Language Variation in Gulf Pidgin Arabic

Mohammad Almoaily

069087387

Abstract

Gulf Pidgin Arabic (GPA) is a contact variety spoken in the Gulf States which has received a relatively little attention in the literature apart from a few descriptive works such as Smart 1990, Hobrom 1996, Wiswal 2002, Gomaa 2007, Almoaily 2008, Naess 2008, Bakir 2010, and Alshammari 2010. Importantly, since GPA is spoken by a non-indigenous workforce over a wide geographical area in a multi-ethnic speech community, language variation seems inevitable. However, to date, there is no account of variation in GPA conditioned by substrate language or length of stay. Therefore, in this thesis I analyse the impact of the first language of the speakers and the number of years of residency in their location in the Gulf as potential factors conditioning language variation in GPA. The data-base for the study consists of interviews with sixteen informants from three linguistic backgrounds: Malayalam, Bengali, and Punjabi. Interviews were conducted in two cities in Saudi Arabia: Riyadh and Alkharj. Half of the data is produced by informants who have spent five or less years in the Gulf while the other half has spent ten or more years in the Gulf by the time they were interviewed.

The analysis is based on ten morpho-syntactic 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. Given the fact that most of the current theories on contact languages have been made on the basis of Indo-European language based pidgins and creoles, analysing the above features in an Arabic-based pidgin promises to be a great addition to the literature of pidgins and creoles.

Results of this thesis show that both first language and number of years of stay in the Gulf seem to have little effect on my informants' choices as regards the studied morpho-syntactic features. There is a significant adaptation to the system of Gulf Arabic (the lexifier language) only with respect to one feature: conjunction markers. This finding could be taken to support Universalist theories of the emergence of contact languages. However, some substratal effect can still be noticed in the data.

Acknowledgements

I would like to express my appreciation to my thesis supervisors: Professor Anders Holmberg and Professor Isabelle Buchstaller for their valuable comments on previous drafts of this thesis. Without their advice and guidance this thesis would not have been produced.

I also deeply thank:

- The participants in this study, who volunteered to take part in it in spite of their full schedules…

- Noah Alqurain, Chairman of Albat'ha Community Centre in Riyadh, for arranging interviews with some of the subjects whom I recorded in this study…

I extend my gratitude to my parents and my wife for their consistent support and encouragement throughout the time it took me to complete this thesis.

List of Abbreviations

1	First person
2	Second person
3	Third person
ACC	Accusative
ADJ	Adjective
ADJP	Adjective phrase
ADV	Adverb(ial)
AGR	Agreement
ART	Article
COP	Copula
DAT	Dative
DEF	Definite
DEM	Demonstrative
DIST	Distal
F	Feminine
FUT	Future
IMP	Imperative
INDF	Indefinite
M	Masculine
NEG	Negation, negative
NP	Noun Phrase
OBJ	Object
PASS	Passive
PL	Plural
POSS	Possessive
PROX	Proximate
PRS	Present
PST	Past
SBJ	Subject
SG	Singular

In glosses:

-	separates morphemes
=	marks for clitic boundaries
.	separates metalanguage elements represented by one word in the source language.

Table of Contents

List of Tables and Figures

1. List of Tables

Chapter 1: Review of Related Literature

Chapter 2: GA and GPA, Definition and Description

Chapter 3: Description of the Substrate Langauges

Chapter 4: Data and Methodology

Chapter 5: Results

Chapter 6: Discussion

2. List of Figures

Chapter 1: Review of Related Literature

Chapter 4: Data and Methodology

Introduction

ma fi faham… Hina kalam sura sura… (M3)

'I don't understand (Gulf Arabic)… (People) here speak quickly…'

In the statement above, M3 – a Gulf Pidgin Arabic (GPA) speaker whom I interviewed in this study – expresses his inability to understand the form of Arabic used by members of the local population in Saudi Arabia when speaking to each other. His failure to understand Gulf Arabic (GA) is possibly not because its speakers speak quickly but rather due to the fact that GPA and GA are two distinct forms of language, with lexical, phonological, syntactical, and morphological differences.

So, what is GPA and how did it emerge? Sakoda and Siegel (2003: 1) write: 'Nowadays, the term "pidgin" has a different meaning in the field of linguistics. It 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'. According to this definition, the situation in the Gulf States is ideal for the birth/emergence of the contact language under investigation, as millions of workers – mainly from the Indian subcontinent – are employed in the Gulf with two-year, extendable, work permits. During their stay in the Gulf, the workers, who come from various linguistic backgrounds and usually do not speak Arabic, get in contact with GA speakers as well as speakers of other Arabic dialects. Since almost all vendors, workers in retail stores, shops, services, etc. in the Gulf are expats (Arab and Asian), there is an urgent need for communication between these two groups (i.e. Arabic-speaking locals and expats on one hand and non-Arabic speaking expats on the other). There have been a few descriptive studies on GPA in several countries of the Arabian Gulf such as UAE (Smart 1990), Kuwait (Wiswall 2002), Saudi Arabia (Almoaily 2008), Oman (Naess 2008), and Qatar (Bakir 2010).

Yet, despite the wide geographical distribution of this pidgin and despite the fact that it has many substrate languages, to date we lack an account of language variation in GPA caused by differences in the morpho-syntactic systems of the substrate languages of GPA or even by variation related to the length of stay in the Gulf. In my MA Dissertation (Almoaily 2008), I attempted to account for ethnic variation in GPA. My results showed that ethnicity had little effect on the linguistic performance of the sample. However, the results can hardly be considered reliable because of the small sample size: there were only four participants in the study, two from Pakistan and two from Bangladesh. Moreover, it could be the case that the impact of the first language of the GPA speakers

and their duration of stay in the Gulf on language variation is stronger than the impact of ethnic variation. Therefore, in the current project, I increased the sample and conducted a more in-depth analysis of GPA in order to investigate two potential factors on language variation in GPA; the first language of the speakers, and the number of years of residency in the Gulf.

The current data-base consists of dyadic and group interviews with sixteen GPA speakers from three linguistic backgrounds, Malayalam (5 speakers), Punjabi (4 speakers), and Bengali (7 speakers). On the basis of these interviews, I created a balanced 12,000 word-corpus, 4,000 words from every language group, 6,000 words from the newly-settled speakers and 6,000 words form the old-staying speakers. 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/hour of speech. The comparison is based on six morpho-syntactic variables: (1) free, bound, or dropped object and possessive pronouns, (2) presence or absence of the GA definiteness marker, (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.

The analysis suggests that the informants' first language and their length of stay in the Gulf seem to only have a weak effect on their choice between GPA linguistic variants. The long-term speakers have made a significant shift to GA after spending ten years in the Gulf in one feature only, conjunction. Similarly, conjunction was the only variable of the six above which exhibited a significant relation between the informants' L1 and their choice among GPA variables. On the other hand, the fact that long-term residents have made shifts towards norms different from those of the superstrate in two features (object and possessive pronouns and verbal agreement) could be an indication that the development of this pidgin is towards a variety different from the superstrate language, GA.

The weak substratal effect in the data of informants I interviewed could be an argument in favour of Universalist theories of pidgin and creole genesis. Yet, the existence of minor substratal effects in some linguistic features suggests that universal as well as substratal factors can be effective in the process of genesis of a contact language.

In Chapter 1, I discuss contact languages and their emergence. I also review some previous works on GPA and other Arabic-lexified contact languages. Chapter 2 aims at providing a morpho-syntactic description of both GPA and its superstrate language, GA.

The third chapter is a descriptive account of the linguistic features in four substrate languages, Bengali, Malayalam, Punjabi, and Urdu. In the fourth chapter, I provide an illustration of the research hypotheses, which are formulated based on differences between the substrate languages. I also discuss the thesis data and methodology. The results are listed in Chapter 5 and analysed in more detail in Chapter 6.

Before reviewing the literature on pidgin and creole languages, I provide a sociolinguistic background of the Arabian Gulf region, where GPA is spoken.

GA refers to the form of Colloquial Arabic spoken by the indigenous people of the Gulf Region (see Map 1 in Appendix D). Despite that there is very little literature describing Gulf Arabic, there seem to be diverging views regarding the definition of this Arabic dialect, its geographical spread, and its indigenous speakers. Qafisheh (1977) refers to Gulf Arabic as the form of Arabic spoken by the indigenous people of Bahrain, Qatar, and the United Arab Emirates. Hence, Qafisheh's strict definition of Gulf Arabic excludes the indigenous people of Kuwait, Saudi Arabia, and Oman. Holes (1990) and Smart (1990), however, have included Omani Arabic into their definition of Gulf Arabic. Indeed, Holes (1990) defines Gulf Arabic as the language spoken by the indigenous people of the Area from Southern Iraq all the way to the United Arab Emirates and Oman, including the dialects of the Eastern province of Saudi Arabia. This definition, however, excludes the variety of Colloquial Arabic spoken in the Centre, West, South, and North of Saudi Arabia. On the other hand, Najdi Arabic, the form of Arabic spoken in the centre of Saudi Arabia, has been referred to as part of Gulf Arabic by Feghali (2004).

Due to the variable views on the perimeters of Gulf Arabic, it should be noted that throughout this thesis I will use this term to refer to the language spoken by the indigenous people of Kuwait, Bahrain, Qatar, the UAE, and Oman and to the dialects spoken in the east and the centre of Saudi Arabia. The reason for excluding the varieties spoken in the West, North, and South of Saudi Arabia from the definition of Gulf Arabic is that they were not included among the GA varieties by the previous authors and that they have slightly different phonological systems (see Jarrah 1993, Al-Mozainy 1981). It should also be noted that since all the fieldwork data were collected in the centre of Saudi Arabia, namely in Riyadh and Alkharj, the description of GA in section 2.1 will mainly focus on the variety of Arabic spoken in the centre of Saudi Arabia. The remainder of this introduction will focus on the linguistic scene in the Gulf.

The official language in all Arabian Gulf States is Arabic. There are two forms of Arabic used side by side. This coexistence of two forms of the same language is known

in the linguistics literature as *diglossia*, a term coined by Charles Ferguson (1959), c.f. Fernández (1993), who claims that the term diglossia was in use prior to Ferguson's 1959 article. The higher form in this diglossic setting is Standard Arabic, which is mainly used in writing as well as in formal oral settings (e.g. education, religious sermons, public speeches, and government decrees). On the other hand, the lower form, Gulf Arabic, is mainly spoken and is rarely written. Gulf Arabic is used in informal settings (e.g. at home, with friends, etc.). All indigenous people in the Gulf fully comprehend Standard Arabic. However, only educated people can speak it fluently.

Due to the large number of non-Arabic speakers, mainly temporary immigrant workers, there are other languages that are used for communication in the Gulf region, the most common among which are English and GPA. Smart (1990) reported that English and Gulf Pidgin Arabic are used in the following situations:

- English: In dialogues between an educated Arab and a European or an educated Indian, between educated Indians, or between Europeans.
- Gulf Pidgin Arabic: In dialogues between a local and an immigrant worker.
- GPA, Hindi, or Urdu: between Indian immigrant workers.

The history, geography, and economy of the region are of primary importance when discussing the linguistic situation in the Gulf Region, due to the fact that they all have contributed in shaping the variety of Arabic known as Gulf Arabic and, more recently, also Gulf Pidgin Arabic. The next three sections shed light on some aspects of the history, geography, and economy of the Gulf region from a linguistic perspective with a focus on some factors such as linguistic simplification and language contact, which could have also led to the emergence of GA and GPA.

A. History

This section draws attention to some historical factors which could have possibly contributed to the emergence of GA. In particular, these factors are *language contact* and the ongoing process of *linguistic simplification*. The language spoken in the Gulf region in the early Islamic era was Classical Arabic with tribal-specific features (refer to Corriente 1976, Versteegh 2001). Versteegh (2004) reports that there were phonological and lexical differences between the so called *lughat* 'dialects' of the Arabian tribes speaking Classical Arabic. The transition from Classical Arabic to Modern Arabic varieties currently spoken all over the Arab world has been a debatable issue in the literature. For instance, Versteegh (1984, 2004) suggests that the Modern Arabic varieties have evolved due to a pidginisation-creolisation-decreolisation (PCD) process. The PCD

model suggests that following the spread of Islam in the Middle East in the seventh century, people in North Africa and in some other parts of the Middle East learned a simplified register of Arabic. This simplified register turned into a Creole when speakers of Arabic intermarried with speakers of the simplified register. Later, this Creole has leveled with Arabic in a decreolisation process.

The historical development of GA seems irrelevant to the PCD model as the language spoken by the indigenous people of the Gulf before the proposed PCD process took place was indeed Arabic, (Versteegh 1984, 2004), unlike the other parts of the Arab world where other languages such as Berber, Coptic, and Syriac were used before the spread of Islam. However, Versteegh (1984, 2004) suggests that the current form of Modern Arabic used in the Arabian Peninsula (i.e. GA) has evolved as a result of language contact between Classical Arabic and some decreolised varieties of Arabic. Some scholars have since shed doubt on Versteegh's PCD model. For instance, some suggest that the transition from Classical Arabic to Modern Arabic was simply due to language change where the *lughat* 'dialects of Arabian tribes' served as input for Modern Arabic (see Fischer 1995, Holes 1986, Al-Agmi 1997). To me, Versteegh's analysis of the transition from Classical Arabic to GA in the Gulf region seems plausible for three reasons. The first is the large number of borrowed lexical items from other languages into GA. There are words from Turkish, English, French, and Persian, which have definitely been transferred into Gulf Arabic as a result of language contact. It is this language contact which could have possibly led to the historic language change of the spoken language in the Gulf region from Classical Arabic to Gulf Arabic. The second reason why I am in favor of Versteegh's PCD model is that he used testimonies of Arab historians like Ibn Khaldoon, who lived in the 11th century, and linguists like Ibn Jinni, who lived in the 10th century, to prove his claim that the language of the Arabians, i.e. Classical Arabic, was influenced by other varieties of Arabic during the prolonged PCD process.

The fact that GA is simplified, compared to Classical Arabic, on the morphological level can also be accounted for as another potential evidence for the PCD model. Simplification is apparent in the loss of case marking in GA. For example, the Classical Arabic case markers used with nouns to indicate the subject and the object (the suffix *–o(n)* for the subject and the suffix *–a(n)* for the object) are dropped in GA. Thus, the Classical Arabic sentence *dharab-a Zaid-un ʕamr-a* 'Zaid hit Amr' would be *dharab-Ø Zaid-Ø ʕamr-Ø* in GA. Thus, unlike Classical Arabic, which has free word order, in GA the first noun indicates the subject and the second indicates the object.

Word order of the subject and the object is thus essential in GA due to the absence of case markers.

In fact, whether GA has evolved as a result of the PCD model, or simply due to historic language change, the claim I made here that the history of the region has played a crucial role in the evolution of GA is still valid.

In addition to the history of the Gulf area, its geography, discussed in the next section, could be one of the main factors which helped shaping GA.

B. Geography

The Arabian Gulf States are located in the centre of the Old World.[1] According to Alaidros (1998), the Gulf States have been a transit hub for trade ships carrying goods between Asia, Africa, and Europe. This brought the indigenous people living near the coast of the Arabian Gulf in contact with sailors from different linguistic backgrounds, which could possibly explain the large amount of loan words in GA from the languages spoken in nearby countries like Persian, Turkish, and Urdu.

The demography of the region, which has been radically changing since in the middle of the 20th century (Feghali 2004), shall also be taken into consideration when discussing the linguistic scene in the Gulf, particularly in terms of the large number of immigrant workers. In Saudi Arabia, for instance, there are 8.5 million foreigners compared to 18.7 million locals according to the 2010 census.[2] The number of immigrant workers in the Arabian Gulf is large for three reasons. The first is that the amount of available jobs cannot be covered by the locals (see the next sub-section). The second reason is that most of the indigenous people do not accept to work in socially low status and low income jobs. Another reason is the geographical proximity of South Asian countries to the Gulf. Due to the steady presence of a large number of immigrant workers from various linguistic backgrounds, the situation has been ideal for a development of a pidgin in the region.

Hence, the demography of the region has been greatly influenced by its economy and both demography and economy have played a role in the linguistic development of GA and GPA. Especially the relatively recent demographic change in the Gulf region has influenced GA and helped in the emergence of GPA. The next section discusses this issue in more detail.

---

[1] Some geographers use the term *Old World* to refer to Asia, Africa, and Europe (see Mundy, Butchart, and Ledger 1992).
[2] Retrieved 29 Dec 2010 from http://www.cdsi.gov.sa

C. Economy

I have argued above that the geographical location of the Gulf States has made the region a centre for trade and that the language contact between locals and traders from various linguistic backgrounds over long periods of time could have played a significant role in the emergence of GA. Indeed, the prosperity the Gulf States have witnessed after the discovery of oil in the region around the middle of the twentieth century has, according to Feghali (2004), influenced the GA varieties in Saudi Arabia. Feghali suggests in particular that the fact that large groups of Saudis have migrated to Riyadh and to the Eastern Province has fostered cultural diversity and interaction between the dialects of the region, the dialects of Arab workers in Saudi Arabia, and the languages of non-Arabs. Over time, this interaction had an influence on GA, leading to the levelling of the GA varieties at the phonological level, (e.g. the affricate allophones [ts] and [tš] of the phoneme /k/ are less used by the younger generation) and morpho-syntactic levels (e.g. the regional dialectal forms of the 2SG.F object and possessive pronoun, *-itš*, *-its*, and *–iš*, are replaced with *-ik* form, see Holes 1990, Bassiouney 2009).

Moreover, the large number of immigrant Asian workers, following the recent massive increase of jobs in the gulf region after the discovery of oil in the region, has led to the emergence of GPA, as locals need to communicate with those Asian workers who are employed in jobs such as shopkeepers, barbers, and bakers on a daily basis. The rest of this section highlights some immigration policies for foreign workers in Saudi Arabia, as stated in the website of the Passports Agency of Saudi Arabia (http://www.gdp.gov.sa, retrieved 30 December 2010). Statements in 1-5 below are some of the immigration rules that foreign workers in Saudi Arabia have to follow:

1) All workers need a work permit issued by the Passports Agency
2) The duration of the work permit is two years.
3) The work permit can be renewed for an unlimited number of times but the duration must not exceed two years every time the work permit is renewed.
4) The following details need to be stated in the work permit full name, nationality, age, job, sponsor/ employer, city of residence.
5) If the employer pays the immigrant worker off, his/her work permit is invalid and he/she has to leave within a week, unless employed by another employer.

Some immigrant workers and employers violate these immigration policies. The common violations are that the immigrant worker works in a different job to the one stated in the work permit, usually by running small businesses or doing DIY jobs. Some immigrant workers also run a business and pay a monthly/ annual sum to their 'fake

employers' in the work permit documents. Furthermore, some immigrant workers enter the country illegally or with a visitor permit and find work.

These violations could have contributed to the emergence of GPA due to the fact that 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), see Bakker (2008).

Another factor that is bound to have played a major role in the emergence of GPA is the linguistic complexity of the GA morpho-syntax as well as its phonology. On the phonological level, GA contains a number of typologically less common phonemes like the pharyngeal phonemes /ʕ/ and/ ħ/ and the fricative voiceless velar phoneme, /χ/, which are candidates for replacement with more typologically common phonemes. Indeed, the GPA phonetic inventory does not include these sounds, which have been replaced with the vowel /ɑ/, with /h/, and with /k/ respectively (refer to Smart 1990, Almoaily 2008, Naess 2008 for a full inventory of GPA and GA phonemes). On the morpho-syntactic level, for example, there is a large number of verbal conjugations to mark for person, gender, and number agreement (see section 2.1.1).

As discussed in section 1.1, linguistic simplification is proposed to be one of the vital factors for the emergence of new pidgins.

Chapter 1: Review of Related Literature

This chapter aims at tackling some issues in the literature of pidgins and creoles concerning their definition and emergence. I will begin by discussing some theories about the genesis of contact languages. This is followed by definitions of some common forms of language contact and the controversies in defining these forms. I then provide a historical overview of the study of contact languages.

In addition, this chapter investigates the linguistic features that the literature tends to ascribe to pidgins and creole languages as well as the typological implications raised by contact languages. I also attempt to report on the linguistic features of Arabic-based pidgins and creoles that have been pointed out in the literature and comparing them with GPA.

1.1 Contact Languages, History and Definitions

When two or more groups of people from various linguistic backgrounds meet and have an interest in communicating but do not have access to a lingua franca, we expect contact varieties to arise. Note that the minimum number of languages required for the creation of a contact language has been debatable (see Romaine 1988, Bakker 2011). Whinnom's (1971) tertiary hybridisation theory, for example, claims that the emergence of pidgins and creoles involves at least three languages. But there is evidence that pidgins or creoles can result from the contact of only two languages. For instance, in his review of Bizri (2010), Bakker (2011) asserts that the emergence of Pidgin Madame – an Arabic-based pidgin spoken in Lebanon – was the result of contact of only two languages, namely Levantine Arabic as the superstrate language and Sinhala as the substrate. Other contact languages which have evolved out of the contact of a lexifier language and only one substrate language are – according to Bakker (2011) – Trio-Ndyuka Pidgin and Berbice Dutch.

Contact languages vary in terms of their linguistic features depending on the duration of the language contact, its intensity, the languages spoken by the groups in contact (i.e. speakers of the higher, more prestigious, language – the superstrate language – and the speakers of the lower status languages, – the substrate languages), as well as many other factors (see Stewart 1965, DeCamp 1971, Hymes 1971, Holm 1988, Romaine 1988, Sebba 1997, Singh 2000, Knapik 2009, inter alia). Consequently, some scholars in the field of pidginisation and creolisation distinguish between a range of outcomes of language contact. Perhaps the most relevant two forms to the current study are the terms *pidgin* and *creole*. For many creolists, the difference between the two is that

pidgins are not native languages, while creoles are (Todd 1990, Muysken and Smith 1995, Sebba 1997). This distinction, however, is contentious as will turn out in the detailed discussion on the definitions of the forms of contact languages in section 1.1.2. But let us now consider how contact languages evolve in the first place.

1.1.1 *The genesis of contact languages*

The literature of pidgins and creoles boasts many theories that propose scenarios leading to the emergence of contact languages. These theories are known in the literature of pidgin and creole languages as the *theories of genesis* (refer to Holm 1988, 2000, McMahon 1994, Arends, Muysken, and Smith 1995, Todd 1990, Singh 2000, Kouwenberg and Singler 2008, and many others). Due to the continuous debate in the literature, some of the theories of genesis such as the European dialect origin hypothesis – which attempts to explain linguistic features observed in European-based pidgins and creoles by linking them back to dialects of the respective European languages (see Faine 1937) – have nowadays fallen out of favour (Basch 2009, Siegel 2010). Since this thesis aims at discovering language variation in GPA resulting from different morpho-syntactic structures of the substrate languages and from length of exposure to GA, I will only discuss the theories of genesis which are most relevant to the current study: Imperfect L2 learning, substrate influence, and Universalist theories.

- **The imperfect L2 learning hypothesis:** Since pidgins are learnt as a second language by adults, some scholars have claimed that pidgins and creoles could have emerged as a result of imperfect second language learning. According to Arends et al. (1995), the imperfect L2 learning hypothesis was first proposed by Coelho (1880), who assigned a possible role for universal aspects of language learning in the emergence of pidgins. This hypothesis is a precursor to what became later known as Universalist approaches (to be discussed in more detail below). The first clear statement about the possible role of transfer (i.e. transmission of elements of a speaker's native language onto the linguistic patterns of the target language, Gass and Selinker 2008) in pidgin and creole genesis was not made, however, until the last decade of the twentieth century by Mufwene (1990: 11), who stated that 'research on transfer in SLA and that on the substrate hypothesis in creolistics may benefit one another.' Some supporting evidence in favour of imperfect L2 learning hypothesis was adduced by Klein and Perdue's study (1997), which investigated the data of naturalistic (i.e. outside the

classroom) learning of English, German, Dutch, or Swedish as a foreign language over a period of thirty months. The learners developed an interlanguage[1], which was called by Klein and Perdue 'the basic variety' i.e. the language necessary for communication. Nearly one third of the informants in their study did not show any development after acquiring the basic variety except for the acquisition of vocabulary. Similarly, Den Besten, Muysken, and Smith (1995) and Singh (2000) listed some of the observed similarities in the data of both pidgins languages and of imperfect L2 learning such as invariant verb forms, fixed word order, and pre-verbal negation markers. Imperfect L2 hypothesis, thus, seems convincing as both contact languages and cases of imperfect second language learning could result from over-simplification and over-generalisation of linguistic rules. Moreover, adult language learners, who learnt a language outside the classroom, are not expected to fully acquire the new system. Instead, factors such as transfer and fossilisation are expected to take place (see White 2003, Siegel 2008b). Indeed, Singh (2000) argues that although some features of pidgins and creoles cannot be accounted for by this theory, the similarities between imperfect L2 learning and pidgins and creoles might not be due to mere chance. This theory of pidgin and creole genesis is favoured by some researchers including DeGraff (1999), Mufwene (1990), Bekker and Veenstra (2003), Winford (2003), and Field (2004). 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'. Thus, more research on the role of language acquisition on pidgin and creole genesis needs to be conducted. I will discuss the potential role of imperfect L2 acquisition on the linguistic structure of GPA in section 6.2.3.

- **The substrate influence hypothesis:** Some researchers believe that the relatively similar structures found across contact languages are in fact due to similar structures of most of the substrate languages of the European language-based pidgins and creoles. For instance, some scholars (such as Holm 1988, Hall 1968, Taylor 1971, 1977) have proposed that some linguistic features could have been carried forward to Atlantic creoles from their substrate languages. This is echoed in Arends et al. (1995), who claim that substrate influence is found in phonology,

---

[1] The term *interlanguage* was coined by Selinker (1972). See also the concept of *approximative system* by Nemser (1971).

morphology/lexicon, syntax, and semantics in Atlantic creoles. For instance, the syllable structure of Surinam creole is different to the European lexifier language and similar to the substrate languages Gbi and Kikongo. On the syntactic level, the typologically rare *serial verbs* are found in some West African-based creoles such as Akan, Gbe, and Kru and they are also a common feature of creole languages (Arends et al. 1995). Since GPA has substrate languages with divergent morpho-syntactic structures (see Chapter 3), it will be interesting to see if these structures can be linked to language variation within GPA (see section 3.4).

- **Universalist theories:** This type of theories focuses on the role of innate cognitive principles on the process of pidgin and creole formation rather than on the influence of the languages in contact. As discussed above, the history of Universalist theories of genesis can be traced back to the late nineteenth century. What seems convincing in such theories is that the structures of pidgins and creole languages display certain similarities irrespective of their different contributing languages. For instance, analytic morphology is attested in Arabic-based pidgins and creoles (see 1.5 below) and in many Indo-European pidgins and creoles, despite the fact that the Arabic-based pidgins and creoles on one hand and the European language-based pidgins and creoles on the other are more or less based on synthetic distinct superstrate languages and on different substrate languages.

  The most famous, though obviously controversial, theory within the Universalist domain is Bickerton's (1981) Language Bioprogram Hypothesis (LBH), which was a predominant theory in the 1980s and early 1990s. According to Veenstra (2008) the LBH was an attempt to answer the question of how the human language has evolved tens of thousands of years ago and how it has developed since then. In order to solve the puzzle, Bickerton tried to create a link between first language acquisition and creolisation. Consequently, the LBH claims that creoles emerge as a result of processes of first language acquisition, whereby children use their parental pidgin input in order to invent creoles. In other words, the offspring born to pidgin speaking parents use their innate linguistic capacities to convert the limited pidgin input to features similar to the ones found in full-fledged languages. In turn, all creoles – according to the LBH – are structurally similar because they were created by utilising universal human

linguistic capacities. Hence, the LBH assumes that comparing creole languages and the language produced by children may provide insights into the evolution of creoles. Thus, this theory distinguishes between the emergence of pidgins, as a second language learnt by adults, and the abrupt emergence of creoles, as a first language made up by infants utilising the parental pidgin input. The LBH has been the subject of much controversy since it was first proposed in the early 1980s. For instance, McMahon (1994) questions the claim that children invent complex structures found in creoles out of the input of less complex structures found in pidgins. Moreover, Seuren (1984) and Siegel (2008a) criticise the LBH for failing to explain the presence of certain features often present in the substrate languages such as serial verbs and pre-verbal tense, mood and aspect (henceforth TMA) marking (see 1.3). In fact, Seuren (1984) argues that Bickerton exaggerated in describing serial verbs and pre-verbal TMA as universal while in fact they are not. The existence of preverbal TMA markers in some Arabic-based pidgins and creoles (see 1.3 below), however, could not be the result of mere chance. This indeed calls for more research on the universality of TMA markers and serial verbs in pidgin and creole languages. One of the strongest arguments against the LBH is evidence from some creoles such as Tok Pisin which did not emerge abruptly (i.e. as a result of first language acquisition over only one generation, see also Eklund 1996, Siegel 2008a, Veenstra 2008). Some proponents of the role of language universals in the emergence of pidgins and creoles are Ferguson (1971), Todd (1974), Bickerton (1981), and Singh (2000). According to Muysken and Veenstra (1995), Universalist theories fall into two types: procedural (i.e. related to psycholinguistic strategies that are assumed to be adopted by speakers of two mutually unintelligible languages in contact) and constitutive universals (i.e. similar structures of contact languages such as SVO word order, pre-verbal particles, and morphologically complex reflexives). If the present study revealed that the sampled informants produced universal features of contact languages which cannot be traced to their first languages, and furthermore that divergent properties of their L1s did not have a significant effect on their production of GPA, this study would support Universalist theories of genesis.

