Sukmawan, R. (2017). Is Sundanese Shifted by Bahasa Indonesia? *International Seminar on Sociolinguistics and Dialectology:* "Changes and Development of Language in Social Life".

Sunderland, J. (2000). Issues of language and gender in second and foreign language education. *Language teaching*, *33*(4), 203-223.

Tagliamonte, S. (2006). *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.

Tahir, I. (2009). Copula in standard English and its Counterpart in Standard Arabic. *Al-Fatih Journal*, *39*, 1-9.

Tamimi. I. (2012). *More gender segregation: Saudi Arabia is planning on establishing four industrial cities for women*. Journomania. Retrieved from: http://journomania.net/opinion/35-opinion/670-more-gender-segregation-saudiarabia-is-planning-establishing-4-industrial-cities-for-women.html.

Tannen, D. (1990). *You just don't understand: Women and men in conversation* (p. 42). New York: Morrow.

Tannen, D. (1994). *Gender and Discourse*. New York: Oxford University Press.

Tannen, D. (2006). *Language and culture*. In In: Fasold, R., Connor-Linton, J. (Eds.), *An Introduction to language and linguistics*. Cambridge University Press, Cambridge, pp. 343e372.

Tannen, D. (ed.) (1993). *Gender and conversational interaction*. Oxford: Oxford University Press.

Tarone, E. (1980). Some influences on the syllable structure of interlanguage phonology. *International review of applied linguistics*, *18*(2), 139-152.

Tatevosov, S. (2005). From Habituals to Futures. In *Perspectives on aspect* (pp. 181-197). Springer, Dordrecht.

Taylor, D. (1971). Grammatical and lexical affinities of creoles. *Pidginization and creolization of languages*, 293-298.

Taylor, D. (1977). *Languages of the West Indies* (No. 498 T3).

Terdal, M. (1993). Watching whole language work. *TESOL Journal*, *2*(3), 25-29.

Thomason, S. G. (2001). An introduction to language contact. *Washington: George*.

Thomason, S. G. (2008). 10 Pidgins/Creoles and Historical Linguistics. *The handbook of pidgin and creole studies*, 242.

Thomason, S. G., & Elgibali, A. (1986). Before the lingua franca: Pidginized Arabic in the eleventh century AD. *Lingua*, *68*(4), 317-349.

Thomason, S. G., & Kaufman, T. (1992). *Language contact, creolization, and genetic linguistics*. Univ of California Press.

Thompson, H. R. (2012). *Bengali* (Vol. 18). John Benjamins Publishing. Thorne, B., & Henley, N. (1975). Language and sex: Difference and dominance.

Thompson, P. (2005). *Spoken language corpora.* In Wynne, M. (ed.) (2005), Developing Linguistic Corpora: A Guide to Good Practice, Oxford: Oxbow Books. Retrived from: http://ahds.ac.uk/linguistic-corpora

Todd, L. (1980). Pidginization and Creolization. *Annual Review of Applied Linguistics*, *1*, 19-24.

Todd, L. (2003). *Pidgins and creoles*. Routledge.

Tolstaja, N. I. (1981). The Panjabi language: A descriptive grammar. *Routledge and Kegan Paul*.

Towell, R., & Hawkins, R. D. (1994). *Approaches to second language acquisition*. Multilingual Matters.

Tryon, D. T., & Charpentier, J. M. (2011). *Pacific pidgins and creoles: origins, growth and development* (Vol. 132). Walter de Gruyter.

Tureaud Jr, U. C., Trammell, D., d Mark, H. T., Tureaud Sr, C., & Tureaud, S. (2018). Octavia Mary Casborn. *Casborn Creoles of Louisiana: Legally Divided In Black and White*, 160.

Turner, R. (2004). *Who is a native speaker and what is it they speak*. In Almoaily, M. (2012). *Language variation in Gulf Pidgin Arabic*, (doctoral thesis). Newcastle University.

USA, October 1989

Valdman, A. (1978). *Le créole: structure, statut et origine* (Vol. 8, No. 1). Institut d'études et de recherches interethniques et interculturelles.

Van Craenenbroeck, J. (2008). Ljiljana Progovac, Kate Paesani, Eugenia Casielles & Ellen Barton (eds.), The syntax of nonsententials: Multidisciplinary perspectives (Linguistics Today 93). Amsterdam & Philadelphia: John Benjamins, 2006. Pp. ix+ 372. *Journal of Linguistics*, *44*(1), 261-267.

Van Teijlingen, E. R., Rennie, A. M., Hundley, V., & Graham, W. (2001). The importance of conducting and reporting pilot studies: the example of the Scottish Births Survey. *Journal of advanced nursing*, 34(3), 289-295.

Vasko, A., (2010a). *Male and female language in Cambridgeshire: differences and similarities. Studies in Variation, Contacts and Change in English 4:* Cambridgeshire Dialect Grammar. Available at: http://www.helsinki.fi/varieng/series/volumes/04/articleA_male_female.html.