1.1.2 *Defining contact languages*

I will now shed some light on debates in the literature concerning the definition of some of the resulting forms of language contact. Limitations stemming from the lack of

consensus in defining and distinguishing between contact languages – pidgins and creoles in particular – are further discussed in section 1.4.

One of the earliest attempts to distinguish between the outcomes of language contact is the jargon-pidgin-creole model, proposed by Hall (1966). It views the development of contact languages as a shift from an unstable form of communication (i.e. jargon) to a more rule-governed form of language (i.e. pidgin), to a nativised, stabilised contact variety (i.e. a creole). Hall's model, which is accepted by many creolists (see Holm 1988, Todd 1990, and Singh 2000), has been amended by some researchers. For example, DeCamp (1971) added a post-creole stage. Moreover, Mühlhäusler (1986) added other possible scenarios to the jargon-pidgin-creole lifecycle. I provide his model in figure 1 below:
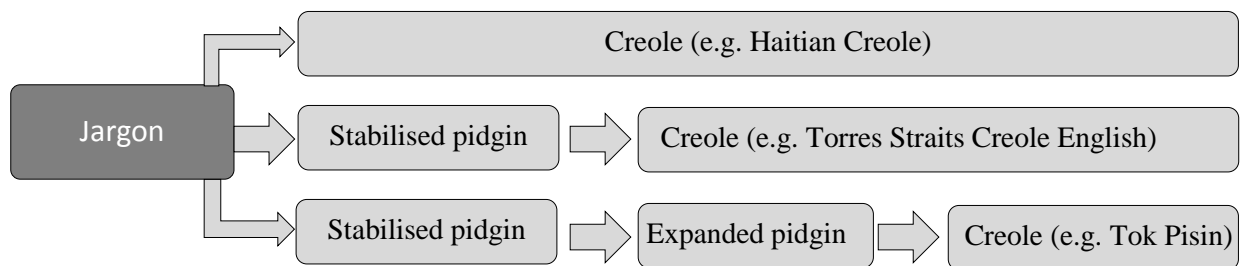
Figure 1: From jargon to creole, possible scenarios (Mühlhäusler 1986).

Hence, according to Mühlhäusler (ibid), there are three possible scenarios, jargon – creole, jargon – stabilised pidgin – creole, and finally, jargon – stabilised pidgin – expanded pidgin – creole. Below I define the terms jargon, pidgin, creole, and some forms of the post-creole continuum in more detail.

At an elementary stage of contact between two speech communities who do not share a language, the two speech communities may use a jargon to communicate with each other. Serrano, Garzón, and Manzanares (2003: 229) define a jargon as 'an unstable pidgin stage'. In addition to being restricted to very limited purposes, jargons have a great amount of lexical and syntactic variability from speaker to speaker since they are used at a primary stage of language contact where no stabilization has taken place or grammatical rules have been established (Singh 2000). Jargons can be distinguished from foreigner talk in that *foreigner talk* is the tendency of native speakers to simplify their speech when speaking to non-natives (Tarone 1980). A jargon is used by both groups in contact, while a foreigner talk is used by one group only, typically the native speakers of the more prestigious language in the contact situation. If the contact between the two speech communities lasts for a more protracted amount of time, the jargon, according to

the model in figure 1 above, may turn into either a pidgin or a creole. The terms *pidgin* and *creole* are discussed below.

Todd (1990: 1) defines a pidgin as a 'marginal language that develops as a means of communication for limited communication purposes between people who do not have a language in common'. Holm (1988: 5) distinguishes between a pidgin and a jargon in that a pidgin is 'more stable and has certain norms of meaning, pronunciation, and grammar'. Todd (1990) discriminates between two types of pidgins: restricted pidgins and expanded pidgins. The first type emerges as a result of limited language contact and is likely to disappear once the communication between the two groups in contact stops. An example of a restricted pidgin is Korean Bamboo English, which was used for communication between Koreans and Americans during the Korean War. Expanded pidgins on the other hand evolve in more protracted language contact, where the pidgin becomes a vital means for linguistic communication among hetero-lingual groups. It is believed that extended pidgins gain more social functions over time, which makes them less likely to disappear. For example, Todd (1990) suggests that West African pidgins were originally means of communication between Europeans and Africans and have only later been used between hetero-lingual groups of Africans living in multilingual communities. It should be noted, however, that the terms *jargon* and *lingua franca* have been employed in the nineteenth and early twentieth century to refer to the resulting forms of language contact now globally recognised as pidgins (Bakker 1995). In the next paragraph, I provide a definition for the term creole and discuss some contentions in the literature in defining this term; particularly as regards the distinction between a creole and a pidgin.

The term *creole* was first used in 1739 by a Moravian missionary to refer to Negerhollands, a Dutch lexifier creole (Muysken and Smith 1995). For many sociolinguists (see Todd 1990, Singh 2000) the term *pidgin* contrasts with the term *creole,* which arises when a pidgin becomes the first language of a speech community. However, some creolists (such as Muysken and Smith 1995, and Bakker 1995) do not link creolisation with nativisation. Indeed, according to Muysken and Smith (ibid), some extended pidgins such as Tok Pisin and Nigerian Pidgin English, have actually gained native speakers. What becomes obvious is that the discussion is as much a debate about the development and classification of pidgin and creole languages as well as about nomenclature: Many contact languages are problematic since they can be classified both as pidgins, given that they are spoken as a second language by some of the speech community, as well as creoles, since they are spoken as a first language by other

members of the speech community. It is not surprising, then, that a contact variety like the English-lexified variety spoken in Melanesia has been classified as a pidgin (i.e. Melanesian Pidgin English) by some researchers because it is a second language to the vast majority of its speakers. On the other hand, others refer to it as a creole because it is a native language for some other speakers and due to the fact that it has a rich morphological system (Siegel 2008b). It follows from the previous discussion that the nature of the debate in defining creoles is mainly between two schools of thought. The first sees creoles as nativised pidgins (as in Hall's 1966 pidgin-creole life-cycle discussed above). The second view suggests that creoles are not necessarily preceded by a pidgin stage (such as Jourdan 1991, Bakker 1995, Mather 2000, cf. the discussion on feature-based pidgin-creole distinction below). In essence, Sebba (1997) proposes a definition for the term *creole* which draws on Mühlhäusler's (1986) scenarios in figure 1 above and attempts to satisfy both views as it accounts for the two opposing views on the emergence of creoles. He defines a creole as 'a language with native speakers which results from contact without normal transmission' (Sebba 1997: 136). He, then, distinguishes between two types of creoles. The first set applies to creoles which resulted from 'abrupt creolisation through a sharp break in the transmission of language in some community' (i.e. without a prior stabilised pidgin) whereas the second type of creoles is those which evolved as a result of 'nativisation'. My concern with this definition, however, is that it does not provide criteria for classifying a newly discovered contact language as a pidgin or as a creole. In other words, it is hard to tell why the creoles which have evolved without a preceding pidgin are considered creoles, not pidgins.

Since the jargon-pidgin-creole model is controversial, some creolists (such as Hymes 1971, Hancock 1980, Markey 1982), have opted for the alternative approach of defining creoles in terms of their common linguistic features, rather than according to whether or not they have become nativised. Some of the features that have been proposed as typical to creoles are: strict SVO word order, use of adverbs to mark for TMA, and the use of reduplication as a word formation process. This list of features has been criticised by McWhorter (2000: 85) as 'insufficient, partly because many creole languages lack a few or even many of these features and partly because there are non-creole languages which combine many of them'. Such criticisms indeed raise a question, which I find myself unable to answer: If we reject the definition of creoles as nativised pidgins and cannot not agree on common features of creoles, then how can we differentiate between pidgins and creoles? This question is dealt with in more detail in section 1.4 below, in

which I discuss some limitations resulting from the lack of consensus in distinguishing pidgins from creoles and whether these limitations can possibly be minimised.

Nevertheless, the view that jargons may turn into pidgins or creoles, whether the emergence is gradual or abrupt and whether nativisation takes place or not, seem to be widely accepted in the literature. Below I discuss the two competing views on the emergence of pidgins (i.e. gradual and abrupt creolisation), as well as other alternative views.

1.1.3 *Gradual vs. abrupt emergence*

The LBH is probably the most famous example of an abrupt creolisation theory. However, despite the fact that this view was predominant in the 1970s and in the 1980s, it was not universally accepted (critics include Alleyne 1971, Chaudenson 1974, and Hancock 1980). The predictions made by the LBH and other adherents of the abrupt creolisation hypothesis are not corroborated by the gradual historical development of Tok Pisin and other contact languages such as Sranan (Bruyn 1993), Saramaccan (Smith 1987), and Haitian (Carden and Stewart 1988). For example, Arends et al. (1995) and Sankoff and Laberge (1974) report that Tok Pisin, has changed gradually and incrementally through the process of expansion over several generations and does not show the radical change in its linguistic structures postulated by Universalists to take place during the process of nativisation within one generation. In fact, as Sankoff and Laberge (1974) and Veenstra (2008) have shown, the stabilisation of Tok Pisin took place before its nativisation. For instance, the addition of a future irrealis marker in Tok Pisin was a result of a gradual process. At an earlier stage the English adverbial expression 'by and by' was used in Tok Pisin as an adverb meaning 'afterwards'. It eventually functioned as a preverbal particle and even as a future prefix. In the meantime, it historically developed from *baimbai*, to *bai*, to *bə*. Another instance of attested gradual creolisation is the split of the Sranan copular system, which took place over a period of around fifty years (see Arends and Bruyn 1995).

Bakker (1995) might be the most radical proposal within the gradualist camp, questioning the development process from a pidgin to a creole altogether, whether abrupt or gradual. He claims that there is no clear historical evidence of a creole that has been preceded by a pidgin other than Hawaii Creole English. Hence, he maintains that creoles have gradually gained their grammatical features without being preceded by a pidgin stage. Indeed, Bakker asserts that the structural differences between pidgins and creoles (such as varying word order in pidgins while almost all creoles have SVO word order)

make it rather difficult to hypothesise that all creoles are derived from pidgins. Hence, he suggests that pidgins and creoles would be more structurally similar if creoles have developed from pidgins. The same can be said about reduplication, which is common in creoles but rare in pidgins. Other scholars (such as Mühlhäusler 1986, Sebba 1997, Véronique 2003, and Siegel 2008a) might be considered an intermediate position, proposing that although some creoles have developed directly from pidgins, other creoles have actually emerged without a preceding pidgin.

1.1.4 *Concluding remarks*

Although some of the theories of genesis seem outdated nowadays, most hypotheses have been, and still are, a matter of ongoing discussion and debate. A particularly embattled frontline runs between two camps, which unsurprisingly epitomise the ongoing trench fight in linguistic theorising: the Universalists and the Substratists (see Holm 2000, Ramat 2009, Siegel 2008a, 2010). It is indeed difficult to favour one of these two opposing views over the other since both are based on assumptions about the emergence of creole languages which lack extensive historical documentation. Hence, often the same 'clue' can be used by proponents of both views. For instance, the tendency of pidgin and creole languages to be analytic, rather than synthetic, has been reported by Bickerton (1984) as a universal feature of adult second language acquisition. However, Holm (2000) warns against this assumption by stating that analytic morpho-syntax is actually a common feature of many African substrate languages. Note also that both the Universalists and the Substratists have constructed their theories of the genesis of pidgins and creoles based on structurally similar languages (i.e. either European superstrate languages or African substrate languages). Hence, assuming that such theories are meant to apply to all contact languages, regardless of their superstrate or substrate, risks the danger of over-interpreting such a narrow data-base (see 1.4). In this thesis, I aim to contribute to this literature by investigating the morpho-syntactic system of an Arabic lexified pidgin, Gulf Pidgin Arabic. I will assess the potential superstratal (i.e. Gulf Arabic) or substratal (a range of South Asian languages) influence on the morpho-syntax of GPA and I will compare these influences to the potential effect of the universal parameters in adult second language acquisition. By doing so, new evidence for some of the above theories might surface. Indeed, when tracking the historical development of the theories of genesis, it is noticeable that with more and newer data available – especially on untypical contact languages – some theories can make a comeback. For example, Siegel (2008b) reports that more creolists today have started to

accept the view that pidgin genesis is related to SLA/FLA acquisition despite the criticisms levelled towards the imperfect L2 acquisition hypothesis in the 1990s. The fact that GPA has gained some creole-like grammatical features such as reduplication and TMA adverbials (see section 6.3.2), while not being spoken as a first language could provide evidence in favour of the gradualist creolisation theory. Thus, observing the real-time development of recently established pidgins such as the pidgin investigated in this study, over a longer period might provide valuable evidence in favour of one of the two competing theories, abrupt vs. gradual emergence. This is especially the case since most research on the genesis of pidgins and creoles has been made on the basis of stabilised pidgins or creoles, but not on recently emerged pidgins (see Thomason 2008).

I will now provide a historical background on the development of the study of pidgin and creole languages, which progressed from scattered observations of what have been thought of as marginal languages to an academic field of research. Hence, I discuss scholarly and public attitudes towards contact languages, both before and after pidgins and creoles were recognised as an academic field of study. I will also identify some major trends in the history of the study of pidgins and creoles.

1.2 Contact Languages as a Field of Research

The study of pidgins and creoles has been neglected for centuries despite the fact that 'language contact seems likely to be nearly as old as language itself' (Holm 1988: 13). This might be due to the fact that in the past pidgins were considered to be broken, low-status varieties of language that do not deserve formal study. Indeed, because of this lack of esteem, pidgins and creoles have attracted humiliating or derogatory names such as nigger French, bastard Portuguese, broken English, cookhouse lingo, and coolie language (McArthur 1998). Singh (2000) claims that the rationale behind considering pidgins as broken forms of full languages is due to the fact that pidgins are not as linguistically sophisticated as full languages. In other words, the tendency of contact languages to have minimal structures, or simplified structures, made some people in the past regard them as primitive languages. It is thus not surprising that pidgins and creoles have not been recognised as a formal field of study until the middle of the last century. Indeed, in spite of the fact that pidgins and creoles have been the subject of linguistic scrutiny as evidenced by the major works by Schuchardt (1842-1927) and Reinecke (1930s), 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). Since then the value of the study of pidgins and creoles has

been gradually recognised by a large number of researchers in an increasing number of linguistic sub-disciplines. For instance, Bickerton (1974) suggested that creole studies are a significant opportunity for the testing and improvement of theories generated within the field of general linguistics. In the field of historical linguistics, McMahon (1994) asserts that the diachronic recency of many contact languages coupled with their rapid development – as compared to full languages – could provide significant observable data since the records usually used in historical linguistics are difficult to test due to the lack of observable evidence (see also Hopper and Traugott 2003). And so, the study of pidgins and creole languages can help in the confirmation or rejection of plenty of hypotheses in historical linguistics. Thus, Lefebvre (2004: 7) described pidgin and creole languages as 'a goldmine for historical linguistics.'

In the remainder of this sub-section, I briefly outline some trends in the history of pidgins and creoles. A more comprehensive historical account can be obtained from Reinecke (1977), Holm (1988), Arends, Muysken, and Smith (1995), and Thomason (2001). According to Holm (1988), the history of pidgin and creole languages can be classified into three major historical eras/trends. The first is the period before the sixteenth century (i.e. before European expansion). The second era is during European expansion, from the sixteenth century to the first half of the twentieth century, and the third era is after the establishment of pidgins and creoles as an academic 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 major turns on the discovery and documentation of contact languages.

Despite the scarcity of documented pidgin and creole languages in the era prior to the European expansion, contact languages are believed to have existed as early as groups of people with heterogeneous languages have started to come into contact with each other (see Hall 1966, Reinecke, et al. 1975, Muysken and Meijer 1979, Holm 1988). Indeed, the circumstances which are believed to be responsible for the emergence of contact languages such as slavery, trade, migration, and colonisation, have existed since ancient human history. Yet, according to Holm (1988), there are only two documented contact languages prior to the European expansion (i.e. before the sixteenth century). The first known text of a contact language (cf. 1.4) is a script of Maridi Arabic, a trade pidgin language thought to be spoken in Mauritania (Thomason and Eljibali 1986), or the Sudan (Owens 1996). This restructured Arabic variety goes back to the eleventh century AD. The substrate language of this pidgin is possibly Nilo-Saharan, yet, we lack clear evidence for this hypothesis and for the place where Maridi Arabic used to be spoken

(Souag 2006). The text was cited by an Arab geographer called al-Bakri, whose negative attitude towards contact varieties is well apparent in the description of the pidgin. In al-Bakri's documentation, Maridi Arabic is referred to as a deterioration of the Arabic language.

The other early attested contact language is Lingua Franca, also known as Sabir. The lexicon of the lingua franca of the Mediterranean is mainly based on a mixture of Southern Romance languages but it contains words from other substrate languages such as Arabic, Berber, and Turkish (Holm 1988). The first available text of this contact language was recorded in 1353 in Djerba, Tunisia. It is likely, however, that Lingua Franca had already been in use at the time Maridi Arabic was recorded (Holm ibid). Wansbrough (1996) even suggests that the lingua franca of the Mediterranean had been used as a language of diplomacy and trade for a period of three thousand years, from 1500 BC to 1500 AD. Nonetheless, as quoted from Holm (1988) earlier, the first available text of Lingua Franca was only recorded in the fourteenth century.

The documentation of contact languages prospered during the European expansion (i.e. from the seventeenth to the twentieth century) due to the growing number of researchers (i.e. historians, anthropologists, preachers, missionaries, etc.) who became interested in contact languages in this era (Holm 1988). And so, major contributions to the study of contact languages were made, even before pidginisation and creolisation was established as an academic field of study in the 1960s. The first documentation of a European-language based contact language I am aware of is of Negro Portuguese Pidgin, which took place in Sub-saharan Africa in 1516 (see Naro 1978). This was followed by a large number of reports/documentations of contact varieties. The first serious study of creole languages, however, did not take place until the 1730s. It was the result of the Moravian Church attempting to communicate with slaves in Suriname in Dutch. When this proved unsuccessful, the Moravian missionaries began to learn Negerhollands, the language spoken by the slaves. By doing so, they were among the first to treat a creole as an independent language (Holm 1988). Despite the generally accepted view in the nineteenth century that contact languages were trivial and rudimentary forms of speech, Greenfield (1830) defended creole languages suggesting that they are not broken forms of language, or degraded tongues, but rather rule-governed languages. Holm (1988) points out that Greenfield's position could be considered a major turn in the attitudes towards pidgin and creole languages. Research in the field of pidgins and creoles, however, only started to flourish several decades after Greenflied, when pidgins and creole languages were recognised as a field in linguistics in the 1960s, as previously

mentioned. As a result, DeCamp (1977) reported that the number of researchers working on pidgin and creole languages had grown from possibly a dozen to hundreds by the end of the 1970s.

Despite the major advances in the field of pidgins and creoles over the past fifty years, there is still a need for more research. For instance, many non-European lexifier contact languages, especially pidgins, remain under-researched. Yet, Versteegh (2008: 161) reports that there is a 'recent shift toward contact linguistics', which he considers an 'improvement'. Examples of works on non-Indo European based pidgins and creoles can be obtained from Heine (1979), Holm (1988), Baker (1996), Bakker (2003), Versteegh (2008), and Haspelmath and Michaelis (to appear). Yet, even in these works, which are mostly typological, the number of European language-based pidgins and creoles is high compared to the non-Indo European language based ones (see the discussion in 1.4). This could be due to the lack of research and documentation of non-Indo European language-based contact languages. Indeed, there is a high possibility that a large number of worldwide pidgins and creoles are undiscovered yet, which in turn calls for more extensive documentation and analysis of pidgins and creoles, particularly as regards the non-Indo European input language such as Chinese, Arabic, or Indonesian. In section 1.5, I will discuss the literature on Arabic-based pidgins and creoles in more detail. Section 1.6 gives a review of the literature of the pidgin under investigation.

But first I will investigate some common traits in the morpho-syntax of pidgin and creole languages, focusing in particular on the question whether atypical contact languages pattern in line with the proposed typological features for pidgins and creoles. In order to answer this question, I examine selected features in the morpho-syntactic systems of ten non-Indo European language-based pidgins to assess their compliance with the proposed typological features of contact languages. Below, I briefly define each one of these non-Indo European language-based pidgins by providing the region where the pidgin is spoken and the languages in contact during its creation. In the cases where it is difficult to determine which language is the lexifier and which language(s) is/are the lexified, I use the term *languages in contact*:

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

> **Ki-tuba** is spoken in the Congo. The lexifier language of Kituba is Kikongo. The substrate languages are French and Lingala (Ethnologue 2011).

Lingala is an expanded pidgin spoken in the Congo (Smith 1995). The lexifier is Bobangi, 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. Other languages in the 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. Pidgin English and other Papuan languages are 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).

In the next section, I review the typological features of contact languages. I also investigate to what extent these typological features are in fact attested in the above non-typical[2] contact languages.

1.3 General Features of Pidgin and Creole Languages

As I have mentioned in section 1.1, most of the theories of genesis were proposed to explain the similarities between different pidgin and creole languages. In this section I investigate the common features which have been reported across pidgin and creole languages. I will concentrate on the level of morpho-syntax as this linguistic level is the focus of this thesis. Please note that the focus will be on pidgin languages; creole features will be discussed in less detail since GPA has been classified as a pidgin by the majority of researchers (see 1.4.3 for a detailed discussion on the classification of GPA).

---

[2] They are referred to as non-typical contact langauges because they are all based on non-Indo European languages.

Compiling a list of the morpho-syntactic features of these two forms of language contact was not a straightforward process for two reasons. The first is that pidgins and creoles seem hard to tease apart (see the discussion in section 1.4). The second source of difficulty, which could be a direct result of the first, relates to the discrepancies in the literature as regards the classification of certain contact varieties as pidgins or as creoles, as discussed above.

An important proviso applies: the term *general features of pidgins and creoles* in the majority of research reported in the literature on pidgins and creoles tends to be based on Indo-European superstrate and West African substrate pidgins and creoles. Thus, the resulting 'general features' might in fact be an artefact of the superstrate and substrate languages involved in the contact situation and thus not be a reflection of the systematic features of all pidgins and creoles spoken worldwide (see the discussion in section 1.4).

What this effectively means is that we cannot assume that the proposed typological features are in fact found in all pidgin and creole languages worldwide (irrespectively of their input languages). I thus endeavoured to test some of these proposed features against the available data from non-Indo European contact languages. As noted by Versteegh (2008) this is difficult to do due to the lack of data on Non-Indo European pidgin and creole languages, but we can still make use of this little amount of data we currently have. Note that, by testing the general assumptions about the structures of pidgin and creole languages I do not intend to invalidate already existing theories. Rather, the aim is to contribute a non Indo-European-centric view of contact languages. Indeed, we should bear in mind that initial results of empirical research by Bakker, Daval-Markussen, and Parkvall (2011) suggest that Indo-European and non-Indo European contact languages do not behave differently as regards linguistic features. [3]

In general, pidgins and creoles are believed to have a reduced, if not absent, inflectional system, reduced derivation, and a small inventory of function words. The amount of reduction, however, is more in pidgins than it is in creoles (Bakker 1995, Muysken 1994, also refer to Sebba 1997 and Siegel 2004 for alternative classifications of simplicity in the morpho-syntactic systems of pidgin languages). These three types of reduction in the morpho-syntax of pidgin and creole languages (i.e. minimal inflection, derivation, and function words) are treated in more detail below. It should be noted that although most of the generalisations reported in the literature have been proposed after careful examination of tens, or sometimes even hundreds, of pidgin and creole languages

---

[3] The authors arrived at this conclusion after examining twenty-nine Indo-European based and five non-Indo European based pidgins.

(e.g. Bakker 2003, Roberts and Bresnan 2008, and Bakker, Daval-Markussen, and Parkvall 2011), the features listed below are subject to future revisions. Indeed, while this thesis by no means intends to invalidate the generalisations about the morpho-syntactic systems of pidgin and creole languages, it is important to triangulate the generally observed features against research based on non-canonical contact languages, which developed from non-Indo-European superstrates and/or non-West African substrates.

1.3.1 *Reduced inflection*

There seems to be a consensus among researchers in the field of pidgin typology that inflection is reduced, or even completely absent, in pidgin languages (refer to Todd 1980, Drecshel 1996, Hudson 1996, Holm 2000, and many others). Roberts and Bresnan (2008: 71) suggest that pidginisation 'may involve a shift from synthetic to analytic morphology' (i.e. having morphological systems which use free morphemes instead of inflections). Analytic morphology also seems to be typical to creole languages. Bakker (1995), for example, argues that reduced inflection is the only common feature between restricted pidgins and creoles. Thus, contact languages are reported to have a reduction of agreement markers (see Romaine 1988). I also show in Chapter 2 that – unlike the morpho-syntactic system of its superstrate language (GA) – the verb in GPA does not agree with the noun in person, number, or gender. Instead, one form of the verb, the third person singular masculine, tends to be used with all subjects. Since the verb in contact languages generally comes in a bare form, some contact varieties – typically creoles – tend to compensate the absence of affixes with TMA adverbials (e.g. *before*, *tomorrow*, *yesterday,* and *soon*), which can be preverbal or postverbal (see Bakker 1995, Roberts 1999, Winford and Migge 2007, and the discussion of contextualisation below). As shown in (1) below[4], Bakker (1995) claims that this phenomenon is also attested in some pidgin languages such as Hottentot Pidgin Dutch and Chinese Pidgin English. Bakker (1995) reports the following example from Chinese Pidgin English:

---

[4] Throughout this thesis, all in-text examples consist of three lines. The first is a transliteration in Roman script, in the second line, I glossed the sentence/utterance using the Leipzig Glossing Rules Conventions[4] (refer to page ii for a full list of conventions used in this study). In the third line, I provide a translation of the extract to English. See the example below, from my interview with the informant P2:

Kalam	same	same	ma	yigdar
Speech	same	same	no	3M-can.SG

'I cannot speak the same' (as Gulf Arabic).

(1)   before   my   sellum   for   ten   dollar

            before   I.SG   sell   for   ten   dollar

            'I sold it for ten dollars'

Bakker (1995: 32) argues that '[a]ll pidgins known to date have at least some derivational morphology. Pidgin morphology is however always simplified compared to the morphological system(s) of the lexifier language(s)'. As far as creoles are concerned, McWhorter (2005) reports examples of creoles with reduced inflections such as Haitian Creole, Korlai Creole Portuguese, and Berbice Dutch.

Below I will compare the ten seleceted non-Indo European pidgins, defined in 1.2 above, in terms of the existence or absence of inflection (see Roberts and Bresnan 2008, and Bakker 2003 for a more detailed account of inflection in pidgin languages). In table 1 below, I list some of the findings of Bakker (2003), Roberts and Bresnan (2008), as well as my own examination of some of the available translated texts of non-Indo European pidgins by Holm (1988), Mühlhäusler, Dutton, and Romaine (2003), and Bakker (2011). Since the purpose is not to describe the pidgin under comparison but to show existence or absence of affixes, detail is kept to a minimum in the comparison below. The sign + indicates that the affixe(s) marking/ indicating the linguistic feature in question exist in the pidgin, while - indicates that affixation is missing.

P/C AFFs for	Fanakalo	Ki-Tuba	Lingala	A-70	Restr. Sango	Restr. Swahili	Pidgin Madame	Hiri-Motu	Naga Pidgin	Pidgin Fijian
Tense	$+ (12 \rightarrow 6)^5$	+	+	-	-	-	-	-	+	-
Mood	-	-	-	-	+ (irrealis tone only)	-	-	-	-	-
Aspect	-	+	+	+	-	+	-	-	-	-
showing S-V Agreement	+	+	-	+	+	-	-	-	+	-

Table 1: Inflection in some non Indo-European pidgin languages (adapted from Holm 1988, Roberts and Bresnan 2008, Bakker 2011).

The data in table 1 show that only three contact varieties out of ten, namely Pidgin Madame, Hiri Motu, and Pidgin Fijian, do not show any form of affixation in their verbal systems, whereas the others use inflection to mark for tense and/or aspect and/or agreement. In fact, the table shows that affixes are used to mark for tense in four of the ten non-European pidgins listed in the table. Similarly, affixes are used in four pidgins in

---

[5] The number of affixes for tense in Fanakalo is reduced from 12 in the superstrate language to only 6 in the pidgin.

the sample to mark for aspect, and in 50% of the sampled pidgins to mark for pluralisation. The only grammatical information that is consistently not marked in this set of non-Indo European Pidgin languages is mood. Thus, with the exception of the morpho-syntactic feature mood, no generalisation can be made about the use of inflectional morphology to mark for any of the categories stated in table 1. Therefore, my findings run somewhat contrary to what has been commonly stated in the literature of pidgins and creoles, namely that pidgins lack affixation.

Bakker (2003) challenges the notion that morphological inflection is either reduced or absent in pidgin languages. After examining the data of more than thirty pidgins, including some less studied pidgins such as Asmara Pidgin Italian and Gulf Pidgin Arabic, he argues that inflection is empirically a common feature in pidgins: half of the pidgins he surveyed contained a form of inflection. For instance, in Asmara Pidgin Italian, the suffix –*ato* is used to mark for past tense. Similarly, in Kenyan Pidgin Swahili the prefix *ta-* is used to mark for future tense. Findings such as these lead him to claim that pidgins have a richer morphology than creoles. However, Roberts and Bresnan (2008) point out that none of the pidgin languages in Bakker's (2003) list had more inflectional morphemes than its lexifier language. Indeed, Roberts and Bresnan's (2008) criticism to Bakker (2003) seems convincing. Thus, so far, the claim that inflection is reduced as compared to the lexifier language has not been convincingly refuted. This raises the need for more analysis of untypical pidgins and creoles.

To sum up this section, there is a widespread belief within pidgin and creole typology that contact languages have extremely reduced inflectional morphology. This assumption has mostly been made on the basis of Indo-European/West African pidgins. Hence, it might be subject to revision. As demonstrated by Bakker (2003) and Roberts and Bresnan (2008) and shown in table 1 above, the hypothesis seems problematic when including some non-Indo European pidgin and creole languages into the typological account. However, the relatively scarce data we currently have on pidgins and creoles developing outside the European and West African contexts make it hard to falsify the claim that contact languages have reduced inflections, especially when considering the statement of Roberts and Bresnan (ibid) that none of the pidgin languages included in their comparison showed more inflection than its lexifier language. In other words, the presence of inflection in some pidgin languages does not contradict the claim that they have reduced systems compared to their lexifiers. Thus, more extensive work which includes as many pidgin languages as possible is required to check the validity of the

claim that the morpho-syntactic systems of pidgin languages contain less affixes as compared to full languages.

1.3.2 *Reduced word formation*

In this section I discuss some of the proposed lexical features across pidgin and creole languages. As detailed below, it has been claimed that pidgin and creole lexicon is characterised by semantic transparency, reduced derivation, and – to a lesser extent – reduplication.

A. Reduction of lexical items

One of the most noticeable features of pidgin languages is the small proportion of lexemes compared to non-contact languages (see Samarin 1971, Romaine 1988, Bakker 1995, Lutzeier 2005). Romaine (1988 and 1992) suggests that this numerically limited set of vocabulary items in pidgin languages is nevertheless still able to equip them with the ability to express all the semantic functions which can be expressed in full languages. This is due to the fact that synonymy is almost absent in pidgins (Sebba 1997) and also because of the wide use of lexical conversion (i.e. multiple meanings of a single lexical item) in pidgin languages (Romaine 1988, Kouwenberg and La Charité 2001). An example for lexical conversion is reported by Bakker (1995), who states that in Chinook Jargon the word *muckamuck* has many equivalents in English; including 'eat, drink, and bite'. Similarly, the term *kato* in Pidgin Fijian covers four lexical items in the Fijian language (Romaine 1988). Sebba (1997) suggests that the utmost exploitation of lexical items is due to their limited number, which rarely exceeds a few thousand words in even in well-established pidgins such as Tok Pisin (if we classify it as a pidgin rather than a creole). In addition, lexical items in pidgin languages also tend to be multifunctional so that one word in a given pidgin can function as a noun, as a verb, and as an adjective. For example, the Tok Pisin *sik 'sick'* can function as a noun as well as an adjective. These two phenomena, i.e. multiple meanings and functions for a single item, are captured by the concept of 'maximum use of a minimum lexicon' by Mühlhäusler (1986: 171). Creole languages, on the other hand, are characterised by lexical expansion, compared to pidgins, due to the intensive implementation of morphological processes such as reduplication (see D below), and compounding (Hancock 1980, Alleyne 1980, Kouwenberg and La Charité 2001).

B. Reduced derivational morphology

Analytic morphology, both in inflection and derivation, has been reported to be a general feature of pidgins and creoles (see Bakker 1995, Sebba 1997, Crowley 2008, and Smith 2008). Some pidgin languages, however, have been reported to use compounds instead of derivational affixes to create new meanings. For instance, according to Mühlhäusler (1986), the words for *man* and *woman* in Tok Pisin are compounded to create a new meaning, as in: *wroko.man* 'work.man' (i.e. worker).

The next part of this sub-section discusses some word formation processes which are used to overcome the shortage of derivational affixes in the lexicon of contact languages. These are: compounding with semantic transparency and reduplication.

C. Semantic Transparency

A language is called linguistically transparent when the meaning of compound words can be easily derived from the meanings of the constituents of the compound itself (Baayen and Schreuder 2003). Sebba (1997) argues that in pidgin languages the relationship between form and meaning is, generally, stronger than in non-contact languages such as English, Spanish, or Arabic. According to Seuren and Wekker (1986) and Sebba (1997), semantic transparency in pidgin languages is an outcome of the general tendency of pidgin languages to prefer simplicity. Thus, a learner of Tok Pisin, for instance, does not have to remember many words for male and female species. Instead, the word *man* is used with male objects and *meri* is used with female ones, examples are *hos man* 'stallion' and *hos meri* 'mare' (Verhaar 1995). These lexemes can also 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' (Sebba 1997). Examples such as these show that semantic transparency is an alternative to the use of derivational affixes such as the English *-er* as in *teacher* and *–ess* as in *Goddess*.

The final part of this sub-section discusses the repetition of a morpheme to create a new meaning, another word formation process frequently reported for pidgin languages.

D. Reduplication

Reduplication, as defined by Holm (1988), is a word formation process which involves the repetition of a word or a part of a word. We can distinguish between reduplication that forms a new meaning and iteration, which is simply the repetition of a word or a part of a word for emphasis. As put forward by Bakker (1995: 33), '[t]he

morphological process of reduplication is common (but not universal) in creole languages, but, strangely enough, rare in pidgins.' Some examples of meaning-forming reduplication in pidgin and creole languages are *vroevroe* in Negerhollands Creole Dutch where the word *vroe* 'early' is reduplicated to express a new meaning 'morning' and *yunyun* in Haitian Creole French, where the reduplication of *yun* 'one' creates a new meaning: 'distribute'. Reduplication has been reported to exist in a range of pidgin languages such as Nigerian Pidgin (Faraclas 1988), Pidgin Maori (Bell and Holmes 1990) and Indian Pidgin English (Mehrutra 1997). Bakker (1995), however, argues that reduplication cannot be conceived of as a global feature of pidgin languages.

In sum, in this section I have argued that pidgin and creole languages typically use less derivation, which is compensated by compounding and by the ability for a single item to cover many meanings as well as reduplication – in some pidgins and creoles – to create new meaning out of the limited lexicon. The next section will address the claim that pidgin languages have a reduced number of function words.

1.3.3 *Reduced inventory of function words*

Due to the fact that at their inception, pidgin languages are 'stripped of everything but the bare essentials necessary for communication' (Romaine 1988: 24), we would expect them to have few, or even lack, function words, especially those contact languages in the early stages of development (see Bakker 1995, Gleitman and Liberman 1995, Luria, Seymour, and Smoke 2006). In this section I discuss the existence vs. absence of three types of function words in pidgin and creole languages: copulas, definite and indefinite articles, and pronouns.

A. copulas

The traditional view of both pidgin and creole languages is that they lack copulas (see Ferguson 1971, Romaine 1988, McWhorter 1995, Sebba 1997). Note, however, that for creoles in particular, this assumption was questioned by some researchers like Arends, Muysken, and Smith (1995). In fact, Holm (1988: 174) even shows that some creole languages have developed forms of the verb *be* which are 'more complex than their lexical source languages'. As for pidgins, there seems to be an agreement that copulas are rare. Siegel (2008b: 26) even takes this criterion as a defining diagnostic of *restricted* pidgin languages by stating: 'Pidgin Fijian also differs from other restricted pidgins in its use of a copula'. Moreover, Rickford (1998) argues that the absence of the copula in African American Vernacular English is evidence that it originates from a

pidgin. The absence of the copula has also been reported in a range of non-typical pidgin languages such as China Coast Pidgin (Ansaldo, Matthews, and Smith 2011) and Romanian Pidgin Arabic (Avram 2010). However, copulas have also been reported to exist in some others, namely Chinese Pidgin English (Baker 1987), Nigerian Pidgin English (Faraclas 1988), and Yimas-Arafundi trade pidgin (Williams 2000). In table 2, I compare the existence vs. absence of copulas in the selected ten non-Indo European languages.

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

Copular verbs exist in eight out of the nine[7] polled non-European based pidgins. On the other hand, there is no copula in only one of the non-European language-based pidgins in this comparison, namely Fanakalo (Asher and Simpson 1994). This evidence, in combination with the fact that the copula exists in some Arabic-based pidgins such as GPA and Pidgin Madame – despite the fact that Arabic does not use a copula in the present tense – could be an argument against the claim that pidgins typically have null copula. Yet, more extensive comparisons would obviously provide a safer database to base these conclusions on.

Hence, unlike what many scholars believe about pidgin languages, copulas do exist in a number of pidgin languages, notably in non-Indo European based ones. The next section investigates the existence versus absence of markers for definiteness or indefiniteness in pidgin and creole languages.

B. Lack of definite or indefinite articles

Pidgins are generally assumed not to have overt definite or indefiniteness markers (see Sebba 1997, Samarin 2000). This is the case in GPA, as the GA definiteness marker is normally dropped in GPA (see Chapter 2).[8] The view that pidgins do not have markers for definiteness or indefiniteness, however, is challenged by Versteegh (1984), who argues that pidgin languages develop definite articles out of demonstratives. As for creoles, Holm (1988) argues that definiteness markers are widely used, where they

---

[6] The copula occurred once in a short translated text by Holm in his (1988) collection of texts.
[7] Pidgin A-70 was excluded due to lack of available data.
[8] There are no indefiniteness markers in the superstrate language, GA, nor in GPA.

usually come after the noun and mark the end of the noun phrase. Table 3 tests the claim that pidgins lack definite articles. It is immediately apparent that Lingala, Restructured Sango, Hiri Motu, and Pidgin Fijian use markers for either definiteness or indefiniteness whereas Fanakalo even 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: Definiteness and indefiniteness markers in some non-Indo European pidgins (adapted from Holm 1988, Roberts and Bresnan 2008, Bakker 2011)

Thus, it appears that the non-existence of definiteness or indefiniteness markers is not a defining feature of pidgin languages since they can be found in a number of the contact languages investigated here. It should be noted, however, that while table 3 is only a reflection of a small polled sample, and can thus not be considered to be representative of pidgins in general, the occurrence of in-/definiteness markers across these languages nevertheless sheds some doubts on generalising claims as regards the group of pidgin languages as such. The next part of this section discusses economy in the pronominal system of pidgins.

C. Reduced pronominal system

According to Mühlhäusler (1986), the number of contrasts in the pronominal systems of pidgin languages is minimised. For instance, in Chinese Pidgin English there are three pronouns only: First, second, and third, lacking the gender, number and politeness distinctions the superstrate language marks. Indeed, Romaine (1988) claims that pronominal gender and case distinctions are generally lost in pidgin languages. Another noticeable feature in the pronominal system of pidgin languages is the dropping of pronouns (see Mühlhäusler 1986, Romaine 1988, and Schumann 1986). Indeed**,** Romaine (1988, 1990) even claims that pro-drop is the unmarked form in pidgin languages. Bresnan (2000) suggests that it is common for pronouns to be free morphemes rather than clitic pronouns. This, she claims, is true even for many pidgins whose input languages have a bound pronominal system such as West African Pidgin Portuguese and Pidgin Hawaiian. As far as creoles are concerned, their pronominal systems are also claimed to be reduced (see Bailey 1966, Valdman 1978). But Holm (1988: 202) suggests that there is 'evidence that creolisation does not necessarily lead to extreme

morphological simplicity in pronominal systems'. For instance, he shows that the Portuguese-based creoles spoken in the Gulf of Guinea have pronominal systems which are as complex as those found in the Portuguese language. Table 4 tests the claim that pidgin languages have free pronouns rather than bound ones. As can be seen in the table, five out of the ten pidgins have bound pronouns.

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

Table 4: Existence vs. absence of pronominal clitics in some non-Indo European pidgins (adapted from Holm 1988, Roberts and Bresnan 2008, Bakker 2011).

The discussion of the proposed common morpho-syntactic features of pidgin and creole languages in subsections 1.3.1 to 1.3.3 reveals that the assumption that morpho-syntactic complexity is reduced in pidgin languages needs to be revised. For instance, the proposed typological feature that pidgin languages lack synthetic morphology seems to be invalid, certainly amongst the small sample used in this study. Moreover, eight of the ten non-Indo European language-based pidgins polled in this study have copulas, in contrast to the traditional view that pidgin languages lack a copula.

To sum up this section, many typological works on pidgins and creoles suggest that these languages have analytic morphology, that creoles have copulas while pidgins normally do not, that creoles have a strict SVO word order whereas pidgins have free word order, and finally that TMA preverbal elements, definite articles, and reduplication are common in creoles but not in pidgins. I would like to suggest, however, that a wider, more encompassing, typological account of the features of pidgin and creole languages might reveal that pidgin and creole typology is different to what it is thought to be. It is indeed crucial to take non-typical contact languages into consideration when proposing general features of contact languages. Indeed, the results of this thesis suggest that we might do well revisiting many of our assumptions about the structure of pidgin languages using a larger and more typologically diverse sample. Please refer to section 6.2.2 for a discussion on the compliance of GPA with the proposed general features listed above.

The discussion in this section – as well as in the two preceding sections – reveal that the literature of pidgin and creole languages still suffers from some inadequacies: (i) lack of agreement in defining pidgins and creoles – and thus in discriminating amongst the two –and (ii) lack of consensus on the features which define pidgins and creoles.

In section 1.4 below I aim to shed some light on the potential causes of these limitations and, if possible, how they might be solved or minimised. The suggestions discussed in this sub-section by no means claim to serve as a road map for overcoming the limitations in the field of pidgin and creole studies or even to resolve the highly controversial issues discussed in the previous three sections. Such a discussion is clearly beyond the confines of a PhD thesis. However, I would like to argue that raising the scholarly awareness of the current limitations in the literature of pidgins and creoles could be our first step towards resolving – or minimising – some of the limitations that pidgin and creole research currently faces.

1.4 Limitations in the Literature of Pidgin and Creole Languages, Causes and Possible Solutions

Research on pidgin and creole languages seems to have suffered from three main limitations. The first is that the literature is pervaded by a European-centric view of contact languages, which seem to have led to a biased view in defining pidgins and creoles, both in terms of proposing theories about their origin and emergence, as well as when it comes to listing their typological features. The other source of limitations in the conceptualisation of contact languages, which is closely related to the above factor, is the insufficiency of available data for pidgins and creoles. Finally, research in the field of contact languages could be held back by the fact that the current definitions of pidgins and creoles do not make a clear distinction between these two types of contact languages. These three factors are discussed in more detail in the sub-sections below.

1.4.1 *European-centric view*

As I have argued above, most of the work on pidgins and creoles has grown out of western scholarship, with the unfortunate result that cultural biases and geographical boundaries hinder the production of a more encompassing account of the world's pidgins and creoles. Although the scholarly awareness of these limitations has grown since the 1980s (see Holm 1988, Romaine 1988), a European-centric bias seems to still pervade this field of linguistics. Indeed, most of what we know today about pidgin and creole languages is 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 1990, Arends et al. 1995, and many others, with a few exceptions such as Bakker 2003, Versteegh 2008).

This bias can be explained by the fact that non-European-language based pidgins and creoles were much more difficult to investigate, or even notice, by the pioneering European scholars who documented most of the contact languages we currently know. Certainly, the contact languages based on African or Atlantic 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. So, for instance, it is easier for a Portuguese researcher (or a researcher speaking another Romance language) to spot a restructured form of Portuguese spoken somewhere in a Portuguese colony in Africa than to discover a restructured form of a local language. For that reason, it is very likely that there are numerous un-documented contact languages based on African or Atlantic languages, let alone the contact languages developing entirely outside the European context.

The European-centric approach to pidgins and creoles results in insufficient investigation of non-European contact languages, which could have in turn had a considerable effect on our conception of pidgins and creoles. For example, Stanford University's reference guide for pidgins and creole languages (2005) defines them, following Bickerton (1981), as follows:

> 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, Mufwene (2008) defines pidgins and creoles as new languages evolving out of the contact of European languages and non-European languages. These two references neatly exemplify a European-centric view, defining pidgin and creole languages as being a 'result of European Colonial expansion' despite compelling evidence that such varieties can arise out of contact of **any** two or more mutually unintelligible languages and that contact languages have evolved even before the European colonial expansion (see Reinecke 1977, Thomason and Eljibali 1986, Holm 1988, Bakker 2003). A similarly biased view can also be found in Samarin (1982, 1986), who claims that non-Indo European contact languages can only evolve under the influence of Europeans. Indeed, some researchers (see Stewart 1962, Whinnom 1965, Granda 1968) have deliberately used alternative names such as pseudo-pidgin, creoloid, semi-pidgin, and secondary hybrid, to denote contact languages which have evolved as a

result of language contact outside the Atlantic context. Versteegh (2008:161) suggests that these terms have been given to non-Indo European, non-Atlantic pidgins and creoles 'in order to avoid assigning true pidginhood or creole status to cases outside the restricted corpus of Atlantic creoles'.

These European-centric views can be challenged in two ways. First, some studies suggest that European and non-European based pidgins and creoles are structurally relatively similar (see Bakker 2003, Avram 2010, Bakker, Daval-Markussen, and Parkvall 2011)[9], also see the description of GPA in 2.1.2 below. Classifying non-European language-based pidgins and creoles as somehow qualitatively different from European-based ones therefore seems like a typological fallacy. The second challenge is the undeniable fact that contact languages, European-based and non-European based alike, have evolved as a result of language contact between at least two groups of people speaking different languages. Hence, if both European and non-European contact languages have evolved in relatively similar circumstances, e.g. as a second language learnt by adults in the case of pidgins, then why do some linguists feel the need to terminologically distinguish between the two?

Importantly, it is very likely that some of our current hypotheses on the emergence (see section 1.1.1) and typology (see section 1.3) of the world's pidgins and creoles have been influenced by this European-centric view, as well as by the shortage of data on non-Indo European contact languages (discussed in more detail in 1.4.2 below). Therefore, a close investigation of pidgins and creoles with Non-European input and a comparison of these with Indo-European based contact languages could broaden our understanding of the nature of pidgins and creoles and help us formulate more accurate theories of how contact languages emerge as well as establish more precise typologies of the typical features of these languages. An example of a European-centric view on the emergence of pidgins and creoles is Lefebvre's (1998) claim that pidgins typically emerge in communities where the majority of the population in contact speak a substrate language and the minority speak a superstrate language. This seems to have been the case for many European-language based pidgins and creoles, but not necessarily for all contact languages. For example, the case is completely the opposite in the development of Gulf Pidgin Arabic (see Almoaily 2008) and Pidgin Madame (see Bizri 2010). In both of these Arabic-based pidgins, the majority speak the superstrate language in contact and only a minority of the speech community speak the substrate languages. This suggests

---

[9] But see the findings in 1.3, where the initial results suggest that they might be not so similar.

that a broader, non-European centric, examination of contact languages might reveal that some pidgins and creoles actually evolve in circumstances different from what many researchers in the field of pidginisation and creolisation have expected to be the norm, based solely on data from Indo-European languages.

Moreover, as I have argued above, a more extensive examination of lesser-described pidgin and creole languages might lead to more accurate inventories of the typological features of contact languages. Some 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 pidgins and creoles. Yet, more comparative work is required as many non-Indo European-based pidgin and creole languages are still under-researched. This brings us to the next limitation in the literature of contact languages: the scarcity of data, most pertinently the lack of data on non-Indo European pidgins and creoles.

1.4.2 *Shortage of data on pidgins and creoles*

Researchers in the field of pidginisation and creolisation are often confronted with situations in which there is not sufficient data for the description of some individual language or where the available data of a contact language is difficult to interpret. This makes it hard to verify or refute theories about the evolvement and typology of pidgins and creoles. The discussion below is divided into three segments. I first discuss the difficulty of interpreting some of the available texts of pidgin and creole languages. I then emphasise the limitation or indeed impossibility to collect more data for a number of pidgins and creoles, due to the fact that they have long since died out. Finally, I argue for the high possibility that there are many pidgins and creoles which are still undiscovered.

One of the most common limitations that researchers in the field of pidginisation and creolisation encounter is the difficulty to interpret, or to test the accuracy of, the available data of a given pidgin or creole. For example, Reinecke et al. (1975) discuss Zyhlar's (1932) suggestion that the hieroglyphic symbols of the ancient Egyptian language show that the language might be a creole that has grown out of a pidgin spoken in the Nile valley. It is very hard to check the validity of this claim due to the controversies surrounding the interpretation of the Egyptian hieroglyphic symbols (see El-Daly 2005). This difficulty might be even more acute if a given text is the only available one for that particular contact language, as in the case of the Maridi Arabic text discussed below.

The other challenge for researchers in the field of pidginisation and creolisation is that a typological analysis of pidgins and creoles needs to be based on the interpretation of texts of as many contact varieties as possible in order to establish a more accurate typological inventory of these types of contact languages. This is obviously a too big task for a single researcher. Thus, Holm (1988: xii-xiiii) describes this predicament, which I shall refer to as the *interpretation paradox*, as follows: 'While having a single author for such a survey permits greater order and consistency, it also presents a fundamental problem: the scope simply exceeds the competence of any individual'. This necessitates relying on descriptive works from different researchers, which bears the problem that – inadvertently – not all linguistic descriptions are accurate or rely on comparable definitions, categorisations and theoretical premises. One solution to this limitation is consulting researchers and experts in the target contact languages. Hence, Holm (1988), for example, reported that he consulted as many as one hundred and fifty experts for his extensive survey of pidgins and creoles. Yet, despite this herculean effort, his survey still misses some of the existing pidgins and creoles (see the discussion below). Indeed, the exhaustive task of accurately documenting as many contact languages as possible, and then drawing typological conclusions about pidgins and creoles from these descriptions, would require the combined effort of large numbers of researchers from various linguistic backgrounds. By doing so, the problems of inconsistent conclusions made by different researchers working separately as well as the possibility of missing some pidgins and creoles are minimised. Indeed, most large scale language typology research is done in teams nowadays. For instance, more than fifty-five authors have collaborated in constructing the World Atlas of Language Structures (WALS), edited by Dryer and Haspelmath (2001). WALS surveys the typology of the phonology and morpho-syntax of most of the currently known languages. Similarly, the World Loanword Database (WOLD) is the outcome of a joint work of around fifty-three authors (Haspelmath and Tadmor 2009). There is a need in the literature of pidgin and creole languages for such such collaborative typological works. Even the promising project of Hapelsmath and Michaelis (to appear), entitled 'Atlas of Pidgin and Creole Structures' (APiCS), does not cover all of the worlds' known pidgins and creoles. Indeed, the APiCS project – which covers seventy-seven well-researched pidgin and creole languages – does not include some of the Arabic-based pidgins such as GPA and Pidgin Madame. One of the aims of this thesis is to determine to what extent a less studied Arabic-lexifier pidgin, GPA, complies with the proposed typological features of pidgins and creoles.

Another limitation when handling the data of pidgin and creole languages is that it is sometimes impossible to collect more data or to check the accuracy of existing texts, for example when analysing old and scarce texts of extinct pidgins and creoles. For instance, Hancock (1977) warns that the earliest available texts of European creoles such as the script of Malayo Portuguese recorded by Meister in 1692, might not be accurate. This is especially relevant at the phonological level, due to the fact that in some cases the spellings were influenced by the first language of the researcher who recorded the creole. Another common drawback when handling old texts is the difficulty to authenticate them. For instance, Souag (2006) questions the accuracy of what is thought to be the first available script of a contact language, the Maridi Arabic script, which dates back to the eleventh century. He warns that there is a high possibility that copyists of the text have altered it in one way or another in order to make it easier to print. It could be the case, for instance, that some phonological features of that text have been omitted. Souag is even hesitant to accept that the text actually exists, given that the book in which Thomason and Elgibali (1986) reported to have found the Maridi Arabic text is not provided in their bibliography.[10] Since written texts – especially those using conventional spellings – often fail to accurately represent the phonological features of a language, the phonological structures of contact languages which have disappeared before they were either phonetically transcribed or audio-recorded are particularly hard to verify. Sebba (1997: 244) writes as follows about two schools of transcribing pidgins and creoles:

> 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 clear advantage of phonetic transcriptions is that they provide the linguist reader with an accurate description of the target contact language. However, Ammon, et al. (2005) suggest that the diacritics used in some phonetic transcriptions make it rather hard to access the target linguistic item using electronic search tools. They thus suggest that compilers of creole speech-based corpora might consider improving the ease of access to the required piece of information by basing their transcription on the

---

[10] Thomason (personal communication, 26 Jun 2011) confirmed that Eljibali, whom I was not able to consult with, was sent a copy of the page containing the Maridi Arabic script by a former mentor of his.

conventional spelling of the language rather than providing full phonetic transcriptions. Since depending on etymological spellings alone might not provide us with phonologically accurate corpora, the transcriptions can be supplemented with audio recordings. I would like to second this argument since using conventional spellings along with audio recordings combines the benefits of accuracy and ease of access. Indeed, the accuracy of the data and its openness to re-checks can be further improved when conversations in the target pidgin or creole are audio-recoded in a digital format such as WAV, which allows for preserving the recorded interviews in a reasonably good quality and makes data handling and distribution maximally easy. In fact, some modern pidgin and creole researchers tend to present the reader/researcher not just with transcribed interviews, but also with sound files of the target contact language (see for instance Huber 1999, who supplemented his book describing Ghanaian Pidgin English with a CD containing recorded interviews). Thus, in the current thesis I recorded the interviews and transcribed them using Arabic spelling (refer to 4.3 for more details of the data analysed in this thesis). Note that such recordings plus transcripts of pidgins can serve as valuable data for future researchers, especially when considering the fact that some pidgins disappear once the contact between the two groups stops (see 1.1.2).

Another shortage in the data of pidgins and creoles could result from the possibility that there are contact languages which have not been discovered by linguists. This means that some of our conclusions, definitions, or descriptions of pidgins and creoles might be inaccurate. Crucially, most of the world's known pidgin and creole languages are located in a geographically limited space and are the product of contact between European languages and African languages or European languages and Caribbean languages (see Versteegh 2008). For example, WALS reports the existence of only thirty-two contact languages, twenty-four of which are either spoken in Africa or near the Caribbean Sea. The rest of the WALS list includes Hawaiian Creole (an English-lexified creole), five pidgins and creoles spoken in Asia (based on Assamese, Chinese, Spanish, Portuguese, and Malay), and six Australian pidgins and creoles (all based on English). This demonstrates that even contact languages spoken outside the African continent and the Atlantic Ocean included in WALS list mostly have a European language lexifier, with the exception of Naga Pidgin (Assamese-based), Russian Chinese Pidgin (Chinese-based), and Betawi (Malay-based). Earlier but similarly, Holm's (1988) survey of the world's pidgins and creoles includes forty-six pidgins and creoles which are either based on English, Portuguese, Spanish, French, or Dutch as opposed to only twenty-four pidgins and creoles based on other languages such as Amerindian, African,

and Asian languages. The number of European-based pidgin and creole languages in Holm's list is thus strikingly high when compared to contact languages based on other languages. In fact, the absence of two of the Arabic-based pidgins, GPA and Pidgin Madame, both in WALS and in Holm (1988), is further evidence that there might be many other undescribed, or lesser-known, contact languages worldwide that have not been taken into consideration when formulating supposedly general tendencies about the linguistic properties of contact languages. If it is indeed the case that what we currently know of the world's pidgins and creoles is just the tip of the iceberg, then most of the existing definitions, theories of genesis, and proposed general features of pidgins and creoles, discussed in 1.1, 1.2, and 1.3, are necessarily subject to revisions and alternations in the future as the processes of discovery, documentation, and analysis of pidgins and creoles continues. Hence, it seems that in order to arrive at a typologically accurate description of the linguistic properties and historical development of pidgin and creole languages, we must actively seek out undiscovered contact languages, especially in highly multilingual areas and in particular involving languages other than the typical European/West African languages.

1.4.3 *Is it a pidgin or a creole?*

In this sub-section I aim to focus my attention on one of the greatest limitations in the field of pidgin and creole linguistics, the lack of consensus in defining and distinguishing between pidgins and creoles. The ongoing battle about the definition, and thus the delimitation of these two, tends to leave researchers faced with the eternal question 'Is it a pidgin or a creole?' when discovering a new contact variety. This question is even harder to answer when the newly discovered variety has some of the linguistic features typical of pidgins and some other features typical of creoles.

The difficulty of making a clear distinction between pidgins and creoles solely based on their linguistic features is stressed by Jourdan (1991: 190-91), who argues that 'no structural characteristics seem to exist that would help discriminate creoles from pidgins apart from the sociohistoric circumstances of their genesis'. Indeed, the inability of most of definitions of pidgins and creoles to provide means of classifying contact languages into either a pidgin or a creole has let some researchers face the difficulty of classifying the contact variety they describe (see for instance Collins' 1980 study on Ambonese Malay and Trengganu Malay). This calls for more precise definitions of pidgins and creoles which help researchers classify a newly discovered contact variety as

a pidgin or as a creole. Alternatively, the distinction between these two terms should be given up.

The discussion in 1.1 has revealed that the possibly clearest, though disputable, proposal for distinguishing between a pidgin and a creole is whether or not the variety under investigation has been nativised or not. (Non-)nativisation provides a reasonable means of classifying a contact language, for example defining it as a pidgin in cases where a given contact language is acquired/spoken by adults past critical age such as the case of GPA, and as a creole when the contact language is acquired by children such as the case of Nubi Arabic (see 1.5 below). However, a possible drawback of the nativisation model is that it replaces one problem with another. Turner (2004), for instance, argues that the terms first language and native speaker are hard, if not impossible, to define clearly. It is tricky for example to determine whether a contact variety is a native language or not for the children who acquired it simultaneously with another language from birth. Hence, Sridhar (1996: 51) argues that '[i]t is neither necessary nor common to find native or near- native competence in all the languages of a multilingual's repertoire'. Furthermore, although the term native speaker is supposed not to include competent L2 learners, definitions of linguistic nativity often fail to distinguish between these two concepts. More controversies surrounding the definition of native speakers and L1 are detailed in Davies (2003), Jenkins (2003), and Turner (2004). Hence, Rampton (1990: 97) favours the use of the term expert instead of native speaker. Note however that the term English expert was criticised by Jenkins (2003) as being subject to personal judgement rather than being a technical classification of language speakers. In other words, by using the terms expert vs. non expert rather than native speaker vs. non-native speaker we are replacing one binary contrast by another and both are really difficult to tease apart.