Veenstra, T. (2008). Creole genesis: The impact of the language bioprogram hypothesis. *The handbook of pidgin and creole studies*, 219-241.

Velupillai, V. (2015). *Pidgins, creoles and mixed languages: An introduction* (Vol. 48). John Benjamins Publishing Company.

Verhaar, J. M. (1995). *Toward a reference grammar of Tok Pisin: An Experiment in corpus linguistics*. Honolulu: University of Hawai 'i Press.

Versteegh, K. (1984). *Pidginization and creolization: The case of Arabic* (Vol. 33). John Benjamins Publishing.

Versteegh, K. (2008). Non-Indo-European pidgins and creoles. *The handbook of pidgin and creole studies*, 158-186.

Versteegh, K. (2014). Pidgin verbs: Infinitives or imperatives. *Pidgins and creoles beyond Africa-Europe encounters*, 141-169.

Vicente, Á. (2002). Une interprétation sociolinguistique d'un dialecte de Jbala: les parlers féminin et masculin dans le dialecte d'Anjra. *Aspects of the Dialects of Arabic Today, Rabat, Amapatril*, 336-344.

Vicente, A. (2011). Gender and language boundaries in the Arab world: Current issues and perspectives.

Voskuil, J. E. (1996). Aspects of verbal morphology in Tagalog. *Klamer, ed, 1996*, 51-67. Retrived from: https://www.ncsu.edu/linguistics/aboutsociolinguistics.php.

Wahba, K. (1996). *Linguistic variation in Alexandrian Arabic: the feature of emphasis* (pp. 103-125). Cairo: American University in Cairo Press.

Walters, K. (1991). Women, men and linguistic variation in the Arab world. *Perspectives on Arabic linguistics*, *3*, 199-229.

Wardhaugh, R. (2011). *An introduction to sociolinguistics* (Vol. 28). John Wiley & Sons.

Watson, J. (2002). *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Webb, D. A., Sweet, D., & Pretty, I. A. (2002). The emotional and psychological impact of mass casualty incidents on forensic odontologists. *Journal of Forensic Science*, 47(3), 539-541.

Wetzer, H. (1996). *The typology of adjectival predication* (Vol. 17). Walter de Gruyter.

Wheeler, M. W. (2007). *Syncope and apocope in the history of Catalan: An Optimality Theory approach.* University of Sussex. Falmer, Brighton.

Whinnom, K. (1956). *Spanish contact vernaculars in the Philippine Islands*. Hong Kong University Press.

Whinnom, K. (1965). The origin of the European-based creoles and pidgins. *Orbis*, *14*(2), 509-27.

Whinnom, K. (1971). Linguistic hybridization and the 'special case'of pidgins and creoles. *Pidginization and creolization of languages*, *91*, 115.

Widdowson, H. (1990). *Aspects of language teaching.* Oxford: Oxford University Press.

Widdowson, H. (2003). *Defining issues in English language teaching*. Oxford University Press.

Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied linguistics*, *21*(1), 3-25.

Winford, D. (2003). Code switching: Linguistic aspects. *An introduction to contact linguistics*, 126, 167.

Wiswall, A. Q. (2002). Gulf pidgin: An expanded analysis. In *Unpublished pro-seminar paper of June* (Vol. 3, p. 2002).

Wolfram, W., & Fasold, R. W. (1974). *The study of social dialects in American English*. Prentice Hall.

Wolfram, W., Montgomery, M. B., Allbritten, R., Bailey, G. H., Brammer, C., Davies, C. E. & Picone, M. D. (2018). *Speaking of Alabama: The History, Diversity, Function, and Change of Language*. University Alabama Press.

Wright, W., & Caspari, C. P. (1896). A grammar of the Arabic language: translated from the German of Caspari, and edited with numerous additions and corrections.

Wynbrandt, J. (2004). *A brief history of Saudi Arabia*. New York, NY: Facts On File.

Wynne, M. (Ed.). (2005). *Developing linguistic corpora: A guide to good practice*. Oxbow Books Limited. Ritrived from: http://icar.cnrs.fr/ecole_thematique/contaci/documents/Baude/wynne.pdf.

Yahya, A. (1989). On the complexity of the initial stages of Arabic text processing. *First Great Lakes Computer Science Conference*; Kalamazoo, Michigan,

Zuhur, S. (2011). Saudi Arabia (ABC. *CLIO*.

Zyhlar, E. (1932). *Ursrpung und Sprachcharacter des Altägpytischen*. Zeitschrift für Eingeborenen-Sprachen 23, 25-45. In Bae, C. H. (2004). Aramaic as a Lingua Franca during the Persian Empire (538-333 BCE). *Journal of Universal Language*, *5*(1), 1-20.