Another proposal that has been put forward to discriminate pidgins from creoles is to classify them based on their linguistic features (see 1.1.2 above). Yet, this method has turned out to be problematic too. As I will show below, for instance, GPA carries features which are thought to be exclusively typical of pidgins as well as other features which are thought to be typical of creoles. Thus, it is impossible to classify the current contact variety as either a pidgin or a creole by looking at its linguistic features alone. I have also shown in 1.3 above that some non-Indo European pidgins and creoles are structurally different to the European language-based ones. Indeed, if the current typological features of pidgins and creoles are in need of revision, then it would be problematic to use them as criteria for distinguishing pidgins from creoles. Hall (1966:

23) in fact argues that '[t]here are no structural criteria which, in themselves, identify a creole as such, in the absence of historical evidence'. In other words, linguistic features alone are not good measures for classifying contact languages as pidgins or as creoles.

It follows from the discussion above that establishing a generalised definition of pidgins and creoles which clearly discriminates one variety from the other seems to be a hard, if not impossible, task. A potential way forward would be to give up the binary distinction (i.e. pidgin vs. creole) and implement a scalar model with pidgin on one end and creole on the other. It would indeed be more straightforward to classify a 'typical' pidgin as a pidgin and a 'typical' creole as a creole and a language in between as an advanced form of the former or as a simplified form of the latter (see figure 2) below. The hybrid form of language between pidgins and creoles was put forward by Bakker (2008: 135), who argues that 'it is better to distinguish between four types – jargons, pidgins, pidgincreoles, and creoles'. However, he defines a pidgincreole in a rather problematic way (see the discussion above for problems related to the term native language), arguing that it 'is a restructured language which is the primary language of a speech community, or which has become the native language for some of its speakers.' (Bakker, ibid: 139). Thus, I would like to argue that it is not necessary to draw a categorical line between any of the three forms (i.e. pidgin, pidgincreole, and creole). Indeed, it is advantageous not to link the term pidgincreole or creole to nativisation. Instead, I will define pidgincreoles as contact languages which carry features proposed to be typical to creoles as well as other features claimed to be typical to pidgins, irrespectively of their status as L1s or L2s (see figure 2). Note that the arrows in the figure below show that pidgincreole is a scalar concept. Hence, contact languages classified as pidgincreoles can be more like a pidgin or like a creole, depending on how close they are from the radical varieties on both sides. Note also that the language classified as a pidgincreole is not necessarily in a transitional stage between pidgins and creoles (see the discussion in 1.1.2 above as regards the idea that creoles are preceded by a pidgin stage).

Restricted pidgin
Free word-order
No copulas
No TMA markers
No reduplication
No definiteness markers

Pidgincreole

mixed   features

Creole
SVO word order
TMA markers
Reduplication
Definiteness markers
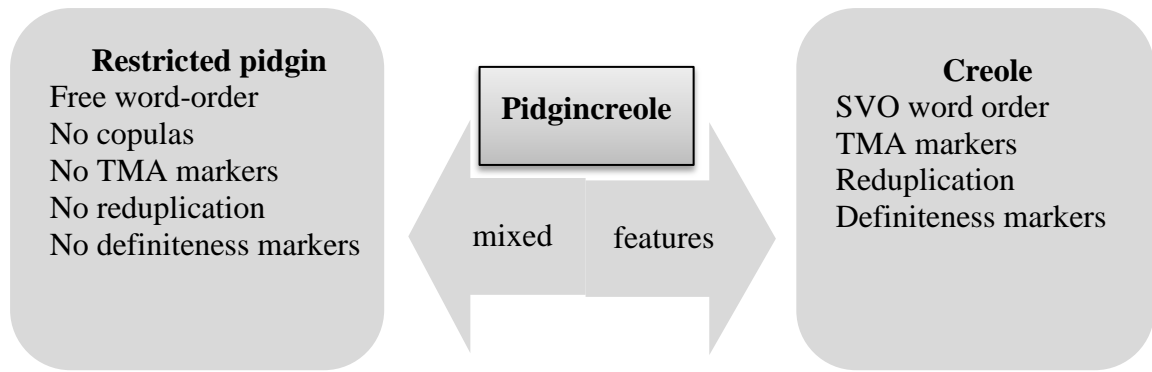
Figure 2: Pidgin-creole scale.

Lefebvre (1998, 2004) similarly suggests that no distinction should be made between pidgins and creoles. The scale above opens new horizons for further research on the nature of pidgin and creole languages. It is indeed subject to further revisions as more pidgin and creole languages are investigated. For example, the distribution of pidgincreoles, or whether they cluster somewhere along the scale, can be determined by studying the features of as many pidgincreoles as possible. Furthermore, it may turn out that contact languages which do not have a mixture of features are rare, and thus most of the world's contact languages are in fact pidgincreoles. Another area that is subject to further research is whether the features are ordered or not. It would interesting to see, for example, whether free word order and lack of TMA markers are always coexistent in contact languages. So far, we lack the data which will help us to answer the above questions with more certainty. For the time being, a contact variety like GPA can be easily classified as a pidgincreole using the scale above, as discussed in detail in the next paragraph.

GPA has been commonly classified as a pidgin in earlier works on the variety (such as Smart 1990, Naess 2008, Almoaily 2008). None of the previous works claimed that this contact variety has native speakers. It is even hard to imagine that it would ever gain native speakers in the near future given that its speakers, mainly adult Asian immigrants, only stay in the Gulf region temporarily and leave their families in their home countries. Thus, a definition that relies on creoles having native speakers would exclude the possibility that GPA ever becomes a creole. The hypothesis that GPA is a foreigner talk variety is also unlikely for two reasons. The first is that GPA is used as a means of communication by both groups in contact (in my case Saudis as well as immigrants), unlike what we expect in the case of a foreigner talk, where only one group

(i.e. speakers of the superstrate language) use a simplified form of the language. Indeed, as suggested by Smart (1990), GPA is not only used for communication between the indigenous people of the Gulf area and immigrant workers, but also between immigrant workers who do not share a common language. This strongly suggests that GPA is not merely foreigner talk but a medium of communication used by several linguistic groups in contact in the Gulf region. Moreover, the fact that GPA has established morpho-syntactic rules, as discussed in section 2.1.2, suggests that this variety is not merely a jargon or a foreigner talk, since both of these simplified codes are not rule-governed. As regards a distinction between pidgins and creoles in terms of their linguistic features, GPA seems hard to classify if we follow the binary model. Yet, GPA can be considered a pidgincreole if we are to follow the scale in figure 2. Given that pidgincreoles can be further classified in accordance with how close they are to either of the two varieties (i.e. pidgincreoles on the left side and others on the right side), GPA can be placed at the middle of the scale above. Indeed, GPA has a balanced mixture of pidgin and creole features. It carries some typical creole features such as the use of adverbs to mark for TMA (see section 2.1.2.1) and the use of reduplication as a word formation process. At the same time GPA exhibits features typical of pidgin languages such as a relatively free word-order and lack of definiteness markers (see section 2.1.2).

To sum up this section, I have argued that the European-centric view on contact languages and the relatively little research on non-European language-based pidgins and creoles could have hindered the production of a balanced account of contact languages. Implementing a scale with pidgin on one end and creole on the other and pidgincreole in the middle, where no boundaries are drawn between the three terms could help overcoming inconsistencies in defining the terms pidgin and creole.

In order to classify GPA in relation to other Arabic-based pidgins and creoles, I examine all Arabic-lexified contact varieties I am aware of in the next section.

1.5 Arabic-based Pidgins and Creoles

Apart from GPA, there are five other Arabic-lexified pidgins and creoles which I will review here: Nubi Arabic, Juba Arabic, Bongor Arabic, Romanian Pidgin Arabic, and Pidgin Madame. As far as I know, these contact varieties, as well as Maridi Arabic[11], are the only documented Arabic-based pidgins. In the first sub-section below, I provide a brief report of each of these five Arabic-based pidgins and creoles. I then concisely

---

[11] Maridi Arabic was not included in the review due to doubts surrounding its existence (see section 1.4).

compare the phonological and morpho-syntactic systems of GPA with the other Arabic-based contact languages.

1.5.1 *A report on Arabic-lexified contact languages*

The emergence of restructured forms of Arabic seems to have resulted mainly from two factors. The first is the spread of Islam in some parts of Africa in the eighth century, where Arabic and its restructured varieties became the language of religion, trade, and education (see Versteegh 1984, Holm 1988). This, as well as the spread of the Ottoman Empire in Eastern Africa in the nineteenth century (Holm ibid), seems to be responsible for the evolvement of the Arabic-based pidgins and creoles spoken in Africa, namely Maridi Arabic, Juba Arabic, Nubi Arabic, and Bongor Arabic. The rest of Arabic-based pidgins and creoles are mainly the result of the imported workforce from South and South-East Asia and Eastern Europe into some areas of the Middle East, as in the case of GPA (see Smart 1990, Wiswall 2002), Pidgin Madame (see Bizri 2010), and Romanian Pidgin Arabic (see Avram 2010). Hence, the Arabic-based pidgins and creoles spoken outside the African continent seem to have evolved decades after their African counterparts. It is not surprising, then, that the African Arabic-based pidgins and creoles are much better researched and more renowned in the literature on contact languages than the rest of Arabic-based pidgins and creoles. Note in this respect the early discussions of Arabic-based contact varieties such as the report of Maridi Arabic in East Africa, which goes back to the eleventh century (see Thomason and Eljibali 1986) and the report of Nubi and Juba Arabic in Reinecke (1937). On the other hand, Smart (1990) suggests that until the 1990s there was no report of an Arabic-lexified pidgin or creole spoken outside Africa. I will thus start my review with the restructured Arabic varieties spoken in Africa.

1.5.1.1 Juba Arabic

The Arabic-based creole known as Juba Arabic is widely spoken as a lingua franca in what is now known as Southern Sudan (see Holm 1988), particularly in Juba, the capital city of this newly-born country. The languages in contact in this creole are Arabic as a lexifier language and a mixture of Sudanic languages (Kaye and Tosco 2001). Miller and Woidich (to appear: 1) suggest that in Juba, Juba Arabic 'is the dominant lingua franca and is used in various contexts such as administration, local courts, preaching, broadcasting, daily talk in the streets or with the neighbours, songs, etc.' The estimated number of speakers of this contact language is 44,000 (Ethnologue

2011[12]). This number, however, might not be precise given that Juba Arabic is a cover term which includes a wide range of local varieties which range from a pidgin on one end and an acrolect on the other end of the spectrum (see Miller and Woidich to appear). Juba Arabic emerged during the control of the Ottoman Empire over Southern Sudan in the era between 1820 and 1870. At that time, it was used as a lingua franca by the multi-ethnic, multilingual, Ottoman army, as well as by the multilingual indigenous people of Juba and other parts of Southern Sudan. Despite the attempts of the British Condominium to 'stop the development of Juba Arabic as the main lingua franca in the highly multilingual Equatorial Province' (Miller and Woidich to appear: 2), it remained in use after the independence of the Sudan from the Ottoman Empire. A more detailed discussion of the history of Juba Arabic can be found in Mahmud (1979), Heine (1982), Holm (1988), Owens (1996), Versteegh (2004), Miller and Woidich (to appear).

Juba Arabic is of particular importance to creolists for two reasons. The first is that it has relatively recently gained native speakers (see Owens 1980). This could be the reason for what Miller and Woidich (to appear: 1) describe as 'massive and quick language changes' which Juba Arabic has undergone in the last thirty five years. The second reason for the importance of studying Juba Arabic is the recent independence of Juba from Northern Sudan. Until July 2011, Juba was under the rule of the Arabic-speaking Northern Sudan. Hence, (Juba) Arabic used to be an important language of trade, education, and other national affairs. Now, after the independence of Southern Sudan, Arabic might have a smaller role to play. Indeed, there are signs that English might take over as a lingua franca. For instance, the Government of the Republic of Southern Sudan website[13] indicates that the official language of this newly born republic is English. As regards the education policy, the Budget Sector Plan (2011: 15)[14] states that '[i]ntensive English training is required for Arabic pattern teachers. Many teachers, particularly in the Northern States, need to be trained to use the English language as medium of instruction'. Thus, it is to be seen whether the social status of Juba Arabic will change as the Republic of Southern Sudan establishes itself as an independent state from the Arabic-speaking Northern Sudan. Hence, monitoring the impact of the split of the Sudan on the maintenance and change in Juba Arabic might reveal exciting data to creolists.

---

[12] Retrieved from: http://www.ethnologue.com/show_language.asp?code=pga, on 17 July 2011
[13] Retrieved from http://www.goss.org/ on 11 October 2011
[14] Retrieved from http://www.goss.org/ on 11 October 2011

As is the case in many other pidgins, Juba Arabic is characterised by reduced phonological[15] (see DeCamp 1977, Miller and Woidich to appear) and morphological (see DeCamp 1977, Holm 1988) systems. For instance, Miller and Woidich (to appear) show that Juba Arabic eliminates most of the Arabic word-formation affixes and replaces them with free morphemes such as the use of *futu* 'to overpass' as a comparative and a superlative marker, instead of the Arabic infixation. Compare the Juba Arabic sentence in (2) below with its corresponding Arabic sentence in (3):

(2) úwo   kebír fútu   éta

     3 SGM   big   overpass   SBJ 2SGM

     'he is bigger than you'

(3) huwa   akbar   min-k

     3SGM.pro   big.COMP   than-OBJ 2SG

More details of the Juba Arabic morpho-syntax will appear below, as I compare GPA with other Arabic-based pidgins and creoles in section 1.5.2.

1.5.1.2 Nubi Arabic

This Arabic-based creole, also known as Ki-Nubi, is spoken by an estimated population of twenty-five thousand speakers in Bombo, Uganda and some parts of Kenya (Ethnologue 2011).[16] According to Wellens (2003), speakers of Nubi Arabic are descendants of multi-ethnic African soldiers recruited by the Egyptian ruler Muhamamd Ali Basha during his military operations in East Africa between 1820 and 1839 (see also Holm 1988, Miller 2002, Kuster 2003 for a detailed historical account of the emergence and development of Nubi Arabic). Similar to Juba Arabic, Nubi Arabic has a reduced phonological system, where the Arbic pharyngeal and uvular phonemes are replaced with more typologically common phonemes such as the velars /g/ and /k/. Moreover, the phonological system of Nubi Arabic is characterised with degemination, with the exception of very few words such as *tenna* 'our' and *yalla* 'well, OK' (Kuster, ibid). Note that due to the linguistic similarities between the two mutually intelligible restructured Arabic varieties *Nubi Arabic* and and *Juba Arabic* (consider also for example the reduction of tense and aspectual inflections and indication of contrasts via preverbal markers such as *gi* 'progressive' and *bi* 'future'), some researchers (such as

---

[15] There are only 21 consonants in Juba Arabic (Miller and Woidich to appear), compared to the 31 consonants in Standard Arabic (see Watson 2002).

[16] Retrieved from http://www.ethnologue.com/show_language.asp?code=kcn on 17 July 2011

Owens 1977, 1980) have suggested these two varieties are two dialects of the same contact language. However, Thomason and Eljibali (1986) argue against the monogenetic origin of these two Arabic-based contact languages. They suggest that the existence of similar features such as the preverbal TMA markers, might be due to influence from the lexifier language (i.e. Arabic) and not to the two contact languages being two dialects of the same creole. A more detailed discussion of the linguistic features of Nubi Arabic is provided in section 1.5.2 below.

1.5.1.3 Bongor Arabic

Bongor Arabic, also known as *túrko* and *túrgo*, is a pidgin spoken in Eastern Chad which has emerged as a result of constant contact between Arabic-speaking traders and locals speaking Chadic and Niger-Congo languages in the nineteenth and the twentieth century. The total number of Bongor Arabic speakers is unknown (Luffin and Woidich to appear, Ethnologue 2011). This Arabic-based pidgin shares some linguistic features with the two other creoles discussed above such as on the phonological level the lack of consonantal gemination, lack of long vowels, and phonological reduction. On the morpho-syntactic level, lack of gender distinction, and little derivational and inflectional morphology are also attested. There are also linguistic features which set Bongor Arabic apart from Nubi and Juba Arabic such as the replacement of the Arabic consonants /f/ and /t/ with /p/ and /d/, respectively.

The lack of research on the previous three Arabic-based pidgins has made Owens (2001) entitle his book 'Arabic Creoles: The Orphan of all Orphans'. Note however that the amount of research on the Arabic contact languages spoken outside the African continent is even scarcer. In the remainder of this sub-section, I report on the little that is known on Romanian Pidgin Arabic and Pidgin Madame. GPA is detailed in section 1.6.

1.5.1.4 Romanian Pidgin Arabic

This pidginised Arabic variety is scarcely discussed in the literature. The only reference I am aware of is Avram's (2007) presentation at the Colloquium on Peripheral Arabic Dialects at the University of Bucharest. This presentation, which was subsequently published as Avram (2010), discusses the emergence of this variety and provides a linguistic description of it. Romanian Pidgin Arabic is an extinct Arabic-lexified pidgin which emerged as a contact language between Romanian, Egyptian, and Iraqi oil workers in Iraq. Its birth can be dated at around 1974, when Iraqi oil companies started employing Romanian technicians and supervisors. The number of Romanian oil

workers, who were mostly non-English speakers, outnumbered their Iraqi and Egyptian co-workers and Romanian Pidgin Arabic has emerged as a consequence of this contact of two large groups without a common language. This pidgin died when Romanian oil workers left Iraq after the outbreak of the Second Gulf War in 1990. Just like the three Arabic-based pidgins discussed earlier, Romanian Pidgin Arabic is characterised by degemination (e.g. *sita* 'six' for the Arabic *sitta*) and replacement of some Arabic consonants such as the pharyngeal voiceless fricative /ħ/ and the velar voiceless fricative /χ/ with more typologically common phonemes: /h/ and /k/, respectively. The loss of gender and number distinctions in the morphology of Romanian Pidgin Arabic is another example of reduction in the linguistic system of this Arabic-based pidgin.

1.5.1.5 Pidgin Madame

The name of this recently-emerged pidgin was coined by Bizri (2005: 53), who describes the rationale of labelling it as Pidgin Madame as follows: 'We are calling it Pidgin Madame because the main actors/creators of the language are the Lebanese Madame and the Sri Lankan maid'. There are only two languages in contact in Pidgin Madame: Lebanese Arabic (superstrate) and Sinhala (substrate). The total number of Sri Lankan female workers in 2002 was 80,000 (Bizri 2010). The number of Pidgin Madame speakers, however, is certainly higher than this figure if we included the Pidgin Madame speaking locals (i.e. Lebanese families employing Sri Lankan maids). Unlike many other Arabic-based pidgins and creoles reviewed above, where the emergence was mainly the result of contact of male Arabic speakers with male recruited soldiers (as in Nubi and Juba Arabic) or male workers (as in GPA and Romanian Pidgin Arabic), Pidgin Madame has emerged as a result of contact between female maids and female 'madams'. This possibly explains the mostly feminine morphology of Pidgin Madame: 'Structurally, many of the linguistic forms are derived from this feminine master-servant relationship' (Bizri 2005: 53). For instance, Bakker (2011) reports the use of feminine adjectives with masculine nouns, as in *ebn-ik massut-ah* (son-your happy-F). 'Your son is happy'. Similarly, the verb in Pidgin Madame is mostly in the feminine past form or the feminine imperative form. For example Bizri (2010) quotes this sentence from her data *mister ana po nemi* (Mister I above sleep.IMP.F 'My employer sleeps upstairs'). Phonological reduction is also attested in pidgin Madame. Bakker (2011: 5) states that 'none of the Arabic uvulars, palatals, pharyngeals, and voiced fricatives is used' in this pidgin.

However, unlike the four Arabic pidgins and creole discussed above, phonemic distinction between long and short vowels is attested in Pidgin Madame (Bakker ibid).

In the next sub-section, I compare the morphology of GPA with the five Arabic-based pidgins reviewed above.

1.5.2 *A comparative account of Arabic-lexified pidgins and creoles*

In this section I attempt to compare the phonology and the morpho-syntax of GPA with the morpho-syntactic systems of other Arabic-lexified pidgins and creoles. Comparative accounts of GPA and African Arabic-based pidgins and creoles can be found in Naess (2008) and Bakir (2010). Other researchers (such as Miller 2002, Kremers 2005, Avram 2010, Luffin and Woidich to appear, Tosco to appear) have also provided typological accounts of Arabic-based pidgins and creoles. None of the researchers above, however, included all the five Arabic-lexified pidgins discussed above, along with GPA, in their comparative analyses. This could be possibly explained by the fact that non-African Arabic-based pidgins are less known in the literature than Nubi Arabic, Bongor Arabic, and Juba Arabic. Another possibility might be the lack of consensus in classifying non-African contact languages as pidgins or foreigner talk varieties. For example, Miller (2002) refers to GPA as a foreigner talk, whereas Smart (1990) and Bakir (2010) refer to it as a pidgin.

Hence, this section aims to compare GPA, Romanian Pidgin Arabic, and Pidgin Madame on the one hand and the Arabic-based African Pidgins and creoles on the other. Such a comprehensive comparison of Arabic-based contact languages is elucidating for two reasons. Firstly, all varieties are lexified by the same language but have emerged in very different situations (i.e. Nubi Arabic and Juba Arabic as military pidgins, Bongor Arabic as a trade pidgin, and GPA, Pidgin Madame, and Romanian Pidgin Arabic as workforce pidgins). Also the substrate and adstrate languages are vastly different (for example the substrate languages for Juba Arabic are Sudanic languages, whereas the substrate languages of GPA and Pidgin Madame are mainly Indian languages and the substrate language of Romanian pidgin Arabic is Romanian). The Bongor Arabic lexicon is influenced by French as an adstrate language. On the other hand, the lexicon of Juba Arabic and GPA has borrowed lexemes from English. The lexicon of Nubi Arabic, however, is enriched with Swahili and English loanwords (see Smart 1990, Wellens 2003, Luffin and Woidich to appear). It will thus be interesting to establish the role of the substrate and adstrate languages on the structures of the Arabic-based contact languages. Thus, we might want to investigate whether the Arabic-based pidgins can be

typologically classified according to the situation in which they have emerged or whether they form a coherent group based on their substrate languages (i.e. African languages vs. Indian languages vs. Romanian). Secondly, a comparative analysis allows us to contrast the linguistic features of relatively recently pidginised forms of Arabic (i.e. non-African Arabic based) with the older, more stabilised, ones (i.e. the African Arabic-based). Such an analysis affords important insights as regards the competing theories of pidgin and creole development. For example, if all the six Arabic-based pidgins and creoles showed similar linguistic features as the typological features of pidgins and creoles discussed in 1.3 above, despite the different situations in which they have emerged, the differences in the speech communities (i.e. natives vs. non-natives), and their different substrate languages, this might give some support to the Universalist theory of pidgin creation. Let us now investigate the linguistic features of these contact varieties.

A common feature of the phonological inventories of the Arabic-based pidgins and creoles is that they are reduced as compared to the lexifier language, Arabic. This reduction is in terms of the type and number of consonants, e.g. loss of pharyngeal and some uvular sounds, and the number of vowels (see Versteegh 1984, Owens 1989, Miller 2002, Naess 2008, Avram 2010, Bizri 2010). Crucially, however, the six Arabic-based pidgins and creoles are not identical in terms of the specific outcome of these reduction processes. They show, for instance, different replacements for some Arabic phonemes. For instance, Avram (2010) reports the replacement of the Arabic voiceless uvular fricative /χ/ with /h/ in Romanian Pidgin Arabic (e.g. *hamsa* 'five' for *χəmsah*). In GPA (see Almoaily 2008), Pidgin Madame (see Bizri 2009), Bongor Arabic (see Luffin and Woidich to appear), Nubi Arabic (see Wellens 2003), and Juba Arabic (see Miller and Woidich to appear), however, /χ/ is replaced with /k/. Another difference as regards the phonological systems of the Arabic-based pidgins and creoles is gemination (i.e. consonantal doubling). In Juba Arabic (see Miller and Woidich to appear), Bongor Arabic (see Luffin and Woidich to appear), and Romanian Pidgin Arabic (see Avram 2010) there is no gemination. On the other hand, consonantal doubling is attested but not common in Nubi Arabic (Wellens 2003). In GPA, however, gemination is common; examples are *fakkar* 'think', *šaggal* 'to play', and *sakkar* 'close'. The absence of gemination in the Arabic-based pidgins and creoles (with the exception of Nubi Arabic and GPA) as well as typologically less common phonemes such as the uvular /χ/ and the pharyngeal /ʕ/ could be evidence of the linguistic simplification that has often been named as a typical process in the emergence of pidgins and creoles (see Ferguson 1971,

Todd 1990, and Siegel 2008b). Table 5 below summarises the comparison of the phonological features in the Arabic-lexified pidgins and creoles.

feature pidgin/ creole	phonological reduction	Phonemic distinction between short and long vowels	gemination
Bongor Arabic	✓	x	x
Nubi Arabic	✓	x	x
Juba Arabic	✓	x	x
GPA	✓	x	✓
Romanian Pidgin Arabic	✓	x	x
Pidgin Madame	✓	✓	x

Table 5: A comparative account of the phonological systems of Arabic-lexified pidgins and creoles.

In the remainder of this section, I compare some morpho-syntactic features in the systems of the Arabic-lexified pidgins.

1.5.2.1 The Noun phrase in Arabic-based pidgins and creoles

While the exact outcome differs from language to language, the number of pronouns is reduced in all six Arabic-based pidgins and creoles. Compare for example, personal pronouns in GA (18) with those in GPA (5). None of the GA clitic pronouns is carried over from GA to GPA (as shown in more detail in 2.1.2.3). In Pidgin Madame, pronouns do not inflect for number or gender (Bizri 2010). Pronouns in Nubi, Juba, and Bongor Arabic inflect for number and person, but not for gender (see Wellens 2003, Miller and Woidich to appear, Luffin and Woidich to appear). The pronominal system of Romanian Pidgin Arabic is even more reduced. It only contains two pronouns: *ana* '1.SG' and *inte* '2.SG'.

The Arabic definite article is not carried over into any of the six Arabic-based pidgins and creoles. In the three African pidgins and creoles it is replaced with demonstratives such as *di*[17] 'this' (Luffin and Woidich to appear). No definite article is attested in Romanian Pidgin Arabic (see Avram 2010), GPA (see 2.1.2.2), and Pidgin Madame (see Bizri 2010).

None of the Arabic-based pidgins and creoles included in this comparative account show full agreement between the noun and the adjective. In Pidgin Madame, the singular feminine form for the adjective is used with all nouns (Bizri ibid). In GPA, Romanian Pidgin Arabic (Avram 2010), and Bongor Arabic (Luffin and Woidich to

---

[17] *Di* could be derived from the Arabic demonstrative for close feminine objects, /ðɪ/.

appear) the unmarked form is the third person singular masculine. In Juba Arabic and Nubi Arabic, on the other hand, the adjective agrees with the noun in number but not in gender (see Wellens 2003, Miller and Woidich to appear). Table 6 below summarises the above comparative account of the noun phrase in the six Arabic lexified pidgins and creoles.

feature / pidgin/ creole	Pronouns reduced?	DEF marker used?	Full N-ADJ AGR	
			Exists?	Unmarked form
Bongor Arabic	yes	no	no	SGM
Nubi Arabic	yes	no	no	SGM or PLM
Juba Arabic	yes	no	no	SGM or PLM
GPA	yes	no	no	SGM
Romanian Pidgin Arabic	yes	no	no	SGM
Pidgin Madame	yes	no	no	SGF

Table 6: A comparative account of the noun phrase of Arabic-lexified pidgins and creoles.

1.5.2.2 The verbal systems of the Arabic-lexified pidgins and creoles

The verb does not show full agreement with the noun in any of the six Arabic-based pidgins and creoles. For instance, Avram (2010) reports that in Romanian pidgin Arabic the Arabic noun is used for verbal function (e.g. *ani shugul* 'I work.N'). In Juba Arabic, the unmarked verb is the masculine third person singular or plural form (Miller and Woidich to appear). Bongor Arabic replaces the whole paradigm of Arabic verbal conjugations with the third person singular past form (Luffin and Woidich to appear). In Nubi Arabic, the unmarked form of the verb is the Arabic singular or plural masculine imperative (Wellens 2003). Pidgin Madame, on the other hand, uses the Arabic imperative and second and third singular feminine forms of the Arabic verb with all nouns. As regards GPA, I show in section 2.1.1.1 below that the verb does not show full agreement with the noun. The unmarked form is the GA third person singular masculine form.[18]

All Arabic-based pidgins and creoles – with the exception of Romanian Pidgin Arabic (Avram 2010) – use adverbs to mark for TMA. Note that this phenomenon is proposed to be a typical feature of creole languages (see 1.3 above). The lexemes that are recruited to show these temporal or aspectual distinctions might be different but have similar functions, compare for example the future markers *ja* 'come' in Nubi Arabic

---

[18] Other forms are also attested in my data such as the use of nouns for verbal functions, imperatives, and third person singular past form of the verb.

(Wellens 2003) with *baden* 'later' in GPA. Both of these two markers are used preverbally to mark for future tense.

Unlike what is supposed to be a general feature of pidgin languages, some Arabic-lexified pidgins and creoles actually have copulas. I show in section 2.1.1.5, for instance, that the copula *fi* 'in/there' is an innovation in the GPA verb phrase compared to the null-copula present tense verb phrase in GA. Similarly, the copula *fi* is used in Nubi Arabic (Wellens 2003) and Pidgin Madame (Bizri 2010). On the other hand, Juba Arabic and Romanian pidgin Arabic lack a copula in their morphological systems.

Table 7 summarises the above discussion on the verbal systems of the Arabic-based pidgins and creoles.

Feature / pidgin/ creole	Full S-V agreement		TMA adverbials	copula
	Exists?	Unmarked form		
Bongor Arabic	No	3SGM	✓	✓
Nubi Arabic	No	SGM or PLM IMP	✓	✓
Juba Arabic	No	3SGM or 3PLM	✓	x
GPA	No	3SGM	✓	✓
Romanian Pidgin Arabic	No	nouns for verbal function	x	x
Pidgin Madame	No	3SGF	✓	✓

Table 7: A comparative account of the morphological systems of Arabic-lexified pidgins and creoles.

It appears from the phonological and morpho-syntactic comparison above that the Arabic-based contact languages, similar to European-based pidgins and creoles (see 1.3), employ reduced linguistic systems. But I have also been able to show that the degree of reduction varies. Crucially, it seems to be greater in the less-stabilised pidgins. Compare, for instance, the number of subject pronouns in Romanian Pidgin Arabic (only two, see Avram 2010) with the more stabilised varieties Nubi Arabic (six, see Wellens 2003) and GPA (five, see Smart 1990).

No other classification can be made, neither based on the geographical location (e.g. African vs. Non-African Arabic pidgins/creoles), nor based on the existence vs. absence of native speakers in the contact language, the situation in which the contact language has emerged in, or the substrate languages. Indeed, tables 5, 6, and 7 show that GPA (a non-African workforce pidgin, with no native speakers, and with Indian substrate languages) is similar to Nubi Arabic (a nativised African military creole with African substrate languages) in having reduced phonology, reduced number of pronouns, lack of Subject-verb agreement, and noun-adjective agreement, but use of a copula and TMA

adverbials), whereas Juba Arabic (another nativised African military creole with African substrate languages) does not have a copula. Moreover, all of the investigated Arabic-based pidgins and creoles show similar features, regardless of differences in their substrate languages or the situations in which they have emerged in such as reduced derivation, lack of subject-verb and noun-adjective agreement, and limited number of pronouns. The fact that these features are also found across European language-based varieties could be an argument in favour of the Universalist theories of genesis discussed in section 1.1.1 above.

1.6 Previous Research on GPA

In this section, I review some of the studies that have been conducted on GPA over the past twenty years in chronological order. My PhD thesis is heavily indebted to these studies insofar as they have provided thorough descriptions of the phonology and morpho-syntax of GPA. Indeed, since I provide a linguistic description of the pidgin in a different section (2.1.2), the review below focuses on the theoretical claims made as regards the structural patterns of GPA as well as some of the methodological shortcomings of previous works. As will become apparent, the bulk of studies conducted on GPA thus far are descriptive and/or theoretical in nature. To date, no 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 morpho-syntax of GPA. The current thesis, therefore, attempts to fill this gap.

The pioneering report on GPA is Smart (1990), which provides an analysis of the phonology, morphology, and syntax of GPA.[19] The linguistic description is preceded by a discussion of the geographical and sociolinguistic situation in the Gulf. Smart claims that there was no mention of any Arabic-based pidgin or creole outside the African continent prior to his article. As far as I am aware, this is true, given that the two other Arabic-based pidgins spoken outside Africa – Romanian Pidgin Arabic and Pidgin Madame – were only reported later by Avram (1993)[20] and Bizri (2009). Smart's investigation is based on two newspapers published in the United Arab Emirates between the years 1986 and 1987. The source material heavily depends on newspaper articles written by Arabic speakers in GPA and published in the Al-Ittihad newspaper in addition to captions of cartoons published in the Al-Khaleej newspaper. Note that it is not common to find written material in GPA or even Non-standard Arabic. Yet, these

---

[19] Note that Smart (1990) refers to GPA as Gulf Pidgin.
[20] Note that Avram started collecting his Romanian Pidgin Arabic data in 1984 (Avram 1993).

materials are written in GPA by native Arabic speakers for the sake of humour. Depending on such data raises two methodological concerns: The first is the use of a written corpus to describe the phonological system of GPA. Bearing in mind the wide fit between Arabic orthography and phonology, such a strategy can risk providing insufficient and misleading data (see also 1.4 above). For example, short vowels are often not accounted for in written Arabic. Hence, the short vowel /ə/, for instance, in the GA word رجال *rədʒæl* 'man' is dropped in the Arabic orthography. This raises the question of how to interpret the vowel quality of sounds that are not represented in the written source. Furthermore, some letters can be pronounced in different ways. For example, the letter ج can be pronounced as a /dʒ / in some dialects of GA and /j/ in other dialects. Similarly, the letter ق can be pronounced in some dialects of GA either as /q/ or as /g/, while it is pronounced as /q/ or /dʒ/ in other dialects. The variation between /q/ and either /g/ or /dʒ/ seems to be relatively unconstrained. Thus, the word رفيقي (friend.my 'my friend') can be either pronounced as *rəfi:qi* or *rəfi:gi* in Najdi Arabic, a sub-dialect of GA, by the same person in different occasions. As such, it is hard to establish the phonology of Arabic or one of its lexified pidgins or creoles on the basis of written scripts. It is not surprising, then, that one finds some questionable examples in Smart's descriptive account of the phonology of GPA. For instance, he reports that instances of the long vowel [u:] are found only in two words in his data: دوكان *du:kan* 'shop' and كورة *ku:rah* 'ball'. Note that these two words are usually spelled in Standard Arabic with a dropping of the short vowel /ʊ/ in the first word and a dropping of the long vowel [u:] in the second. The rare occurrence of the long vowel [u:] in Smart's (1990) data suggests that the difference between the short /ʊ/ and the long [u:] vowels is not phonemic in GPA.[21] Indeed, the infrequent occurrence of the long vowel [u:] could be challenged by Naess's (2008: 42) own conclusion that '[v]owel length does not seem to be phonemic in GPA, as words with the same meanings are pronounced with both short and long vowels'. In other words, [u:] in GPA is an allophone rather than a phoneme. Smart's (1990) article thus illustrates the problems faced by pidgin and creole research that is based on written material. Note that Smart (1990: 87) himself shows 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]'.

---

[21] It could thus be the case that the two occurrences of the vowel /u:/ are a result of typos in the written material. Thus, typists could have misspelled the word for *shop* (i.e. دوكان *du:kan* instead of دكان *dukan*) and also misspelletd the word for ball (i.e. كورة *ku:rah* instead of the proper spelling كرة *kurah*).

The second methodological observation concerns the accuracy of the data used by Smart to describe GPA. The source material in Smart's article is problematic for two reasons: Firstly, the material used is jocular. Due to the fact that comedy typically relies on linguistic exaggerations, there is a high possibility that the data do not accurately represent the linguistic system of GPA. The second reason for questioning the accuracy of Smart's data is that the materials he bases his analysis on were not produced in spontaneous spoken interaction, ideally when GPA is the medium of communication between a local and an expat, or between two expats with two distinct first languages. Having been written by native speakers of Arabic for native Arabic readers, the material is at the risk of bearing frequent shifts to Gulf Arabic or even to Standard Arabic. Smart (1990: 114) reports that some of these shifts have actually been made purposefully to 'heighten the comic effect by contrast'. He describes some of the shifts in his data such as الجامعة العربية, *al-dʒamiʕɘh al-ʕarabjɘh*, 'The Arab League' and لقد واحد جول, *lagad wahid gul*, 'Someone has said...' In the first example, the definiteness marker *al-* is prefixed to both nouns (i.e. Arab and league), which is typical of GA and Standard Arabic, but not of GPA. The definiteness marker is normally dropped in GPA as detailed in section 2.1.2. In the other example, the Standard Arabic perfective marker لقد is used by the cartoonists in a GPA conversation.[22] In spite of these limitations, however, the results of Smart's analysis correspond to a great extent to the subsequent works on GPA discussed below.

The above methodological reservations illustrate the need to use spontaneous spoken data when describing a language, especially a contact variety. Thus, all the data in the current thesis – both for the descriptive account of GPA in section 2.1.2 below as well as for the quantitative analysis in Chapter 4 – are collected via interviews with informants who do not speak Gulf Arabic as their first language (refer to Chapter 4 for more details on the data used in this study).

Twelve years after Smart's (1990) paper, Wiswall (2002) conducted a comparative analysis of lexical borrowing in GPA as well as the use of the copula *fi*, 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. His subjects were divided into two groups, ten locals and nine expats. Members of the two groups were asked to translate sentences from

---

[22] As a speaker of GA and GPA, I can confirm Smart's (1990) claim that the perfective marker لقد is not carried over from Standard Arabic to GPA or even to GA and thus using it in a GPA conversation would make the jocular material even funnier.

English to GPA and then Wiswall compared the use of these linguistic features in the two groups. Since the three syntactic features above are typical to GPA only, GA speakers are expected to shift to the superstrate language (GA) when they speak GPA (i.e. drop the copula, less use of the verb *sawwi*, and less use of the possessive *mal* with pronouns). Wiswall's comparative investigation can be credited as the first study on GPA – and perhaps the only one thus far – which compares expats' and locals' linguistic production of GPA. In addition, Wiswall strove to maximise the accuracy of his data via a clearly defined speaker sample: His local informants had to meet certain conditions to qualify for participating in his study such as constant contact with expat GPA speakers. This eliminates the possibility of including non-GPA speaking subjects into his sample. The methodology implemented by Wiswall, however, raises some concerns. For instance, the 'observer's paradox', discussed in section 4.4.2, could have influenced the linguistic production of the participating informants, as they can be expected to be more careful about their linguistic production in a translation task (see Schäffner 2002 and Tennent 2005 for problems surrounding translation tasks).

Overall, Wiswall (2002) surprisingly finds that local native GA speakers use more of his selected morphosyntactic features of GPA (i.e. the copula *fi,* the possessive *mal,* and the command verb *sawwi*) than the expat Gulf Arabic speakers. We might hypothesise that this is due to what Labov (1966) refers to as *hypercorrect speech.* It could be the case that the polled GA-speaking informants overuse GPA features, trying to approximate the typical GPA speech. On the other hand, Wiswall's study suggests that lexical borrowings from English and the substrate languages of GPA are more frequent in the expats' translations to GPA. This is explainable by the fact that the majority of expat workers in the Gulf are multilingual. Indeed, some participants in the current study speak as many as four languages, see 4.4.1.1.

The third piece of research on GPA I will discuss here is Naess's (2008) MA dissertation, which provides a detailed account of GPA phonology, possession, negation, the copula *fi,* and the verb phrase. The description of GPA in Naess's dissertation is preceded by a theoretical discussion on the emergence of pidgin languages, in which she classifies GPA as a workforce pidgin. The phonetic and morpho-syntactic report is based on interviews with GPA speakers conducted by Naess in Buraimi, Oman and in the Emirati city Al-Ein. The use of fieldwork data gives Naess's work a thorough empirical grounding. Furthermore, unlike other works on GPA (such as Almoaily 2008, Alshammari 2010, and Bakir 2010), where informants are mainly from a lower class, the informants in Naess' study come from different socio-economic backgrounds. Her social

sample comprises equal numbers for each class (eight mid-class and eight low class participants) and each gender (eight males and eight females). A description that is based on interviews with female and male informants who come from various linguistic and social backgrounds is certainly one of the great advantages of this study. Despite this perfectly balanced sample, Naess does not, unfortunately, provide any variationist quantitative analysis. Moreover, the fact that Naess is not a speaker of Gulf Arabic herself could have affected the quality of her data. Note that the medium of communication between Asian workers in the Gulf and English speaking foreigners is typically English whereas GPA is normally used for communication between non-English speaking locals and expats (see Smart 1990). Indeed, the fact that Naess is not a speaker of Gulf Arabic herself could have made her informants code-switch to English more often than the norm. Naess (2008: 10) is aware of this impediment: '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'. In order to overcome this problem, Naess agreed with her informants to use GPA instead of English. Hence, although her fieldwork method elicited GPA data, the problem of code-switching is still unsolved, especially given its sub-conscious nature (see Myers-Scotton 1989, Edmondson 2004). Naess's (2008) data could have been more spontaneous if she had employed locals or expats to conduct the interviews. Indeed, in a study on Hawaiian Creole in which she faced the problem of modified speech when speaking to foreigners, 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'.

My MA dissertation (Almoaily 2008) provides a descriptive and quantitative account of GPA phonology and morphosyntax based on interviews conducted with Pakistani and Bengali expats in the city of Alkharj, located in the centre of Saudi Arabia. The fact that all informants are males, have similar low-status and relatively low education provided a good testing ground for the factor 'ethnicity' as a source for linguistic variation in GPA. The study provided the first quantitative analysis of GPA that uses data collected via interviews between a local speaker of GPA and expats. My findings suggest that ethnicity had little effect on the speech production of the sample. However, the results of my initial quantitative analysis can hardly be considered reliable because of the small sample size: there were only four participants in the study, two from Pakistan and two from Bangladesh and they only produced 4000 words each. In fact,

based on my initial findings, I have argued that the nationality of the GPA speaker might have had less effect on linguistic variation than the variation caused by first language or length of exposure to GA. To test some of these factors, the current study includes informants from three different linguistic backgrounds (Bengali, Punjabi, and Malayalam), who have been living in the Gulf for either five or less years or ten or more years. I also increased the number of informants to 16 and the overall corpus to 12,000 words, 4000 words per language group (refer to section 4.3 for more details on the corpus).

My 2008 MA dissertation also attempted to discover the locals' frequency of use of GPA when conversing with GPA-speaking expats in Saudi Arabia, the locals' self-professed competency in this pidginised form of GA, and the sociolinguistic status of this variety. To this aim, I conducted a questionnaire distributed to 77 Saudi students in the UK. Surprisingly perhaps, given that GPA seems to have evolved as a means of communication between GA speakers and Asian workers in the Gulf Area only over the past forty years, the variety is now widely used in the Gulf. More than 60% of participants in this questionnaire claimed that they do not find it difficult to communicate in GPA. Indeed, only 6% reported that it was very difficult to communicate in this variety, whereas 32% reported a slight difficulty (see table 8 below).

I find it difficult to communicate in GPA	No. of responses	Percentage
Yes, very difficult	5	6%
Yes, I somewhat difficult	25	32%
Yes, slightly difficult	35	45%
Not difficult at all	12	16%

Table 8: Locals' competence in GPA (Almoaily 2008: 57).

We can therefore assume that – certainly amongst the younger generation of Saudis – GPA has become a lingua franca. Locals regularly use it in communication with foreign workers, who use it for communication with foreign workers of a different linguistic background. This is supported by the respondents' answer to the question "I use GPA with speakers of Indian languages, but not with speakers of other Asian languages". Only 9% of informants claimed that they do not speak GPA at all (see table 9 below).

I use GPA with speakers of Indian languages, but not with speakers of other Asian languages		
	No. of responses	Percentage
Yes	21	27%
No	49	64%
I do not speak GPA at all	7	9%

Table 9: Frequency of use of GPA among locals (Almoaily 2008: 59)

Note that the tendency for pidgins and creoles to be regarded as inferior and low-status varieties by the speakers of the superstrate language, discussed in section 1.1.2 above, was also apparent in the respondents' answers to my survey. None of my seventy-seven respondents strongly agreed with the statement '*I support the permanence of GPA*'. Only one respondent agreed, whereas sixty-nine informants in the sample either disagreed or strongly disagreed. In fact, another question in my questionnaire revealed that only 16% of the polled sample considered GPA to be part of the Arabic language whereas 32% considered that GPA is not part of the Arabic language at all. The rest of the sample considered it a mixture of Arabic and other languages (see tables 10 and 11 below).

I support the permanence of GPA	No. of responses	Percentage
Strongly agree	0	0%
Agree	1	1%
I do not know	7	9%
Disagree	21	27%
Strongly disagree	48	62%

Table 10: Locals' attitudes towards the persistence of GPA (Almoaily 2008: 56)

I consider GPA to be:	No. of responses	Percentage
Part of the Arabic language	12	16%
A mixture of Arabic and other languages	40	52%
Not part of the Arabic language at all	25	32%

Table 11: Locals' attitudes towards classifying GPA as Arabic (Almoaily 2008: 56)

These findings support Smart's (1990) claim that GPA is a low status contact variety and that some of its temporary immigrant speakers suffer lack of respect by the indigenous people.

In a more recent paper, Bakir (2010) provides a description of the verbal system of GPA, based on fieldwork data collected in Doha, Qatar. Bakir's paper includes an interesting theoretical discussion on the pidgin status of GPA, where he lists some of the features GPA shares with other Arabic-based pidgins and creoles (refer to the discussion in 1.5 above for details). As far as I am aware, this study is the first detailed comparative analysis of GPA and other Arabic-lexified pidgins and creoles. The ten informants polled in Bakir's study come from several linguistic backgrounds, including Malayalam, Sinhala, Bengali, Tagalog, Tamil, and Hindi. Just like Naess's (2008) study above, there

was an equal distribution of gender. The length of stay of informant ranges between two and thirty years. Surprisingly, despite the diversity of gender, first language, and length of stay in the Gulf area in Bakir's sample, no attempt was made to tease apart the importance of social and developmental factors on language production. Hence, Bakir makes generalising statements about GPA which could be statistically supported or indeed disproven if a variationist analysis had been carried out on his data. For instance, Bakir (2010: 204) claims 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'. This claim could have been made so much stronger if Bakir had attempted to show a significant shift towards GA among the long-time residents. Indeed, my own analysis reported in this thesis, which investigates the relevance of length of stay on linguistic production (see Chapter 5), does not support Bakir's claim.

Finally, Alshammari (2010) provides a morpho-syntactic description of GPA based on the analysis of linguistic data produced by nine informants who come from three linguistic backgrounds, namely Pashtu, Bengali, and Malayalam. The description comes with a comparison of GPA against the typological features of pidgin languages. The informants interviewed in Alshammari's study were all males who spent between five and fifteen years in Saudi Arabia. The interviews were conducted in two cities of the north of Saudi Arabia, Hail and Sakaka. Reporting the existence of GPA in Sakaka, a city located in the far north of Saudi Arabia, is further evidence for the wide geographical distribution of this contact language. The fact that Alshammari is a native speaker of GA could have helped collecting spontaneous data. The phonemic transcription of the interviews in Alshammari's study offers the reader a better idea of the phonology of GPA (cf. 1.4 above). Findings of Alshammari's study support Smart (1990) insofar as the morpho-syntactic system of GPA complies with the typological features of Indo-European lexified pidgins such as analytic morphology and indication of tense via adverbs.

Notably, however, as in the other descriptive studies, Alshammari does not investigate potential variation patterns across the informants coming from different linguistic backgrounds.

In the next section I discuss the issue of language variation, particularly in contact languages.

1.7 Linguistic Variation

The study of linguistic variation in contact languages can make a valuable contribution to the field of (socio)-linguistic variation and change. Yet, within the field of pidgin and creole research – with the exception of a small number of studies, some of which I review in 1.7.2 below – linguistic variability is hardly ever investigated. For instance, while Avram (2010) reports on the existence of phonological variation in Romanian Pidgin Arabic, no quantitative variationist analysis of the variability is provided (consider also the description of the phonology and syntax of Juba Arabic in Miller and Woidich [to appear], where the unmarked forms are listed without any reference to their variants).

1.7.1 *Analysing linguistic variation*

Although native speakers of English or Arabic may find no difficulty in communicating with other native speakers of that language, a close look at any linguistic system reveals geographical, dialectal, sociolinguistic, stylistic, and even personal variations in the production of any language (see Weinreich, Labov, Herzog 1968, Labov 1972, Bailey 2002, inter alia). Tagliamonte (2006) asserts that linguistic variation is attestable in all levels, from phonetics to discourse (see also Chambers, Trudgill and Schilling-Estes 2004). A distinction between linguistic, when language-internal factors condition variability, and extralinguistic variation, when language-external (social or cognitive) factors have a bearing on language use, can be made here. In the discussion below I provide a brief historical sketch of the study of (socio)-linguistic variation.

Observations about language variation made by linguists, philologists, and anthropologists can be found in the literature from the late nineteenth century onwards. For instance, Schuchardt (1885) stated that variations are found in the speech of any individual. Sapir (1921: 38) also expressed his awareness of the failure of grammars to accurately account for the entirety of the language: 'Unfortunately or luckily, no language is tyrannically consistent. All grammars leak'. Until the 1950s, however, the general assumption was that linguistic variation is unpredictable (see Fries and Pike 1949, Weinreich et al. 1968). In other words, language variation was thought to be free from the influence of factors such as age, gender, social class, etc. The sociologist John Fischer (1958) was amongst the first, if not the first at all, to conduct a systematic study on the social conditioning of linguistic variation. His research on variation in the pronunciation of the suffix -ing (i.e. /ɪn/ versus /ɪŋ/) in the speech of New England schoolchildren lead him to postulate a link between social factors and the choice of

linguistic variables. But it is mainly the contributions of William Labov (and his collaborators) that have defined the study of linguistic variation since the 1960s and thus it is not surprising that the quantitative analysis of linguistic variation has often been referred to in the literature as the Labovian Paradigm (see Trudgill 1986, Hudson 1996).

While the variationist sociolinguistic paradigm continues to flourish, the systematic study and analysis of linguistic variation has been almost entirely restricted to a small number of languages, in particular to English and Canadian French (and to a certain extent also Portuguese). For example, in the introduction of his book *Variation and Change in Spanish*, Penny writes: 'The main aim of this book is to apply certain theoretical insights into linguistic variation and change (insights often derived from studies of English and other Germanic languages) to the Spanish-speaking world….' (Penny 2000: ix). Studies on the linguistic variation in non-Indo-European language such as Arabic (see Skousen 1989, Wahba 1996) and Korean (see Hong 1991) are few and far between. The upshot of this rather restricted focus is that we lack – to date – consolidated knowledge about linguistically and socially conditioned variability in non-European languages. The same is true for contact languages, and even more so for non-Indo-European contact languages.

1.7.2 *Linguistic variation in contact languages*

As discussed in section 1.1.2, one of the main differences between a jargon and a pidgin or a creole is that the first characteristically lacks linguistic rules whereas in the other two, community-wide speech conventions have been established. Hence, Singh (2000: 3) suggests that 'the people who use a jargon need to speak to each other but do not constitute a stable speech community who together develop and share consistent linguistic norms'. Pidgins and creoles, on the other hand, are usually described as having a lesser amount of inter-speaker variation due to the establishment of linguistic rules (see Mühlhäusler 1986, Singh 2000). Yet, as discussed above, linguistic variation is an integral part of the phonology, morphology, and syntax of every linguistic system. This implies that jargons possibly exhibit even more and – crucially – less principled (i.e. explainable by linguistic, cognitive, and social factors) variation than pidgins and creoles languages.

Given the fact that language variation also exists in contact languages, it is not surprising that quantitative analyses have been conducted on contact varieties, albeit to a lesser degree than on their lexifiers (see Guy 1981, Siegel 1982, Lucas, Bailey and Valli 2001, Mufwene 2001). Muysken and Smith (1995: 9) suggest that the extent of linguistic

variation in contact languages is more than it is in non-contact languages, for two reasons. The first is their swift change – as compared to non-contact languages – and the second is that pidgins and creoles are 'highly dynamic language systems and often coexist with their lexifier languages in the same speech community'. Patrick (2009: 469) concurs, claiming that 'the argument that systematic variation is greater among P/Cs during their development and expansion appears strong'. I would like to argue, however, that we simply do not have the empirical database to test the suggestion that contact languages are characterised by more linguistic variation than non-contact languages. An investigation of this hypothesis would need to be based on thorough variationist studies, not just on contact languages but also on the lexifiers of these languages. Indeed – since contact languages tend to encode less redundant morpho-syntactic information than their superstrate languages – it might actually be the case that comparing contact languages with their lexifiers reveals that contact languages are less varied than their lexifiers. These observations, however, are not meant to suggest that pidgin and creole languages are not good candidates for variationist analyses. Indeed, they might be valuable resources for variationist linguistics due to the reasons stated above (i.e. that contact languages normally evolve in multi-ethnic communities and that they develop rapidly). Yet, linguistic variation is still not accounted for in many studies on pidgins and creoles, as discussed at the beginning of this section. Note, however, that Muysken and Smith (1995: 13) argue quite the opposite, namely that 'many creolists have tended to put variation and change at the centre of attention'. This statement might be true for researchers at the variationist end of the spectrum of pidgin and creole studies such as Gillian Sankoff, John Rickford, Peter Patrick, Suzanne Laberge, and Miriam Meyerhoff, who endeavour to apply the Labovian Paradigm to contact languages. For instance, in one of the earliest variationist studies on pidgin and creole languages, Sankoff and Laberge (1974) examined a corpus of 234 examples of the Tok Pisin adverb *bai* and its variant *baimbai*. They found that *baimbai* occurred only once in the data of the 'adults' group and was never produced by any of the members of the 'children' group. This suggests a potential historical replacement of the adverb *baimbai* by the shorter form *bai*. The data also revealed that the adult members polled in the study tend to stress the adverb *bai* more than the members of the 'children' group (27.8% vs. 49.6% at the stress level 3). Evidence of a significant relation between age and stress in Tok Pisin presented in this study has opened the door to subsequent variationist analyses investigating the development of pidgin and creole languages. For example, Sankoff and Brown (1976) analysed linguistic variation in the use of *ia* 'here' in Tok Pisin as a focus marker or as a

relative marker, separating the main clause from the relative clause. The examination of corpus data collected from 26 informants revealed more usage of *ia* as a focus marker among the younger generation in the sample, suggesting a diachronic shift from a relative marker to a focus marker. Similarly, in a more recent study, Sidnell (2002) examined the use of the imperfective aspect in Guyanese Creole. The data set, which was collected in a rural Indo-Guyanese village of about seven-hundred inhabitants and from a nearby settlement, comprises of a total of 1001 occurrences of the imperfective *a* or one of its variants, *doz, Ø, Ø+V+ing* , and *yuustu*. The study showed significant differences in the selection of variables between 'working' and 'non-working' groups. The first group preferred the imperfective *a,* whereas the latter preferred the variant *doz.* Other studies on (socio)-linguistic variation in pidgins and creoles can be found in Aceto and Williams (2003), Deumert and Durrleman (2006), Meyerhoff (2008), and Patrick (2009). Indeed, it appears that there has been valuable cross-pollination between variationist and creolist research. For instance, Patrick (2009: 461) claims that concepts such as 'implicational scaling, grammaticalisation, and the variable rule framework... have been introduced to variationist work or flowered there in the consideration of P/C data'.

The evidence of a significant effect of factors such as age, socio-economic standing, and place of residence on linguistic variation in pidgins and creoles gives good reason for conducting a quantitative study on GPA. Such an analysis would allow me to test, for instance, Bakir's (2010) claim that Asian workers in the Gulf shift to GA after spending some time there. If findings of my quantitative analysis prove that Bakir's claim is true, then the variety under question might be considered a pidgin for the newcomers and a depidginised variety for the long-staying workers.

Note however that – while variationist analyses on contact languages date back to the 1970s – it is difficult to find studies on linguistic variation in non-European language based pidgins and creoles. Examples of the rare quantitative accounts made on language variation in non-typical pidgins and creoles are Siegel's (1982) comparative study of lexical variation in Pidgin Fijian and Wiswall's (2002) study on variation in GPA, detailed in 1.6 above. Generally, however, researchers have merely reported the existence of linguistic variation but refrain from conducting quantitative analyses of the factors that condition the heterogeneity (see Samarin 1986, Mesthrie 2002, and Avram 2010). This lack of quantitatively accountable analyses on non-Indo European pidgins calls for more research on linguistic variation in un-typical contact languages. Hence, in this study I attempt to provide a quantitative analysis which aims to discover the potential effect of the two factors *the speakers' first language* and *the number of years spent in the Gulf* on

variability in GPA morpho-syntax (see Chapter 4 for a more detailed discussion on these two factors, the data, and the linguistic variables analysed in this thesis).

In the next chapter, I provide a morpho-syntactic description of selected linguistic features of GPA and GA, the superstrate language of the pidgin under investigation. For each feature, I exemplify for the linguistic variable and its variants.

Chapter 2: GA and GPA, Definition and Description

This chapter provides a description of selected aspects of the morpho-syntax of GPA, the pidgin under investigation in this project, and its lexifier, GA.

2.1 Description of GA and GPA

The description of the languages in contact in this chapter and in the subsequent chapter will be restricted to the morpho-syntactic features, which I will consider in my quantitative investigation (see section 4.1). These features are: agreement in the verb phrase and in the noun and adjective phrase, definiteness and indefiniteness, pronouns, coordination, and copular verbs. I will discuss each in turn in the following sub-sections.

2.1.1 *Linguistic description of selected features of Gulf Arabic*

2.1.1.1 Agreement

(A) Verbal agreement

In Gulf Arabic, the verb inflects for gender, number, tense, and person. In this section I describe agreement following Feghali's (2004) classification, which distinguishes between three forms of the GA verb (see also Holes 1990, Qafisheh 1977 for alternative classifications of the GA verb). All examples below are provided by myself, a native speaker of GA.

Importantly, the GA verb root and tense determine the way in which subject-verb agreement is applied in gender, number, and person. Hence, for every verbal classification below, there is a brief illustration followed by a table listing its various conjugations (i.e. the form the verb takes in agreement with the noun in number, gender, and person). There are two tables for every classification, one for past forms of the GA verb and another for present forms.

Note that there are some considerations that are worth mentioning. All examples illustrate the most common regular subject-verb agreement forms in GA. It should be noted as well that irregular forms are also found in GA. Furthermore, the description provided here is not meant to be exhaustive. It only accounts for major dialectal variation in GA. Other GA dialects may have different forms to those shown below. Hence, the purpose of the examples in this section is not to list all GA subject-verb agreement forms but rather to exemplify how the GA verb inflects for number, gender and person.

Let us now have a look at the structure of the GA verb starting with verbs of a three consonant root then verbs of a four consonant root.

1- Verbs of a three-phoneme verb root (tri-phonemic verbs)

In Arabic, verbs typically have three consonants as their root. Various verb conjugations are applied via the addition of consonants and vowels to the verb root (Schramm 1962, Cavalli, Soudi, and Mitamura 2000). According to Feghali (2004), the tri-phonemic past form of the Gulf Arabic verb is applied via prefixes or suffixes attached to the verb in addition to infixed consonants and vowels. Hence, the tri-phonemic GA verb falls into three classes, each of which takes different present, past, future, and imperative forms. These three classes are strong, doubled, and weak verbs. The subsections below explain these three classes and provide examples for them:

a) Strong verbs: these are the verbs whose root is composed of three consonants where the last two consonants are not alike (e.g. d-r-b 'to hit', and r-b-t 'to tie'). Table 1 shows how the various forms of the strong verb j-l-s 'to sit' agree with the subject in number, gender, and person.[1] The transliterations in the third column represent two of the common ways to pronounce the GA verb.

Subject	Arabic derivation	Transliteration	Meaning
1SG	جلست	*jalas-t*	'sat-1SG'
1PL	جلسنا	*jlas-na/ jalas-na*	'sat-1PL'
2SGM	جلست	*jlas-t*	'sat-2SG.M'
2SGF	جلستي	*jalast-i/ jlast-i*	'sat-2SG.F'
2PL	جلستو	*jlas-tu/ jalas-tu*	'sat-2PL'
3SGM	جلس	*jalas/ jlas*	'sat.3SG.M'
3SGF	جلست	*jlis-at/ jlas-at*	'sat.3SG.F'
3PLM	جلسو	*jlis-aw/ jlas-aw*	'sat-3PL.M'
3PLF	جلسن	*jlas-an/ jlis-an*	'sat-3PL.F

Table 1: Various agreement forms of the GA verb stem (j-l-s) in the past tense (adapted from Feghali 2004).

Applying the inflections above to other GA strong verbs, for instance g-ʕ-d 'to sit', can be done by replacing the three consonants of the stem and retaining the rest of consonants and phonemes in analogy with the j-l-s stem exemplified in table 1. Hence, the third person plural masculine form of the verb root *g-ʕ-d* is *gʕid-aw* or *gʕad-aw* and so on. Similarly, this procedure of adding prefixes, suffixes, and affixes to the stem applies to all GA verb forms listed in the tables from 2 to 6 below.

Note that in the present tense, the verb stem j-l-s takes different forms from those stated in table 1 above. But the procedure which creates these forms is still the same: prefixation, suffixation, and/or infixation. The various forms are shown in table 2 below.

---

[1] Note that other, less common, dialectal possibilities are not mentioned (e.g. *jilisti* and *jlisti* 'sit-2SGF').

Subject	Arabic Conjugation	Transliteration	Meaning
1SG	أجلس	*a-jlis*	'PRS.1SG-sit'
1PL	نجلس	*na-jlis/ ni-jlis*	'PRS.1PL-sit'
2SG.M 3SG.F	تجلس	*ta-jlis/ ti-jlis*	'PRS.2SG.M-sit' 'PRS.3SG.F-sit'
2SG.F	تجلسين	*ta-jlis-iin/ ti-jlis-iin*	'PRS.SGF-sit-2SG.F'
3SG.M	يجلس	*ya-jlis/ yi-jlis*	'PRS.2SG.M-sit'
2PL	تجلسون	*ta-jlis-uun/ ti-jlis-uun*	'PRS.2PL-sit-2PL'
3PL.M	يجلسون	*ya-jlis-uun/ yi-jlis-uun*	'PRS.PLM-sit-PLM'
3PL.F[2]	يجلسن	*ya-jlis-an/ yi-jlis-an*	'PRS.PLF-sit-PLF'

Table 2: Various agreement forms of the verb stem (j-l-s) in the present tense (adapted from Feghali 2004).

In the future tense, the future marker /b/ is attached as a prefix to the forms listed in table 2 (i.e. present tense forms). Thus, the future form for *ajlis* is *b-ajlis* 'I will sit', the future form for *najlis* is *b-n-ijlis* 'we will sit', and so on.

Importantly, the conjugations listed in tables 1 and 2 above are of the GA strong verb. Verbs of other verbal stems may take different conjugations. These verbal categories are doubled and weak verbs. The rest of this sub-section will highlight these verbs as illustrated by Feghali (2004).

b) Doubled verbs: These are verbs in which the middle consonant of the tri-consonantal stem is doubled (i.e. the middle and the last consonants in the stem are one doubled consonant). Table 3 below shows how the verb ʕ-d-d 'to count' agrees in person, gender, and number with its subject in the past tense:

Subject	Arabic Conjugation	Transliteration	Meaning
1SG 2SG.M	عديت	*ʕadd-eet*	'counted-1SG' 'counted-2SG.M'
1PL	عدينا	*ʕadd-eena*	'counted-1PL'
2SG.F	عديتي	*ʕadd-eeti*	'counted-2SGF'
2PL.M	عديتو	*ʕadd-eetu*	'counted-2PL.M'
2PL.F[3]	عديتن	*ʕadd-eetin*	'counted-2PL.F'
3SG.M	عد	*ʕadd*	'counted-3SG.M'
3SG.F	عدت	*ʕadd-at*	'counted-3SG.F'
3PL.F	عدّن	*ʕadd-an*	'counted-3PL.F'
3PL.M	عدّو	*ʕadd-u*	'counted-3PL.M'

Table 3: agreement between doubled verbs and their subjects in the past tense (adapted from Feghali 2004).

---

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

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

In the present tense, the GA doubled verb has different forms to the ones illustrated in table 3. These verb forms are shown in table 4:

Subject	Arabic Conjugation	Transliteration	Meaning
1SG	أعدّ	*a-ʕidd*	'PRS.1SG-count'
1PL	نعدّ	*ni-ʕidd*	'PRS.1PL-count'
2SG.F	تعدّين	*ti-ʕidd-iin*	'PRS.2SG.F-count-2SG.F'
2SG.M 3SG.F	تعدّ	*ti-ʕidd*	'PRS.2SG.M-count' 'PRS.3SG.F-count'
2PL.M	تعدّون	*ti-ʕidd-uun*	'PRS.2PL.M-count-2PL.M'
2PL.F[4]	تعدّن	*ti-ʕidd-an*	'PRS.2PL.F-count-2PL.F'
3SG.M	يعدّ	*yi-ʕidd*	'PRS.3SG.M-count'
3PL.F	يعدّن	*yi-ʕidd-an*	'PRS.3PL.F-count-3PL.F'
3PL.M	يعدّون	*yi-ʕidd-uun*	'PRS.3PL.M-count-3PL.M'

Table 4: Agreement between doubled verbs and their subjects in the present tense (adapted from Feghali 2004).

c) Weak verbs: In this category of GA verbs, the root contains at least one vowel. These verbs take different forms to those shown in tables 1-4 above. Weak verbs can be further subcategorised into three subtypes, namely defective, hollow, and hamzated verbs. Just as strong verbs, all weak verbs inflect for tense, person, number, voice, and gender. The root of a GA defective verb either begins or ends with the semivowel /w/ (e.g. w-s-l 'to arrive') or ends with a short /a/ (e.g. g-r-a 'to read'). Hollow verbs are verbs whose root is composed of the vowel /a/ followed and preceded by a final consonant (e.g. j-a-b 'to bring). Finally, the root of a hamzated verb consists of at least one glottal stop[5] (Feghali 2004). Table 5 below exemplifies the inflection of the GA weak defective verb w-s-l for person, number, gender, both in the present and past tenses:

Subject	Present	Past	Meaning present/ past
1SG	*oo-sal*	*wi-sal-t*	'I arrive/ I arrived'
1PL	*noo-sal*	*wi-sal-na*	We arrive/ we arrived'
2SG.M 3SG.F	*too-sal*	*wi-sal-t* *w-sal-at*	'you/she arrive(s)/ I/she arrived'
2SG.F	*to-sal-ain*	*wi-sal-ti*	'You-SGF arrive /arrived'
2PL	*to-sal-oon*	*wi-sal-to*	'you-PL arrive/ arrived'
3SG.M	*yoo-sal*	*wi-sal*	'he arrives/ he arrived'
3PL.F	*yoo-sal-an*	*w-sal-an*	'they-F arrive/ arrived'
3PL.M	*yo-sal-oon*	*w-sal-aw*	'they-M arrive/ arrived'

Table 5: Subject-verb agreement in the GA defective verb (adapted from Feghali 2004).

---

[4] For many GA speakers, the 3.PL.M form is used interchangeably with this form in a random manner.
[5] Note that in many cases, the glottal stop in GA is converted to the short vowel /a/.

2- Verbs of a four-consonant verb root (Quadriliteral verbs):

Just like the GA verbs of a tri-consonantal verb root, quadrilateral verbs inflect for person, gender, and number. Table 6 demonstrates various inflections for the verb root *χ-r-b-ṭ* 'to mess up objects'. Since the purpose of this discussion is to show that agreement exists at all GA verb levels, only one example for subject-verb agreement for GA quadrilateral verbs will be provided. More details on Arabic quadrilateral verbs can be found in Qafisheh (1977), Holes (1990), and Zadeh and Winder (2003).

Subject	Present	Past	Meaning
1SG	*a-χarbiṭ*	*χarbaṭ-t*	mess-1SG
1PL	*n-χarbiṭ*	*χarbaṭ-na*	mess-1PL
2SG.M	*t-χarbiṭ*	*χarbaṭ-at*	mess-2SG.M
3SG.F		*χarbaṭ-t*	mess-3SG.F
2SG.F	*t-χarbiṭ-iin*	*χarbaṭ-ti*	mess-2SG.F
3SG.M	*y-χarbiṭ*	*χarbaṭ*	mess-3SG.M
2PL	*t-χarbiṭ-uun*	*χarbaṭ-tu*	mess-2PL
3PL.F	*yi-χarbiṭ-in*	*χarbaṭ-an*	mess-3PL.F
3PL.M	*y-χarbiṭ-uun*	*χarbaṭ-taw*	mess-3PL.M

Table 6: Subject-verb agreement in GA quadrilateral verbs (adapted from Qafisheh 1977).

This section has shed light on subject-verb agreement in the GA verb. It is clear from the examples in tables 1 to 6 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 noun phrase and adjective phrase.

(B) Agreement in the NP and in the ADJP

This section highlights agreement in GA between the noun and the adjective, between the noun or adjective and the numeral, and between the noun or adjective and the demonstrative:

1- Agreement between the noun and the adjective(s)

In GA, the adjective or set of consecutive adjectives in the same noun phrase agree with the head noun in gender, number, and definiteness (cf. Feghali 2004, Smart 1990, Almoaily 2008). One important difference between agreement in number in the VP and in the NP and the ADJP shall be mentioned here. In the VP verbal agreement inflects for number as singular versus plural, whereas in the NP and in the ADJP there are three inflections for number: singular, dual, and plural. Examples (1-4) below provide an

illustration of noun-adjective agreement forms in GA for the noun *tˤalib* 'student' and the adjective *jidiid* 'new'. For every example, there are three lines: the first is the GA excerpt, the second is the gloss, and the third is the meaning in English:

(1) tˤalib-ah jidiid-ah[6]

    Student-SGF.INDEF new-SGF.INDEF

    'A new female student.'

(2) itˤ-tullab il-judad

    DEF-student.PL.M DEF-new.PL.M

    'The new male students.'

(3) tˤalib-tain jidid-tain

    Student-DUF.INDEF new-DUF.INDEF

    'Two new female students.'

When combining two head nouns of different genders, the adjective is in the masculine form. Agreement in number and in definiteness is still in effect, as shown in (4) below:

(4) it-tˤullab w=it-tˤalib-at il-judad

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

    'The new male and female students'.

The noun/adjective agreement with the number in GA takes a slightly different form to the noun-adjective agreement discussed here. Hence the following subsection is dedicated to describing agreement between the noun/adjective and the number.

2- Agreement between the noun/adjective and the number

In GA, numbers from one to ten inflect for gender, while numbers larger than ten always come in one form (a masculine-like form).[7] In table 7 I list GA feminine and masculine cardinal numbers from one to eleven.

---

[6] Refer to section 1.3.1for an illustration of the presentation of examples in this thesis.
[7] Lacking the feminine suffix –ah (e.g. *χamsiin walad w=χamsiin bint* 'fifty boys and fifty girls').

No.	Masculine		Feminine	
1	واحد	*waḥid*	وحده	*waḥdah*
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	*ḥdaʕaš*	حدعش	*ḥdaʕaš*

Table 7: GA cardinal numbers

Feghali (2004) suggests that the feminine forms of GA numbers (three to ten) quantify masculine nouns and adjectives, whereas masculine numbers quantify feminine nouns or adjectives. This seems to be true as a general rule. However, as a native speaker of GA, I can attest that using the masculine form of the numbers from one to ten with a masculine noun or adjective and the feminine form with the feminine noun or adjective is still accepted and indeed widely used.

The quantifying number and the quantified can either be definite or indefinite. If the quantified element is singular or dual, the definiteness marker (discussed in section 2.1.1.2 below) is attached to the optional number and to the quantified element, as in (5) and (6) below. The definiteness marker, however, is attached to the number only if the quantified element is plural, as in example (7).

(5) el-yad   (el-waḥd-ah)   ma   t-saffig

DEF-hand.SG   (DEF-one-F)   NEG   PRS-3SG.F.clap

'The one hand does not clap' (An Arabic proverb).

(6) šarērt   el-liʕbi-tēn   (eθ:-intēn)

Buy.1SG.PST   DEF-toy-DU   (DEF-two)

'I bought the two toys'

(7) el-arbaʕ-at   aglam   ðˤaʕaw

DEF-four-F   pen.PLM.INDF   loose.PST.3PL.M

'I lost the four pens'

If the quantified element is indefinite, the definiteness marker is neither attached to the number nor to the quantified element, as in (8) and (9) below. In sentences like (5), (6), and (8), where the number is redundant, the number is normally dropped both in definite and in indefinite phrases.

(8) rajjal-ēn   (ithnēn)

   man-DU   (two)

   'Two men'

(9) sitt   ban-at

   six   girl-PLF

   'Six girls'

Thus, as shown in examples (5-7) above, the quantified lexeme in GA only agrees with the number in definiteness if it is singular or dual. Similarly, quantified lexemes do not always agree with the numeral in number. If the quantified noun is between three and ten it comes in the plural form as exemplified in (10). If the number is larger than ten, however, the quantified comes in the singular form as in (11) (Almoaily 2008):

(10)   iθ-alaθ-at   mudarris-iin

      DEF-three-F   INDF.teacher-PLM

      'The three teachers'

(11)   el-ʔišriin   mudarris

      DEF-twenty   INDF.teacher.SGM

      'The twenty teachers'.

This subsection has discussed agreement between the numeral and the quantified element in GA. As has become clear from the examples (5) to (11), agreement between the numeral and the quantified in number, gender, and definiteness is rather complicated, since it exists in some cases and is absent in others. The following subsection shall highlight agreement between the noun or adjective and the demonstrative.

3- Agreement between the noun/adjective and the demonstrative:

The GA demonstrative pronoun inflects for number, gender, and proximity (Qafisheh 1977, Holes 1990). More precisely, the singular forms of the GA demonstrative pronoun inflect for gender and proximity (*haða* 'this.M' and *haði* 'this.F' vs. *haðak* 'that.M' and *haðik* 'that.F').

Examples (12-15) illustrate the single GA demonstrative inflection for gender and proximity:

(12)   haða   l-kita:b

      this.MSG.PROX   DEF-book.MSG

      'this book'

(13) haðak   l-kitab
     this.MSG.DIST   DEF-book.SG.M
     'that book'.

(14) haði   i-š:ašah
     this.SG.F.PROX   DEF-screen.SG.F
     'this screen'

(15) haðik   i-š:aš-ah
     this.SG.F.DIST   DEF-screen-SG.F
     'that screen'.

The plural forms of the GA demonstrative, on the other hand, inflect for proximity, but not for gender. Thus, both the plural demonstrative for proximate objects *haðoli* and the plural demonstrative for distant objects *haðolik* are used with both feminine and masculine objects. Consider examples (16) and (17) below:

(16) haðoli   el-awlad
     this.PL.PROX   DEF-boy.PL.M
     'these boys'.

(17) haðolik   i-š:aš-at
     this.PL.DIST   DEF-screen-PL.F
     'those screens'.

The fact that the verb root determines the verbal inflection in the GA verb has left us with tens of verbal conjugations in GA. Moreover, agreement in the GA noun phrase and adjective phrase involves a large number of other conjugations. As discussed in section 1.3.2, we would expect this elaborate agreement system in GA to be one of the major sites for restructuring in GPA.

The next section discusses definiteness in GA, which involves another affixation in the GA morphological system.

2.1.1.2 Definiteness and indefiniteness

Definiteness: As discussed above, the marker for definiteness in GA is the prefix *al-*[8], exemplified in (18) below. Feghali (2004) suggests other, indirect, ways of expressing definiteness in GA. For instance, when an indefinite noun is specified by a

---

[8] Since the consonant /l/ in the prefix *al-* assimilates with the following consonant, there are various forms of GA definite marker depending on the following consonant (e.g. al-, el-, and /ɪ /followed by a geminated consonant).

definite noun, i.e. the so called *construct state*[9], the indefinite noun(s) is/are considered definite as in example (19) below. Definiteness can also be expressed by adding a pronominal clitic to the indefinite element as in (20). Finally, an indefinite noun is considered definite when it is specified by a proper noun as in (21). Hence, the nouns in (19-21) below are interpreted as definite although they are structurally indefinite:

(18)   el-gitˁar

          DEF-train

          'the train'

(19)   suug   l-khudrah

          market   DEF-vegetable.PL

          'the vegetables market.'

(20)   siyart-i

          car-1SG.POSS

          'my car'

(21)   matˁa:r   Jeddah

          airport   Jeddah

          'The airport of Jeddah'

Indefiniteness: In GA, the absence of the definiteness marker *al-* from the noun or adjective typically marks indefiniteness, as shown in example (22) below. Other ways of expressing indefiniteness have been reported by Feghali (2004). For instance, when the definite element is preceded by the word *aħad* 'one of/ someone' or its feminine form *iħda,* the definite noun is considered to be indefinite as exemplified in (23) below. Indefiniteness can also be expressed by adding the phrase *waħid min* 'one-M of' or its feminine form *waħd-ah min 'one-F of'* to the definite element as illustrated in (24):

(22)   gitˁar

          train

          'a train'

(23)   aħad   at-tˁullab

          one of.M   DEF-student.PL

          'One of the male students'

---

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

(24)   waħid   min   el-Kuwait

        Someone.M   from   DEF-Kuwait

        'A male from Kuwait'

This section has provided a discussion of definiteness and indefiniteness in GA. It has been argued in this section that the only overt marker for definiteness in GA is the prefix *al-* but definiteness in GA can also be expressed via other means such as *idafah* 'addition' or pronominal clitics. There are no overt markers for indefiniteness in GA. However it can be expressed indirectly.

The next section discusses subject, object, and possessive pronouns in GA.

2.1.1.3 Personal pronouns

This section briefly describes and exemplifies the GA subject, object, and possessive pronouns. GA personal pronouns inflect for number, person, and gender. Subject pronouns in GA are free morphemes whereas object and possessive pronouns are always bound morphemes. In the following three subsections, I shall list the various gender, number and person inflections for the GA personal pronouns:

A) Subject pronouns:

Table 8 lists the forms of the GA subject pronon:

Pronoun	Transliteration	Examples
1SG	*Ana*	*ana hina* 'I am here'
1PL	*ħinna (inna)*	*ħinna hina* 'We are here'
2SG.M	*int (inta/ anta)*	*int hina* 'you-SGM are here'
2SG.F	*inti (anti)*	*inti hina* 'you-SGF are here.'
2PL	*intum (into)*	*intum hina* 'you-PL are here.'
3SG.M	*Hu*	*hu hina* 'he-SGM is here.'
3SG.F	*Hi*	*hi hina* 'she-SGF is here.'
3PL.M	*Hum*	*hum hina* 'they-PLM are here.'
3PL.F	*Hn*	*hin hina* 'they-F are here.'

Table 8: GA subject pronouns (adapted from Feghali 2004).

Importantly, subject pronouns in GA are optional (refer to examples 26 and 28 below). The subject pronoun is generally silent, except if it is in focus or if it is a shifted topic. This is linguistically known as subject *pro drop*, which exists in many languages including Chinese, Spanish, Persian, and Italian. Please refer to Fassi Fehri (1993) for more on pro drop in Arabic.

B) Object and possessive pronouns

Unlike subject pronouns, these pronouns are suffixes attached to the verb, noun, active participles, and to particles (Qafisheh (1977). When attached to nouns, they indicate possession, and when attached to verbs, active particles, or prepositions they function as object pronouns. Table 9 below lists GA object and possessive pronouns:

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

Table 9: Object and possessive pronouns in GA (adapted from Feghali 2004).

The discussion in the last few sections has revealed that GA has an elaborate pattern of agreement in the verbal and nominal complex whereby affixes and infixes are used to mark for gender, number and definiteness. In terms of personal pronouns, whereas subject pronouns are free, object and possessive pronouns in GA are suffixes attached to the noun or to the verb. The following two sections look at coordinating conjunctions and copulas, which are free morphemes, in GA.

2.1.1.4 Coordinating conjunction

According to Feghali (2004), GA has a range of coordinating conjunction markers. The most commonly used amongst them are *wa* 'and', *laakin* 'but', and *aw* 'or'. But there are also less common coordinating conjunction markers such as *f(a)-* 'then', *willa* 'or', *ya ... ya..* 'either.. or..', *bass* 'but', and *la... wala* 'not.. nor..'. Below is a brief discussion of these GA conjunction markers

I- w(a[10]) *'and':* this marker is used to link words, phrases, clauses, as well as sentences (Feghali 2004). As can be seen in example (25) below, *w(a)* is a prefix attached to the linked element. In multiple coordination, as in (26), *wa* is attached to every linked element.

(25) laʕab   Ahmad   w=sajjal   goal

      played   Ahmad   and=scored   goal

      'Ahmad played and scored a goal'

(26) šif-t   Ahmad   w=ʕali   w=Nasser

      Saw-1SG   Ahmad   and=Ali   w=Nasser

      'I saw Ahmad, Ali, and Nasser.'

II- Laakin *'but'***:** This conjunction marker is used to link sentences, phrases, and clauses, but not words. Suffixed pronouns, discussed in section 2.1.1.3 above, can be optionally attached to this conjunction marker. They cannot be attached to any other GA conjunction marker (Feghali 2004):

(27) šif-t   l-mbarah   laakin-(ni)   ma   šift   l-goal

      Saw-1SG   DEF-match   but-(1SG.OBJ) not   saw   DEF-goal

      'I saw the match but I didn't see the goal.'

III- bass 'but'**:** *bass* is synonymous to *laakin*. The main difference between them is that suffix pronouns can be attached to *laakin*, but cannot be attached to *bass*:

(28) ʕazam-ni   Khaled   bass   ma   riħ-t

      invited-me   Khaled   but   not   go-PST.1SG

      'Khaled has invited me but I didn't go'.

IV- Aw and **willa:** These two markers are synonymous, both meaning *or*. They can be used to link words, phrases, clauses, and sentences and are used repeatedly in multiple conjunctions (Qafisheh 1977). Consider examples (28) and (29) below:

(29) Hatha   l-maħal   yisakkir   is-saʕah   tisiʕ aw ʕašir?

      This   DEF-shop   closes   DEF-hour   nine   or   ten?

      'Does this shop close at nine or at ten?'

---

[10] *Wa* can be reduced in rapid speech and pronounced as /w/.

(30) taʕal   el-jimʕah   aw   is-sabt   aw   il-aħad

      come   DEF-Friday   or   DEF-Saturday or   DEF-Sunday

      'I can come on Friday, Saturday, or, Sunday'.

V- f(a)-: This coordinating conjunction is rarely used in GA. It links two sentences to indicate the occurrence of two consecutive actions (Feghali 2004). Thus, in example (30) below, the second action, going with the Khalid, occurred immediately after the speaker was asked by Khalid to go with him:

(31) Khalid   gal-l-i   taʕal   f-riħ-t   maʕ-ah

      Khalid   said-to-me   come   and-go-PRS.1SG   with-him

      'I went with Khalid soon after he asked me to go with him.'

VI- ya.. ya.. 'either.. or': This conjunction is used to indicate a choice between two possible options. It is used with words, phrases, clauses and with sentences. In (31), it is used to conjoin two sentences:

(32) ya   n-ruuħ   li-rriyadh   ya   n-ijliis   hina

      either PRS.1PL-go   to-riyadh   or   PRS.1PL-stay   here

      'We either go to Riyadh or stay here'

VII- la... wala.. 'not.. neither/ no.. and not': According to Qafisheh (1977), this coordination conjunction is used to coordinate two negative elements as in (33). When the first of these elements is a verb, then the second is a consequence of that verb, as demonstrated in (34).

(33) la   t-ru:ħ   la   minna   wa=la   minna

      No   PRS.2SG-go   Not   this.way   and=not   this.way

      'Don't go anywhere'

(34) la   titsawwag   kiθiir wa=la   int   b-mðˤayyeʔ   floos-ik

      no   shop   a lot   and=not   you   FUT-lose   money-your

      'Don't do a lot of shopping and you will not lose your money'.

      To summarise, in this section we looked at eight GA coordinating conjunction markers. Apart from two markers, *wa-* and *fa-,* all GA conjunction markers are free morphemes. The next section discusses copular verbs in GA.

2.1.1.5 Copula

According to Holes (1990) The GA copula is used overtly only in past and future sentences whereas it is covert in the present tense. The copulas *kaan* 'be.PST' and *sˤaar* 'became' agree with the subject in number and gender. This agreement, however, is subject to sociolinguistic variation. According to Holes (1990), agreement between the GA copula and the subject is more common in the speech of educated people, whereas in the speech of less educated people we tend to find one invariant form of the copula, the first person singular masculine. Examples (35) to (38) display the presence of the copulas *kaan* and *sˤaar* in the past and future tenses, while example (39) demonstrates copula absence in the present tense:

(35)   Kaan-at   Sarah   mariiðˤ-ah   ams
     COP.PST-SG.F   Sarah   sick-SG.F   yesterday
     'Sarah was sick yesterday.'

(36)   B-itikuun   hina   bukrah?
     FUT- COP.2SG.M   here   tomorrow?
     'Are you going to be here tomorrow?'

(37)   sˤaar   el-labtob   gideem
     COP.PST.SG.M   DEF-laptop   old.M
     'The laptop became old.'

(38)   B-yi-sˤiir   el-labtob   gideem   baʕd   ʕašir   sanaw-at
     FUT-SG.M-COP   DEF-laptop   old.SG.M   after   ten.M   years-PL.F
     'The laptop will be outdated in ten years'

(39)   Ø   alħiin   ana   jayeʕ
     Ø   now   I   hungry
     'I am hungry now'/'I am getting hungry'

Note that the dropped copula in sentence in example (39) could be *kaan* or *saar*. Thus, the sentence, as shown in the translation above, is ambiguous. Also note that *kaan* and *sˤaar* also inflect for tense, number, gender, and person. Some conjugations for these two copulas are: *kin-t* 'I was' and *kana-u* 'they were' and for *sˤaar* e.g. *tisˤiir* 'it becomes' and *sˤar-an* 'they-F became'.

The discussion of GA morpho-syntax above revealed that inflection is heavily used in GA morpho-syntax. In fact, three out of the five morpho-syntactic features investigated here are consistently expressed as bound morphemes (the definiteness marker *al*, agreement markers, and object and possessive pronouns). Inflection also exists in the remaining two features, coordination and copular verbs.

The paradigmatic complexity in the GA morpho-syntax has been demonstrated in the above discussion of the subject-verb agreement rules and the object and possessive pronoun forms. As discussed in section 1.3 in Chapter 1, morphological simplicity is the norm in pidgin and Creole varieties. Hence, we would expect less inflection and more economy of morpho-syntactic rules in GPA. Indeed, the descriptive section of GPA system reveals that it contains fewer inflections compared to GA by employing a reduced agreement system (as demonstrated in 2.1.2.1). Moreover, as shown in 2.1.2.3, GPA is more analytic: Whereas GA object and possessive pronouns are bound morphemes, they are free morphemes in GPA.

In the following section I describe GPA with respect to the morpho-syntactic features investigated above, illustrating linguistic simplification in the GPA morpho-syntax.

2.1.2 *Gulf Pidgin Arabic*

In this section, I provide a descriptive account of GPA. I will mainly focus on the five features listed in 2.1 above. All the examples below are from my fieldwork data. The code of the informant is placed immediately after each example.

2.1.2.1 Agreement

In this section I provide a description of subject-verb agreement in addition to agreement in the NP and in the ADJP in GPA. In section 1.3.2 in Chapter 1, it has been suggested that agreement markers are very rare in pidgin languages. As will become evident in the discussion below, GPA uses a reduced agreement system.

(A) Subject-verb Agreement

There is less subject-verb agreement in GPA than in GA. In fact, my fieldwork data reveals that verbal forms are expressed variably via a number of strategies: (i) Informants either use the third person singular masculine present or (ii) past form of the GA verb, regardless of the gender, number, or person of the subject. In a considerable number of cases, my informants also (iii) use the imperative form of the GA verb in place of the GA present or past verb, (iv) replace the verb with a noun, (v) or even drop the verb when the meaning can be inferred from the context. The following examples illustrate the five strategies used by my speakers.

I- Generalization of the third person singular masculine present form of the GA verb over other subjects

(40)   ana   yi-dris   kulliah (M2)

        I   PRS.3SG.M-study   college

        'I studied in college.'

In GA, the 1SG form of the verb is used with the 1SG pronoun *ana*. Thus the GA form of sentence (40) would be:

        ana   daras-t   fi   el-kulliah

        I   study-1SG.PST   in   DEF-college

II- Generalization of the GA third person singular masculine past form over other subjects

(41)   Jawal   alatuul   kallam (P1)

        Mobile   always   speak.3SG.PST

        'I always call my family in Pakistan using my mobile phone.'

        In GA, the verb in example (41) would be the 1SG present form of the verb.

Thus, the GA version of (41) would be:

        ʕalatuul   a-kallim   b-il-jawwal

        Always   1SG.PRS-speak   with-DEF-mobile

III- Use of GA singular masculine imperative form of the verb instead of the inflected verb

(42)   Madrasah   ruuh   Quran   maalom (B1):

        School   go.SG.M.IMP   Quran   known

        'I went to school and studied Quran.'

        In example (42) above, B1 reports that he went to school and studied Quran when he was young. In GA, the form of the verb used in this case would be the 1SGM past form, as demonstrated below:

        riħ-t   l-il-midrisah   w=darast   el-quran

        Go-1SG.M.PST to-DEF-school   and=study.1SGM.PST   DEF-Quran

IV- Replacement of the verb with a noun

(43) inta   kalam   haða   kiða   haða   kiða (B4)

You.SGM   speech   this   such   this   such

'You tell me the Arabic words for this and that'.

In (43), the informant has replaced the GA inflected verb *t-gu:l* 'PRS2SGM-say' with the noun *kalam* 'speech'. Hence, when translated to GA, (43) would be:

t-gu:l   l-i   maʕna   haða   w=haða

PRS.2SG.M-say   to-me meaning   this   and=this

V- Verb Deletion

As discussed in section 1.3 above, contextualization is a typical trait in pidgin and creole languages. GPA seems to be in line with other pidgin languages in this phenomenon as the verb is deleted in some situations where the meaning is clear without the use of a verb. Given the absence of any verbal form, no account for subject-verb agreement can be provided in such cases, see the example below:

(44) Radio   Ø   mumkin   sayarah   bass (P4)

Radio   Ø   maybe   car   only

'I may only listen to the radio in the car'

In GA, the verb is *ʔasma* 'hear-1SG PRS'. Thus, (44) in GA would be:

yimkin   a-smaʕ   ir-radio   bass   fi-s-siyarah

maybe   1SG.PRS-listenDEF-radio   only   in-DEF-car

The forms of GPA verbs illustrated in (I) and (II) show that the verb in GPA only agrees with the subject if it is a third person singular. Indeed, it is this third person singular masculine form of the GA verb that is used with all other subjects regardless of their gender, number, or person. Alternatively, as shown in (III), (IV), and (V) respectively, the verb in GPA can be the GA singular masculine imperative form of the verb instead of the inflected verb, replaced with a noun, or deleted. In the next section I discuss tense marking in GPA in more detail.

VI- Tense marking (past, present, and future) in GPA

Although my data reveal that there are past and present tense forms of the GA verb which are used in GPA (as shown above in examples 40 and 41), it should be stressed here that in most instances the time reference of the sentence can only be understood from the context. In fact, my data contains many cases where the third person singular masculine present form of the GA verb is used in place of the past form of the

86

verb and vice versa (see 41). Examples (45-46) further illustrate non-congruous tense choices in GPA. In (45), the informant uses the 3SGM past form of the GA verb instead of the 2SGM simple present tense, *tħassil* 'you find'. In example (46), M1 uses the 3SGM present form of the verb instead of the 3PLM PST past form of the verb, *dfaʕaw* 'they paid'.

(45) Ay   makan moyah   mawjood   hassal   samak (B1)

Any   place water   available   find.3SG.M.PST   fish

'In any place in Bangladesh where there is water, you find fish.'

(46) Fi   θneen nafar awal kullu ma   yi-dfaa   fuluus (M1)

There two   people first all   not   PRS.3SG.M-pay   money

'There were two people who used to live with me. None of them has paid money for the landlord.'

As discussed in section 1.3.2, it is a typical feature of creole languages to use adverbials in order to specify temporal reference. GPA seems to follow this general pattern of pidgin and creole languages. Indeed, when the time reference of the verb is not clear from the context or when GPA speakers want to stress the time of occurrence of the action, they tend to use lexemes such as *awal* 'first/past' to indicate the past tense or *baaden* 'after' to indicate the order of two actions or to indicate the future tense. The use of time adverbials to indicate the order of actions is exemplified in (47) and (48) respectively. In (47), two adverbials are used to deliver the meaning: 'there was a benefit in the past, but now there is no benefit'.

(47) Alhen ma   fi   faydah...   Awal   fi (B2A)

Now   not   there benefit…   past   there

'I used to [learn Arabic quickly], but now there is no progress' [in learning Arabic].

(48) Inta   ruuh   Pakistan inta   baaden   kalam Urdu   maalom (P3)

You   go.SGM.IMP   Pakistan you   then   speech Urdu   known

'If you go to Pakistan you will learn Urdu'

Overall, this section has discussed tense marking in GPA verbs. It has demonstrated that GPA has typical pidgin structures, namely the lack of subject-verb agreement and increased dependence on the context. This effectively means that the GPA morpho-syntactic system relies less on inflection and more on temporal adverbials and contextualisation for denoting the time of events. Overall, there is less redundancy of

information in the GPA system, another typical pidgin feature as discussed in section 1.3 in Chapter 1 above. The next subsection discusses agreement in the GPA noun phrase and adjective phrase.

(B) Agreement in the Noun Phrase and in the adjective phrase

As discussed in 2.1.1.1 above, the GA adjective agrees with the noun in definiteness, gender, and number. Also, the GA demonstrative, which inflects for proximity, agrees with the noun in number and gender. However, as I have illustrated in 2.1.1.1 above, the system is not completely regular since the noun in GA is not consistent in its agreement with the numeral both in number, gender, and definiteness. This section illustrates agreement in the NP and in the ADJP in GPA. The three subsections below illustrate to what extent the complex system of noun-adjective agreement, numeral and quantified agreement, and noun/adjective agreement of GA is carried over into GPA.

I- Noun-adjective agreement

In GPA, the richness of the nominal and adjectival paradigm of the lexifier language is greatly reduced: the noun or adjective following a number comes in the non-definite, singular, masculine form, regardless of the number or the gender of the noun in the NP (Almoaily 2008). Thus, the adjective in GPA is invariant and does not agree with the noun in definiteness, gender, or number (Smart 1990). Note also that the adjective-noun word order is free in GPA. Below is an example from my fieldwork data:

(49)   Ana   mama   mawjood (P1)

         I   mother   available.SG.M

         'My mother is still alive'.

In this excerpt, P1 used the masculine form of the adjective with the feminine noun *mother*. The superstrate language form of the adjective in (49) would be *mawjood-ah* 'available-SG.F'.

II- Agreement between the noun or adjective and the numeral

As shown in 2.1.1.1, in GA the noun or adjective following the numeral can be singular, dual, or plural, feminine or masculine, and definite or indefinite. In contrast, in GPA the noun or adjective following a number is invariably in the indefinite, singular, masculine form, regardless of the number.

Whereas in GA the number can be definite or indefinite, feminine or masculine[11], the number in GPA is always in the indefinite form. Note, however, that, as exemplified in (50), the number in GPA inflects for gender. In the following excerpt, the singular form of the noun *lughah* 'language' is used with the feminine form of the number *arbaah* 'four':

(50)   Ana   mawjood   arba-ah   lughah (M5)

        I   exist   four-F   language.SG.F

        'I speak four languages.'

The sentence in (50) could be translated to GA as follows:

        Ana   atkallam   arbaʕ   luġ-at

        I   PRS.speak.1SG   four.M   language-PL.F

III- Demonstratives

According to Smart (1990), demonstratives in GPA, unlike in GA, do not agree with the noun or with the adjective in number and gender, neither they do inflect for proximity. Instead, the singular masculine proximate demonstrative *haða* 'this' is used with all nouns and all adjectives irrespective of their gender, number, or distance from the speaker. This finding is supported by my fieldwork data. My interviews and focus groups contained a direct elicitation section for demonstratives during which I asked seven of my informants to use the demonstrative with objects of different genders, numbers, and proximities. In fact, all the seven informants used the singular masculine proximal demonstrative *hatha* 'this' with all objects irrespectively of their gender, number, or proximity. Refer to Appendix B for a full account of this direct elicitation exercise.

The next section discusses the representation of definiteness and indefiniteness in GPA.

2.1.2.2 Definiteness and indefiniteness

The GA definiteness marker is normally dropped in GPA, with the result that in GPA there is no observable marker for definiteness (Smart 1990, Almoaily 2008). This could be taken to mean that the notion of definiteness and indefiniteness is fully contextualised in GPA. However, informants in the current study occasionally use the

---

[11] This applies to numbers from (1-10) only (refer to 2.1.1.1).

GA definiteness marker *al-* to express definiteness (refer to Chapter 5 for the number of instances for each informant). Examples (51) and (52) from my data exemplify dropping and retaining the GA definiteness marker *al-* in GPA, respectively.

(51) Ø   Mudiir   kalam   sakkar (B2)

      Ø   Boss   speech   close.3SG.M.PST

      'The boss asks me to close the shop (at night).'

In GA, the sentence in (51) would be:

      el-mudiir   yi-tˤlib   min-ni   a-sakkir   el-maħal

      DEF-boss   3SG.M.PRS-ask   from-me   1SG.PRS-close DEF-shop

(52) Fi   zawaj   al-hamdu   l-illah

      COP   marriage   DEF-thank   to-God

      'I am married, thanks God.'

In summary, the discussion in this section has revealed that GPA uses less inflection. This is due to the fact that GPA has less agreement both in the verb phrase and in the noun phrase. This feature is also observed in many pidgin languages as discussed in section 1.3.2. The next section examines personal pronouns in GPA.

2.1.2.3 Personal pronouns

Overall, the number of personal pronouns in GPA has reduced from eighteen GA pronouns, illustrated in tables 8-9 above, to only five. According to Smart (1990), the personal pronouns in GPA are *ana* 'I', *inta/anta* 'you', *hu(wa)* 'he/they', *hiy:a* 'she', and *nihn* 'we'. GPA pronouns seem to have been borrowed from Standard Arabic rather than from GA discussed in 2.1.1.3 above. The fact that GPA pronouns are pronounced more like the Standard Arabic pronouns supports this claim. For instance, *anta* could have been borrowed from the Standard Arabic pronoun *anta* 'you.SGM', *huwa* could have been borrowed from the homophonous Standard Arabic pronoun for 'he.SGM', *hiya* is also used in Standard Arabic as a third person single feminine pronoun, and *naħn* is used as a first person plural pronoun. Compare these pronouns with their corresponding GA counterparts, namely *int* 'you.SGM', *hu:* 'he', *hi:* 'she', and *ħinna* 'we'.

The reason for this reduction in GPA pronouns is that there is an overgeneralization of some Arabic pronouns in the GPA pronominal system. For example, in (53), there is no agreement in number; the third person singular masculine

subject pronoun *huwa* is used instead of the third person plural masculine subject pronoun.

(53) Kashmir   huwa   kalam   lahal šwayah (P1)
       Kashmir   he   speech   alone little
       'Kashmiris, their language is a bit different.'

In GA, the pronoun *hu* is only used as a third person singular masculine subject pronoun whereas in GPA, *huwa* is used both as a third person singular masculine pronoun and as a third person plural pronoun. In other words, the two GA pronouns *hum* 'they.M' and *hin* 'they.F' are replaced with *huwa* 'he'. Similarly, the second person singular pronoun in GA inflects for gender (i.e. *int* 'you.SGM' and *inti* 'you.SGF'). In GPA, however, *inta* can be used with male and female interlocutors. The following examples further demonstrate the levelling of the pronominal paradigm. Here, the same forms are used as subject, object, and as possessive pronouns (see also Smart 1990 and Almoaily 2008). In (54), the pronoun *ana* functions as a subject pronoun, in (55) it functions as a possessive pronoun, and finally in (56) it functions as an object pronoun. In GA, however, the pronoun *ana* in (55) and (56) would be the clitic *-i* in *ebo-i* 'father-my' and *mithl-i* 'same-me'. As such, in GPA the distinction between personal, object, and possessive pronouns is fully contextualised.

(54) Ana   ma   yi-rif (M1)
       I   NEG   3SG.M.PRS-know
       'I don't know'

(55) Baba   ana   moot (P1)
       Dad   I   death
       'My father died'

(56) Kalam   ma   yi-gdar   same same ana (P2)
       Speech   NEG   3SG.M.PRS-can   same same I
       'He (informant's brother) can't speak (GPA) like me'

The discussion in section 1.3.6 has suggested that pidgin languages tend to have an economic system, which uses as few pronouns as possible. Furthermore, as can be seen in examples (55) and (56) and discussed in section 1.3.6, GPA seems to correspond to the global structural feature of pidgin languages to use free pronouns rather than clitic ones. In fact, I found my informants occasionally drop the object and possessive pronoun when the referents can be derived from the context, a feature that is also attested in other pidgins/creoles. Below are some examples for null object and possessive pronoun:

(57) Walad   Ø   bint   Ø   madrasah (B2B)

      Son   Ø   daughter   Ø   school

      'My son and daughter go to school'

(58) Maa   fi   gariib   bass   Ø   maalom Ø (P2)

      Not   copula   close   but   Ø   know   Ø

      'We are not relatives but I know them'

In (57), informant B2B drops the 1SG possessive pronoun *ana* both after *walad* and *bint*. In (58) there are two dropped pronouns: the first is the 1SG subject pronoun *ana* and the second is the 3PL object pronoun *huwa*. The next section introduces another deletion in GPA, namely the dropping of coordinating conjunction markers.

2.1.2.4 Coordination

In GPA, as in other pidgin languages, asyndetic linkage is the norm (Smart 1990, also refer to section 1.3.7 in Chapter 1). However, my data shows that, while the general tendency is indeed for GPA to drop conjunction markers, some of the GA coordination conjunction markers are used by a number of GPA speakers. It appears that the GA markers *wa* 'and', *laakin* 'but', and *willa* 'or' are the most commonly used by GPA informants. The tendency of GPA to apply asyndetic linkage is exemplified in (59-60) below, while (61-62) demonstrate some conjunction markers which appeared in my fieldwork data:

(59) Fi   akhu   university   Ø   sister   university   (B4)

      COP   brother   university   Ø   sister   university

      'My brother and sister study at the university'

(60) Mumkin   θalaθah   Ø   arbaah nafar   mawjuud (M1)

      Maybe   three   Ø   four   person there

      'Maybe there are three or four persons there.'

(61) Wahid   w=nuss   fi   el-leel (M3)

      One   and=half   in   DEF-night

      'one and a half at night.'

(62) Ma   fi   maalom   laakin Sikh   fih (P3)

      NEG   COP   know   but   Sikh   there

      'I don't know but there are Sikhs…'

To sum up the previous sections, it is noticeable throughout the previous descriptive sections that GPA displays features typical for contact varieties in that it is much less highly inflected and that there is more economy of linguistic rules in the

morpho-syntax of GPA when compared to GA. The next subsection, on the other hand, introduces a rather uncommon feature of pidgin languages, namely the use of a copula.

2.1.2.5 Copula

As discussed in section 2.1.1.5 above, the copula is always missing in the GA present tense. It is used, however, in the past and future tenses. Although it is not common in pidgin languages to have copulas (see 1.3.3), a copula does exist in GPA, namely *fi*, which is also used in the present tense. Smart (1990) suggests that this GPA copula is derived from the GA existential particle *fih,* meaning 'there'. The use of the word *fih* in GA as an existential particle is exemplified in Sentence (63):

(63)   El-bēt   ma   fih   aħad
       DEF-house   NEG   there   anyone
       'There is no one in the house'

In GPA, the word *fi(h)* might be used as a copula, as in (65), or as an existential particle, as in (64). In this project, I am only investigating the use of *fi(h)* as a copula in GPA.

(64)   Ana   alhen   fi   sēkal (B2B)
       I   now   there   bike
       'I have a bike now'

(65)   Nafar   fi   kabiir yi-ji   šughul   kēf (P4)
       Person cop   old   3SG.M.PRS.come   work   how
       'If a person is old, how can he come and work here.'

My fieldwork data reveals that the use of the copula is optional in GPA. All informants who participated in this study have a tendency to drop the copula, both in the present and in the past tenses, except for the informants labelled (B3), who had a preponderance to produce the copula in the present tense rather than to drop it.[12]

Example (66) is an illustration of copula dropping in the past tense in GPA (more examples can be obtained from Naess 2008, Almoaily 2008).

(66)   ams   ana   Ø   mašgu:l (P3)
       yesterday   I   Ø   busy
       'I was busy yesterday'

---

[12] Refer to section 5.3 for more numerical and statistical details on the use of copula among the interviewed GPA speakers.

Note that in GA the copula *kan* – which inflects for person, number, and gender – is used in the past tense. Thus, (66) would be translated into GA as follows:

kint	mašġu:l	ams
was.1SG	busy	yesterday

Overall, this chapter has provided a morpho-syntactic description of GA, the superstrate language of GPA, and GPA, the pidgin under investigation. The next Chapter defines the substrate languages of GPA. I will introduce the methodology of determining the substrate languages in this project before moving on to describing the relevant morpho-syntactic features in these languages. The description in the next chapter will be limited to the same morpho-syntactic features discussed in this chapter.

Chapter 3: Substrate Languages, Definition and Description

This chapter aims at describing four of the main substrate languages of GPA: Bengali, Malayalam, Punjabi, and Urdu. The chapter starts with a discussion of the substrate languages and the methodology of determining the three languages, apart from GA, with the largest number of speakers in the Gulf region. I will then describe the morpho-syntactic features relevant to this project (i.e. agreement, pronouns, definiteness/ indefiniteness, coordination, and copula). Each descriptive section starts with a brief sociolinguistic, typological, and geographic account of the language under investigation. This chapter concludes with a comparison between GA, GPA, and the described languages in this chapter and a discussion of how similarities/differences between the substrate languages may possibly play a role in the variation encountered between the speakers of GPA.

3.1 Substrate Languages of GPA

GPA developed in a rather complex linguistic situation, and there are several substrate languages involved, among which are Urdu, Punjabi, Malayalam, Bengali, Tagalog, Pashtu, and Indonesian. Testing all the substrate languages of GPA as potential sources of linguistic variation in GPA would seem impossible. Therefore, I opted to investigate those languages which are spoken by the majority of workers as a mother tongue. Determining the most common languages, however, was not straightforward due to the lack of statistics about the ethnic backgrounds of expats in Saudi Arabia. The Saudi Central Department of Statistics-Demographic Survey Report (2004)[1] only gives statistics of the number of immigrants without any reference to the countries they come from or the languages they speak. In fact, the number of speakers for every substrate language in Saudi Arabia is constantly changing as workers tend to work under a two-year work extendable permit (refer to the introduction). Indeed, the available numbers of the population of potential GPA speakers differ widely. For instance, Ethnologue (2010)[2] estimates the population of Bengali speakers in Saudi Arabia as fifteen-thousand speakers, whereas according to the Bengali ambassador in Saudi Arabia[3] there are more than 1,300,000 Bengali workers in Saudi Arabia. Therefore, the approach I implemented

---

[1] Retrieved 31 January 2011 from http://www.cdsi.gov.sa/pdf/atlas2502.pdf
[2] Retrieved 10 December 2010 from http://www.ethnologue.com/
[3] In an interview with Alarabia.net in 31 Oct 2010.

in determining the three largest language groups was asking the expats themselves. The next section demonstrates the methodology which I followed in order to find out the three largest substrate language groups of GPA.

3.1.1 *Determining the three substrate languages with the largest number of speakers in Saudi Arabia*

I started my fieldwork in August-September 2009 by asking eighteen expats (shopkeepers, barbers, tailors, laundry workers, etc) from the countries where the majority of temporary immigrant workers come from (namely India, Pakistan, and Bangladesh), what they think is the largest language group from their country currently living in Saudi Arabia. The expats in this pilot study were from various linguistic backgrounds such as Bengali, Malayalam, Urdu, Punjabi, and Pashtu. I asked the following questions orally in short interviews with the informants:

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

This pilot study revealed that the three largest language groups are believed to be Bengali, Malayalam, and Punjabi.

All the Bengali informants in the pilot study claimed that there is only one language in Bangladesh, Bengali. Ethnologue (2010), however, suggests that there are 42 indigenous languages in Bangladesh. What is a language and what is a dialect is obviously a matter of definition, and often a highly controversial issue. There are indications, though, that Ethnologue favours classifying related varieties as languages rather than dialects of the same language.[4] This might be the reason why there is a difference between the number of languages Ethnologue has reported to exist in Bangladesh and the popular view, reflected in my pilot study. Even if we suppose that all the 42 languages are independent linguistic systems and not varieties of the same language, Bengali is by far the biggest single language spoken in Bangladesh, according to Ethnologue, 110,000,000 speakers out of 153,281,000, in the 2001 census. According to this statistic, 72% of the population in Bangladesh speaks Bengali as their first language. In addition to Bengali, only three of the 42 languages have over one million

---

[4] For instance, despite the fact that the varieties of Modern Arabic are treated by most linguists as Arabic dialects (refer to Ferguson 1989, Kenstowicz 1989, Holes 1990), Ethnologue referred to Gulf Arabic, Nadji Arabic, and Hijazi Arabic, three varieties of Arabic spoken on the Arabian peninsula, as languages rather than dialects of Modern Arabic.

speakers as reported by Ethnologue: Chittagonian (13m), Rangpuri (10m), and Sylheti (7m). Most of the rest of these languages have less than 10,000 speakers. Indeed, all the three languages with more than a million speakers are fairly closely related to Bengali since all four languages belong to the Bengali Assamese subgroup of Indo-Aryan.

Indicating the next two main language groups was more complicated since both India and Pakistan are highly multilingual countries. However, the seven Indian expats I asked agreed that the majority of expats from India speak Malayalam as their mother tongue. As regards the Pakistani community, five out of six of the Pakistani expats I queried claimed that the majority of Pakistanis who work in Saudi Arabia speak Punjabi as their mother tongue. Their claim is consistent with the fact that Punjabi is the mother tongue of 42 per cent of the population of Pakistan (Ethnologue 2010).

On the basis of these findings, it could be supposed that the largest three language groups which GPA speakers in Saudi Arabia speak as a first language are Bengali, Malayalam, and Punjabi. I will discuss these languages in turn regarding the morpho-syntactic features that are the focus of this study. In addition to these three languages, I will provide a description for Urdu for its potential effect as a second language and a lingua franca for the majority of Asian workers in Saudi Arabia (refer to 3.2.4).

3.2 Description of the Substrate Languages of GPA

As stated at the beginning of this chapter, each descriptive section starts with a brief introduction of the language under investigation.

3.2.1 *Bengali*

This Indo-Aryan language (belonging to the Bengali-Assamese group) is spoken in Bangladesh, India, Nepal and other countries (refer to Map 2 in Appendix D). Like most of other Indo-Aryan languages, including the two languages discussed later (i.e. Urdu and Punjabi), Bengali is an SOV language with a split-ergative case and agreement system (Bhatt 2007 and Holmberg, personal communication, 25 November 2010). According to Banglapedia (2006) and W. Wurff (personal communication, 16 March 2010), there are two forms of Bengali: Sadhu and Chalita. The latter is the form used both in daily communication and in formal settings, whereas Sadhu[5]– the old form – is only rarely used and only in formal settings such as newspaper titles and government

---

[5] Also known as Chaste.

announcements. This diglossic linguistic context is similar to the Gulf Region discussed in section 2.1.4 above.

For the purposes of the current study, we are thus interested in the Modern Standard form of Bengali, Chalita, because any possible effect on the Bengali informants' production of GPA will be likely to be due to Chalita rather than Sadhu. The following subsections highlight the morphological features of Bengali under investigation (see Ray, Abdul Hai, and Ray 1966 and Nasrin and Wurff 2009).

A) Agreement
1-Verbal agreement

Bengali has an affix-rich morphological system which comprises more than fifty different affixes to inflect for tense, mood, aspect, and person. Perhaps surprisingly, despite the large number of affixes in Bengali morphology, verbs in Bengali do not inflect for number and gender.

2-Nominal agreement

Nouns in Bengali, as shown in the definiteness subsection below, inflect for number, size, and proximity, but not for gender or animacy. Ray et al. (1966) suggest that adjectives in Bengali are distinguished from nouns only on the semantic level. Hence, the suffixes attached to nouns are also attached to adjectives. Nouns and adjectives of a Sanskrit origin also show agreement in gender. For example, they both take the suffixes for femininity *-i* as in: *sandar-i* 'good looking-F' *naar-i* 'woman'. Yet, Milne (1993:120) argues that 'it is only adjectives of pure Sanskrit origin which are inflected in the feminine and then only in the high literary style of writing'.

The tens of forms of demonstratives in Bengali 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.

B) Definiteness

According to Ray, et al. (1966), definiteness in Bengali is expressed indirectly using the suffixed demonstrative *-ta* and its various forms discussed above, which can be translated into English as 'that/those individual unit(s)'. It should be noted however that the suffix *-ta* and its various forms are mostly used with inanimate nouns (W. Wurff, personal communication). Ray et al. (1966) reported other demonstratives in Bengali

which can be used to express definiteness indirectly (e.g. *-a* and *-i* 'the unit fraction/ piece', and *gulo*/*guli* 'these/those'). Examples (1) and (2) below illustrate the Bengali demonstratives *–ta*, *-a -i*, and *gulo*, respectively. All examples are adopted from Ray et al. (1966).

(1) cabi-ta

    key-DEM

    'the particular key'

(2) gach-a   gach-i

    tall-DEM   tall-DEM

    'The tall piece'

(3) chele-gulo

    boys-DEM

    'those boys'

Indefiniteness in Bengali is expressed using the prefix demonstrative *ek-* and its various forms such as *ki* and *ko*, as in *ke-mon* and *ke-to*.[6] The prefix *ek-* means according to Ray et al. (ibid): 'an indefinite unspecified identity or measure'. Thus, *ek-* could be translated to English as: 'some'.

C) Pronouns

Subject Pronouns in Bengali are optional according to Ferguson (1991), i.e. Bengali is a subject pro-drop language. They inflect for person, politeness, and number, but not for gender. Pronouns in Bengali are free morphemes (e.g. *ami* 'first person', *tuy* 'a second person of a lower rank', *tumi* 'a second person of a similar rank', *apni* 'a second person of a higher rank', and *ini* 'third person'. Affixes are attached to the pronouns above to inflect for number and case (dative, genitive, and locative). Thus, the forms of the pronoun *ami* are: *ami* 'I', *amra* 'we', *amake* 'me', *amar* 'my', *amader* 'our, us', and *amate* 'between me and you'.

D) Coordinating conjunction

In Bengali, conjunction is normally expressed using asyndetic linkage as in (4). Another way, exemplified in (5), is to use conjunction markers such as *ba* 'or' and *ar* 'and'. Similar to English, the conjunction marker is put before the last coordinated element.

---

[6] The conditions under which these forms are used were not provided by Ray et. al. (1966).

(4) Ram   Shyam   Jadu

    'Ram, Shyam, and Jadu' or 'Ram Shyam, or Jadu'

(5) Ram   Shyam   ba   Jadu

    Ram   Shyam   or   Jadu

    'Ram, Shyam, or Jadu'.

E) Copula

According to (Ray, et al. 1966), Bengali does not have a copula in positive unemphatic sentences in the simple present tense. However, Finch (2001) claims that the copula can actually be overt in some cases. In fact, Finch suggests that the copula in Bengali can be either covert or overt, depending on the following element. Hence, Finch distinguishes between the true copula (covert copula), and the overt copular verb *achh*. The copula *achh* in Bengali is only used with stage level (i.e. temporary) predicates (e.g. 'X is sad'). Otherwise, the copula is covert.

3.2.2 *Punjabi*

Punjabi is an Indo-Aryan language spoken mainly in the Punjab province in Pakistan, refer to Map 3 in Appendix D. Punjabi is also spoken in Afghanistan and in India (the Indian variety is referred to in Ethonologue as Western Panjabi). The population of Punjabi speakers worldwide is estimated to be around 62 million speakers (Ethnologue 2010).[7]

Punjabi is believed to have been influenced by many languages, including English, Urdu, Persian, and Arabic (Shackle 1970). Indeed, Shackle points out that in Lahore, the capital of the province of Punjab, three languages are concurrently spoken, namely English, Punjabi, and Urdu. Punjabi has the lowest rank of the three. English is the language used in formal settings and amongst highly-educated people in the state of Lahore. Similarly, Urdu is widely used in formal settings such as public speeches and intellectual discussions, whereas Punjabi is typically the language of uneducated people. Generally, Punjabi is the colloquial language used by lower class and lower middle class family members. Note that the difference between this diglossic situation and its counterparts in Bangladesh and in the Arab countries is that this context involves the use of different languages whereas in the diglossic situations both in Bangladesh and in the Arab countries different varieties of the same language are used in parallel.

---

[7] Retrieved 10 December, 2010 from: http://www.ethnologue.com/

Below is a sketch of the morphological features in Punjabi (based on Bhatia 1993).

A) Agreement

1- Verbal agreement

There are three categories of verbs in Punjabi: Simple, conjunct, and compound. Verbal agreement occurs in simple verbs and in compound verbs only:

a) Simple (the verbal root):

The verb agrees with the subject in number and gender if the subject is in direct case (i.e. nominative), which is the case in the present/imperfective tenses; otherwise, the verb agrees with the object. The simple verb also inflects for voice, tense, mood and aspect. In the present indicative tense, for instance, the verb is formed with the present participial form of the verb. A suffix, which inflects for gender and number, is attached to the verb.

b) Compound verbs are sequences of verbs, classified and analysed in Bukhari (1999) as serial verbs. Agreement suffixes are attached to the final verb in the sequence.

2-Nominal agreement

Adjectives ending with -*aa* agree with the noun in number and gender. For instance, the adjective *cangaa* 'good' may take the following forms: *cangaa*-SGM, *cange*-PLM, *cangii*, SGF, and *cangiaa*-PLF. On the other hand, adjectives borrowed from other languages like *kaafii* 'enough' (borrowed from Arabic) and *xush* 'happy' (borrowed from Persian) do not agree with their head noun.

B) Definiteness

Punjabi does not use markers for definiteness/indefiniteness. Bhatia (1993: 99) states: 'the concept of definiteness and indefiniteness is expressed indirectly by means of pronouns and the numeral *ikk* 'one'. Similar to Bengali above, definiteness can also be expressed indirectly using demonstratives such as *é* for close objects (i.e. 'this/these') and *ó* for distant objects (i.e. 'that/those'), followed by the optional adjective *saaraa* 'all' as in (6) below:

(6) é   saar-iãã   káán-iãã   nüü   kaun náîî   jaan-daa?

    DEM   all-PL.F   story-PL.F   ACC/DAT who NEG   know-PRS.SG.M

    'Who does not know all these stories?'

In (7) below, the adjective is dropped:

(7) é   camra   vecko

     DEM   room   see.IMP

     'See this room.'

C) Pronouns

All Punjabi pronouns are free morphemes. In the following I will briefly sketch the most important contrasts that are marked in the pronominal system this language (see Bhatia 1993, Aglsoft 2010).

1. **Subject Pronouns** in Punjabi inflect for person and number. First person pronouns are: *maim* 'I' and *asim* 'we'. Second person pronouns are: *tum* 'you-SG', and *tusim* 'you-PL'. Demonstratives such as *ih* 'this' and *uh* 'that' are used as third person pronouns. Subject pronouns in Punjabi, according to Brown and Ogilvie 2009, are optional. Thus, Punjabi is a pro drop language.

2. **Object pronouns**: The first and second person object pronouns take the same form as the subject pronouns above. However, a different form is used for third person object pronouns: *us* for the singular third person and *is* for the plural.

3. **Possessive pronouns**: A different set of possessive pronouns is used to inflect for person, number, and gender. Consider the following examples of possessive pronouns in Punjabi: *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'.

D) Coordinating conjunction

Punjabi uses the following conjunction markers: *te*, 'and', *ki* and *jaa* 'or', *lekan, magar*, and *par* 'but' and *vii* 'also' to conjoin sentences, phrases, clauses and words. They are placed before the last coordinated element. Examples (8) and (9) illustrate the use of *te* 'and' to conjoin NPs and sentences:

(8) Mai   kamm kiit-aa   te   dost   ne   araam kiit-aa

     I   work   do-PST.SG.M   and   friend ERG   rest   do-PST.SG.M

     'I worked and my fiend rested'

(9) Munda   te   kurii   jaarae san

   Boy   and   girl   going were

   'A boy and a girl were going'

E) Copula

According to Bhatia (1993), a copula is used in positive sentences and deleted in negative ones. The unmarked word order in sentences with a copular verb is: Subject-complement-copula. Just like English, the copula agrees with the subject in number. As shown in the examples below, if the subject is singular, the copula *ai* is used. The copula *ne*, on the other hand, is used with plural subjects:

(10)   kurii   nek   ai

        girl   good   is

        'She is a good girl.'

(11)   Kuria   nek   ne

        Girls   good are

        'They are good girls'

3.2.3 *Malayalam*

Malayalam is the official language spoken in the South-Western Indian Region called Kerala, refer to Map 4 in Appendix D. It is classified as a Dravidian language. Malayalam is believed to be the mother tongue of more than thirty-five million people, as claimed by The Official Web Portal of Government of Kerala, 2010.[8] According to Subramoniam (1997) one unique phenomenon of Malayalam is the correlation between the religion of speaker groups and their dialects. The difference between the dialects of Malayalam is mainly on the lexical level in terms of the language from which words are borrowed, namely from Hebrew into Judeo-Malayalam, from Arabic into Malabar, and from Portuguese into Mappilah, respectively. This dialectal difference will be taken into consideration in the descriptive section below since the informants in the current study speak the Malabar dialect of Malayalam. Hence, if there are differences between the dialects of Malayalam on the morpho-syntactic level, only Malabar will be accounted for here. The subsections below describe the five morpho-syntactic features of Malayalam relevant to this project.

---

[8] http://www.kerala.gov.in, accessed 17 March 2010

A) Agreement

1.   Verbal agreement

Unlike other Dravidian languages, in Malayalam there is no agreement between the subject and the verb (Asher and Kumari 1997). Tense is expressed via suffixes attached to the final verb. According to Rajaraja and Roy (1999), the suffix –*unnu* marks present tense, the suffix -*i* marks past tense, while -*um* marks future tense. In compound verbs, the pre-final verb always remains invariant (Jayaseelan 2004).

2.   Nominal agreement

In Malayalam, the set of third person singular pronouns are used as demonstrative pronouns. Thus, demonstrative pronouns in Malayalam inflect for proximity, number, gender, and case. The prefix *i*- is attached to demonstrative pronouns to mark for closeness, whereas the prefix -*a* is attached to the pronoun to mark for remoteness (also refer to the subsection on pronouns below).

Note also that attributive adjectives (which are always pre-nominal) do not show agreement with the noun (Jayaseelan, personal communication, 2010). According to Asher and Kumari (1997), there is a small set of invariant adjectives which can occur in the pronominal position in Malayalam, among which are: *nalla* 'good', *ceriya* 'small', and *valiya* 'big'.[9] Predicative adjectives in Malayalam, on the other hand, agree with the noun they modify in gender, number, and person. Thus, when used predicatively, the adjective *nalla* 'good' may take the forms: *nalla-van* 'good-SG.M', *nalla-val* 'good-SG.F', *nalla-var* 'good-PL.HUM', *nalla-tə* 'good-SG.N', and *nalla-va* 'good-PL.N'.

B) Definiteness

There are no definiteness markers in Malayalam. However, definiteness can be expressed indirectly using the demonstratives *ii* 'this' and *aa* 'that'. On the other hand, there are markers for indefiniteness in Malayalam, which take the following form:

1. *oru* is only used with singular count nouns (e.g. *oru aal vanniṭṭunṭə* 'a man come')

2. Quantifiers like *cila* 'few', *pala* 'several', and *kuracce* 'little', which are used with plurals and mass nouns. An example for the use of quantifiers to express indefiniteness in Malayalam is *cila-rə niŋŋale kaaŋaan vanniṭṭuŋṭə* 'some.people you see come'.

---

[9] The list only includes nine adjectives, which Asher and Kumari (1997) also referred to as 'inherent adjectives'.

C) Pronouns

According to Jayaseelan (2000), Malayalam is a pro-drop language, with subject as well as object pro-drop. Pronouns in Malayalam, which are always free morphemes, inflect for person (e.g. *ɲaan* 'I' vs. *nii* 'you.SG' vs. *aval* 'she'), case (discussed below), number (e.g. *nii* 'you.SG' vs. *niŋŋal* 'you.PL'), gender (e.g. *avan* 'he' vs. *aval* 'she'), politeness, and proximity. They have the same case suffixes as those used with nouns. Thus, suffixes are attached to the subject pronoun to create direct object pronouns, indirect object pronouns, and possessive pronouns. The suffix *-ure* is used to create possessive pronouns. The suffix *-e* marks for accusative case (i.e. direct object of a transitive verb), while the suffixes *-kkə* and *-(n)ə* mark for dative case (Asher and Kumari 1997).

D) Coordination

Asher and Kumari (1997) state that coordination is expressed via the suffixes *-um* and *–oo*, in the following ways:

1. -um 'and' is attached to both coordinated elements. When coordinating sentences, the suffix is attached to verbs. Consider the example below:

(12)   Raaghavan=um   kumaar=um   vannu

        Raaghavan=and   Kumar=and   come-PST

        'Raghavan and Kumar came.'

2. -oo 'or' coordinates objects, as illustrated in (13)*:*

(13)   niŋŋalkkə   kitakkayil=oo   paayayil=oo   kitakkaam

        you   bed=or   mat=or   lie

        'you can lie on the bed or on the mat'

Jayaseelan (personal communication 2010) notes that the suffixes *–um* and *–oo* cannot be attached to a tensed clause. Instead, they are attached to the infinitive form of the verb and the tense is attached to the following auxiliary verb.

E) Copula

According to Asher and Kumari (1997) Malayalam uses the copulas *aakuka* and *aaŋtə*. The latter is more widely used, especially with nominal complements. These two copulas are used interchangeably. Although it is possible to drop the copula if the tense is clear from the context, it is still used almost all the time (Jayaseelan, personal communication, 2010). If it is impossible to express the tense without the copula then the copula is obligatory. Example (14) illustrates the use of the copula in Malayalam:

(14)   avan   oru   ṭiiccar   (aaŋtə)

        he   a   teacher   (be.PRS)

        'He is a teacher'

3.2.4 *Urdu*

Urdu is an Indo-Aryan language spoken in Pakistan, India, Bangladesh, Mauritius, and South Africa, and in many other countries worldwide (Ethnologue 2010). Urdu uses Arabic script and most of its formal vocabulary is borrowed from Arabic. According to Humayoun and Ranta (2007), Arabic influence on Urdu is not only on the orthographic and lexical levels. It is also found in the morphological system of the language. Urdu has become a lingua franca for immigrants from India, Pakistan, Bangladesh, Nepal, and Afghanistan living in the Middle East, Europe, the United States, and Canada (Schmidt 1999). Hence, authors like Bhatia and Koul (2000) suggest that the total number of Urdu speakers, both as a first or a second language, is around six-hundred million people. It is not surprising, then, that almost all the informants who participated in my study claimed to speak Urdu as a second language. For example, informant B2A reported that he learned Urdu only after he started working in Saudi Arabia, while M3 claimed that he became more competent in Urdu after working in Saudi Arabia.

Although none of the informants in this study speaks Urdu as a first language, it is included here due to its potential cross-linguistic influence on the linguistic production of informants participating in this study, given that Urdu is a second language to most of them and GPA is a third language. Tremblay (2006), reports that there is an increasing interest among researchers in language acquisition in the potential influence of the L2 on the acquisition of the third language (See also Leung 2005, Long and Doughty 2009, Bardel and Lindqvist 2010 for more discussions on third language acquisition).

Another reason for including Urdu here is the influence Arabic has had on Urdu in orthography, lexicon, as well as morphology, which in turn may have affected its role in the formation of GPA. The similarities between Urdu and Arabic could have led the first wave of immigrants to the Gulf, who speak Urdu either as a first or as a second language, to use Urdu patterns when trying to speak Arabic. If this assumption is true, Urdu could have helped in the evolution of the pidgin language spoken in the Gulf, GPA. One means to check whether this assumption is true or not would be to compare the linguistic data of informants who competently speak Urdu and informants who are either not competent in Urdu or do not speak it at all (see section 6.2.1).

In the next subsections, I provide a morpho-syntactic description of Urdu for the relevant features to this project.

A) Agreement

1. Verbal Agreement

Schmidt (1999) reports that finite verbs in Urdu are formed by adding suffixes to the verb root, which is inflected for tense, mood, aspect, gender, and number. Intransitive verbs in Urdu agree with the subject both in gender and number, but not in person. Transitive verbs, however, agree with the subject in gender, number, and person only in present/imperfective tenses, when the subject has nominative case. In past/perfective tenses the subject has ergative case, in which case the verb may agree with the object, if it has absolute case.

2. Agreement in the NP and in the ADJP

Adjectives in Urdu agree with the noun in number and gender (Schmidt 1999). Consider the following examples:

(15)   Bar-a   ghor-a

      Big-SG.M   horse-SG.M

      'A big horse'

(16)   Bar-i   ghor-i

      Big-SG.F   mare-SG.F

      'A big mare'

B) Definiteness

There appear to be no articles for definiteness or indefiniteness in Urdu, except for borrowed definite Arabic words, which retain their definiteness marker *al-*.

C) Pronouns

Urdu Pronouns inflect for person, number, politeness, and case but not for gender. Both object and possessive pronouns are free morphemes that are not identical to their subject pronoun counterparts, except for first and second person object pronouns (Schmidt 1990). Compare the pronouns below:

1) **Singular SBJ and OBJ pronouns**: *maim* 'I', *maim ne* 'me', *vo* 'he, she, it', *us ne* 'him, her, it'.

2) **Plural pronouns**: *ham* 'we', *ham ne* 'us', *vo* 'they', *inhom ne* 'them'. It should be noted here that the pronoun *vo* 'he, she it' is also used as a demonstrative pronoun in Urdu as in (17) below:

(17) vo   kya   hai

      this   what   COP

      'What is this'

D) Coordinating Conjunction:

The conjunction *aur* 'and' is used to link words and phrases. It can also be used as an adjective meaning 'more' or 'other'. The conjunction marker meaning 'or' in Urdu is *ya*. Both *aur* and *ya* are placed between the elements they coordinate as shown in the examples below. The sentence in (18) exemplifies using *aur* to coordinate phrases while (19) demonstrates using *aur* to coordinate words:

(18) bara   patila   aur   karahi

      big   frying pan   and   saucepan

      'A big saucepan and frying pan'

(19) Ge   ya   kafi

      'Tea   or   coffee'

E) Copula

According to Schmidt (1999) the copular verb in Urdu is sentence final. It inflects for person, number, and tense. The copula is dropped only in unemphatic negative sentences. Some forms of the Urdu copula are: *hai* 'is', *tehe*, 'was/were' and *haim* 'are'.

Now that I have provided a morpho-syntactic description for GPA and for the major languages in contact in the case of GPA, I try to provide a concise and an easily accessed cross linguistic comparison of Gulf Pidgin Arabic, Gulf Arabic, and the major substrate languages of GPA in table 1 below:

Language \ Feature	GPA	GA	Bengali	Punjabi	Malayalam	Urdu
Definiteness	ø Markers	al- (prefix)	Expressed indirectly (DEMs e.g. –ta)	ø Markers (expressed indirectly)	ø Markers (expressed indirectly)	ø Markers
Indefiniteness	ø Markers	ø Markers (Expressed indirectly)	prefix *ek-* 'some'	ø Markers (expressed indirectly)	*oru* (only with SG count Ns)	ø Markers
Coordinating Conjunction	ø Markers	Markers used	Optional CONJ markers	*ate* 'and' *jar* 'or'	Markers are used	*aur* 'and' *ya* 'or'
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)	*hōnā:* inflects for P,N, and G
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	G and N
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 and G
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	inflect for P, N, politeness, and case
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	inflect for P, N, politeness, and case

Table 1: A cross-linguistic comparison of the morpho-syntax of GPA, Gulf Arabic, and the three substrate languages.

As can be seen in table 1 above, the description of the substrate languages reveals some differences in the morpho-syntactic systems of these three substrate languages. For instance, Malayalam is characterised by an absence of subject-verb agreement, whereas in Punjabi the verb agrees with the subject in person, number, and gender. In Bengali, however, the verb agrees with the subject in person only. Another example of different morpho-syntactic structures in these substrate languages is in the existence of definiteness markers. Descriptive grammars reveal that Bengali is the only substrate

language among the three which uses a marker for definiteness, though expressed indirectly. Refer to section 4.2 for a discussion on expected language variation patterns in GPA based on the morpho-syntactic differences between the substrate languages.

In the next chapter, I detail the methodology and data used in my study.

Chapter 4: Data and Methodology

In this chapter, I provide a detailed description of the current study. The chapter is divided into two main parts. In the first part (sections 4.1 and 4.2) I describe the purpose of this study and illustrate the structure of its corpus. Then I discuss the process of creating the corpus in the second part (i.e. sampling, preparing for, conducting, and transcribing the interviews, as well as the procedures followed in counting and labelling the tokens) in the second part (sections 4.3 to 4.6). I conclude this chapter with a discussion on some potential limitations in the data used for this study.

4.1 Description of the Current Study

In section 1.6, I reviewed various studies describing, and hence proving the existence of, GPA in several countries of the Gulf such as the UAE (Smart 1990), Kuwait (Wiswall 2002), Saudi Arabia (Almoaily 2008, Alshammari 2010), Oman (Naess 2008), and Qatar (Bakir 2010). Since GPA is spoken over a wide geographical area in a multi-ethnic speech community (see the introduction), language variation in GPA seems inevitable. However, to date, there is no account of language variation in GPA caused by differences in the morpho-syntactic systems of the substrate languages of GPA despite the large number of substrate languages (see section 3.1.1). We also lack information about the effect of the length of stay in the Gulf, in spite of the fact that some of the foreign workers in have been in Saudi Arabia for more than twenty years (see table 1 below). The current thesis aims to provide an analysis of language variation in GPA based on different morpho-syntactic systems of the substrate languages of GPA and the duration of stay in the Gulf. In Chapter 3 above I detailed the morpho-syntactic features of the substrate languages. The hypotheses that follow from these differences are discussed in section 4.2. In 1.6 above, I reported the claim of Bakir (2010) that long-staying speakers of GPA tend to shift to GA. Bakir, however, did not use quantitative data to support his claim. Hence, in the current study I compare the data of newcomers to the Gulf area (i.e. GPA speakers who spent five years or less at the time I interviewed them) with that of the long-term residents in the Gulf (i.e. those who spent ten years or more at the time they were interviewed). This allows me to investigate the question whether GPA speakers actually shift towards GA after spending more than ten years in the Gulf.

Cross-linguistic interference, known as *language transfer*, is widely discussed in the literature of Second Language Acquisition (refer to Odlin 1989, Han 2004, and

Sabourin, Stowe, and Han 2006). Eckman's (1977: 321) Markedness Differential Hypothesis (MDH) proposes that 'the areas of difficulty that a language learner will have can be predicted on the basis of a systematic comparison of the grammars of the native language, the target language, and the markedness relations stated in Universal Grammar'. Investigations of the role of language transfer were also carried out on pidgin and creole languages. This suggests that language transfer is not restricted to full-fledged languages, but may also occur in contact languages. For example, Thomason and Kaufman (1991) discussed the role of language transfer in pidgin genesis. Siegel (1999, 2003) examined transfer constraints and substrate influence on Melanesian Pidgin.

As for the influence of the length of stay, this study aims to answer the question of whether GPA speakers shift to GA after spending some time in the Gulf or not. According to Versteegh (to appear), one of the main differences between the language acquisition of children and that of pidgin speakers is that child speech is characterised by a shift towards the target language while that of pidgin speakers tends to fossilise at a certain stage of language acquisition. In another view, however, Bakir (2010) argues that GPA speakers shift towards GA after spending some time in the Gulf. Bakir's (ibid) claim could be verified, or refuted, by comparing the data of newly settled GPA speakers with that of the speakers who have stayed longer in the Gulf. Hence, half of the informants polled in the current study have spent five or less years in the Gulf at the time I interviewed them and the other half have spent ten or more years (see table 1 below for details on the exact number of years spent in the Gulf for each informant). The data collected from the newcomers of each language group will be compared with that of long-term residents (e.g. newly-settled Punjabi speakers vs. Punjabi speakers who spent more than a decade in the Gulf). In other words, I will be investigating whether the long-term residents have actually shifted towards GA or not by comparing their proportional use of GA tokens with that produced by their newly settled counterparts.

It should be noted here that it is unlikely that the speech of long-term speakers has been influenced by other factors resulting from their length of stay, such as chronological changes in the quality and quantity of GA/GPA input. In other words, I assume that the GPA speakers who worked in the Gulf twenty years ago have gone through a similar experience of those currently working in the Gulf. Thus, what causes language variation in the data of the two groups (i.e. long-term and short-term informants) is possibly their length of stay in the Gulf, rather than differences in the input which speakers of the two groups have been exposed to. Indeed, this assumption is buttressed by the fact that GPA has been widely used in communications between locals

and immigrant workers in the Gulf over the past twenty-five years (also see earlier works on GPA, such as Smart 1990 and Hobrom 1996). Thus, the possibility that older speakers have been less exposed to GPA in their early years of residence in the Gulf, as compared to newly-settled speakers, is far-fetched (also refer to the discussion on the target of GPA speakers in section 6.1).

4.2 Hypotheses

The analysis is based on the informants' use of the variants of five selected morpho-syntactic phenomena, see section 2.1 above. I briefly recapitulate these features below.

(1) Free or bound object or possessive pronouns

In GA, object and possessive pronouns are suffixes attached to the noun. In GPA, however, the unmarked pronominal choice is the use of free pronouns to replace the GA bound object and possessive pronouns (see Smart 1990). For example *kitab ana* 'book I' instead of the GA *kitab-i* 'book-my'. GA bound possessive and object pronouns, however, may also occur in the speech of GPA. Hence, there are three variants for the GPA object and possessive pronouns (free pronouns, bound pronouns and dropping the pronoun).

(2) Absence or presence of the Arabic definiteness marker *al*[1]

The GA definiteness marker *al-* is normally dropped in GPA, but it may be infrequently used by some GPA speakers. For example, a speaker of GPA may say *ana hassal kitab* 'I found.3SGM book' or – less frequently – use the GA definiteness marker *al-*, as in: *ana hassal il-kitab* 'I found.3SGM the-book'.

(3) Presence or absence of Arabic conjunction markers

GA uses conjunction markers such as *willa* and *aw* 'or' and *wa* 'and'. These markers are normally dropped in GPA. Hence, the two variants in GPA are dropping the conjunction marker, as in: *walad ana Ø bint ana* 'son I daughter I' and – less often – using the conjunction markers, as in *walad ana wa bint ana* 'son I and daughter I', both of these sentences can be translated to English as 'my son and daughter'.

---

[1] I used the generic form *al-* to refer to the GA definite marker. Note, however, that the Arabic definite marker *al-* is pronounced in GA as /ɪl/. The phoneme /l/ assimilates with some consonants such as /n/, /r/, and /d/. Thus /ɪl/ is pronounced /ɪn:/ in the word /ɪn:u:r/ 'the light', see Ryding (2005).

(4) Presence or absence of the GPA copula *fi*

There is no copula in GA in the present tense, whereas in GPA there is an optional copula. For example, the English sentence 'I am a student' can either be translated to GPA as *ana fi taleb* or *ana Ø taleb*.

(5) Presence or absence of agreement in the VP and the ADJP

In GA, the verb agrees with the subject in gender, number, and person. In GPA, however, the third person singular masculine form of the verb is usually used with all subjects. Hence, the equivalent of the GA sentence *hi t-ruuħ l-il-midrisah* 'she 3SGF.PRES-go to-the-school' in GPA is *hiya y-ruuh madrasah* 'she 3SGM-go school'. One can rarely find tokens in GPA where the verb agrees with the noun in person, gender, and number (see section 5.1). Thus, the variants are: (i) use of the GA agreement markers with the agreeing noun, (ii) use of GA agreement markers, but the marker does not agree with the noun (iii) use of verb-less utterances.

As for the noun-adjective agreement, the GA adjective agrees with the noun in number and gender, as in *is-siyar-ah gidi:m-ah* 'the-car-SGF old-SGF', whereas in GPA the adjective is normally in the singular masculine form, regardless of its head noun (e.g. *say:ar-ah gadiim* 'car-SGF old.SGM'). Less frequently, the GPA adjective may agree with the noun in number and/or gender.

The hypotheses listed below have been formulated to test the effect of speakers' substrate languages and the length of stay effects on language variation in GPA. Differences in the substrate languages can be expected to have an effect on the choice among the available GPA variants. Hence, in the current project I investigate whether there is any indication in the data that the participating informants use morpho-syntactic features similar to the ones found in their L1s when they speak GPA (see the discussion on cross-linguistic interference in section 4.1). Prior to each set of hypotheses, I explain why they have been formulated as such. Note that a detailed description of each of the five morpho-syntactic features below can be found in Chapter 3. Please be reminded that the contrasts marked in the substrate languages are summarised in table 1 in Chapter 3 above.

4.2.1 *Variation in agreement*

A. Subject-verb agreement

The morpho-syntactic description in 3.2 above revealed that there is no subject-verb agreement Malayalam (Asher and Kumari 1997), while in Bengali, the verb agrees with the subject in person (Ray, Abdul Hai, and Ray 1966), and in Punjabi the verb agrees with the subject in number, gender, and person (Bhatia 1993). Therefore, we might expect Punjabi and Bengali speakers of GPA to produce more cases of subject-verb agreement than the Malayalam language group:

H_1 Malayalam speakers have less subject-verb agreement compared to Bengali and Punjabi speakers.

H_0 There is no difference in subject-verb agreement among the three language groups.

B. Agreement in the NP and ADJP

In Malayalam, the predicative adjective agrees with the noun in person, number, and gender (Asher and Kumari 1997) and in Punjabi adjectives agree with their head in number and gender, except for loan words (Bhatia 1993). Yet, in Bengali adjectives are distinguished from nouns only on the semantic level (Ray et al. 1966). Bengali adjectives of a Sanskrit origin inflect for gender only in literary writings. Apart from that, adjectives in Bengali do not inflect for number or gender; the singular masculine form is used with all nouns (Milne 1993). Thus, I formulated the set of hypotheses below:

H_1 Malayalam and Punjabi speakers have more noun-adjective agreement in the AP than Bengali speakers

H_0 There is no difference in noun-adjective agreement among the three language groups

4.2.2 *Variation in definiteness*

Bengali is the only substrate language of GPA which uses a marker for definiteness (see Asher and Kumari 1997, Ray et al. 1966, and Bhatia 1993). Therefore, Bengali informants might be expected to use the Arabic definiteness marker *al-* more frequently than their Punjabi and Malayalam counterparts when they speak GPA. Hence, the following set of hypotheses has been formulated:

H_1 Bengali speakers use the Arabic definiteness marker *al-* more frequently than Malayalam and Punjabi speakers.

H_0 There is no difference in the use of the Arabic definiteness marker *al-* among the three language groups.

4.2.3 *Variation in the use of object and possessive pronouns*

Pronouns are free in all the three substrate languages polled in the current study (see Chapter 3). Hence, the following hypothesis expects no difference in the use of object and possessive pronouns among the speakers of Bengali, Malayalam, and Punjabi:

H₁ There is no difference in the use of GPA possessive pronouns among the speakers Punjabi, Malayalam, and Bengali.

H₀ Speakers of one of the substrate languages use the GA bound pronouns more/less frequently than the speakers of the other two languages.

4.2.4 *Variation in coordination*

All of the GPA substrate languages under investigation use conjunction markers. However, Bengali is the only language where the use of conjunction markers is optional (see Ray et al. 1966). Thus, the hypothesis below expects this difference in the substrate languages to cause an effect on the informants' use of conjunctions in GPA:

H₁ Bengali speakers drop conjunction markers more frequently than Malayalam and Punjabi speakers.

H₀ There is no difference between speakers of the three substrate languages in using conjunction when speaking GPA.

4.2.5 *Variation in copular verbs*

In Malayalam, the copula is used without restrictions (Asher and Kumari 1997), whereas in Bengali it is used with short-term adjectives only (e.g. *happy*, *sad*, *here*), but not with nouns or long-term adjectives such as *tall*, *short*, and *bold* (Finch 2001). In Punjabi, the copula is used with positive sentences only (Bhatia 1993). Thus, Malayalam speakers are expected to produce more tokens of the copula than the speakers of the two other languages:

H₁ Malayalam speakers use the GPA copula *fi* more frequently than Bengali and Punjabi speakers.

H₀ There is no difference in the frequency of using the GPA copula among speakers of the three substrate languages.

4.2.6 *Length of stay*[2]

In addition to the hypotheses outlined above, we can also formulate the following hypothesis as regards variation between long-tern speakers of GPA and the newcomers.

H_1 Long-term residents shift to GA when speaking GPA.

H_0 There is no difference between newly-settled GPA speakers and those who spent 10 or more years in the Gulf.

4.3 The Corpus

The corpus consists of the speech of informants participating in the interviews which I conducted in Saudi Arabia in two field trips in August – September 2009 and in June – August 2010. The data-base consists of interviews with sixteen GPA speaking informants from three linguistic backgrounds, Malayalam, Bengali, and Punjabi. Interviews were conducted in two cities, Riyadh and Alkharj, which are located in the Central Province of Saudi Arabia. Half of the speakers in the sample have spent five or less years in the Gulf while the other half have spent ten or more years in the Gulf by the time they were interviewed.

Due to the absence of general principles for quantification of variability above the level of phonology (Macaulay 2002), researchers have come up with various methods for the quantification of tokens. Some authors quantified the tokens per number of words. For instance, Precht (2008) quantified gender similarities and differences per 1000 words in American English conversations. Similarly, Cheshire, Kerswill and Williams (2005) calculated variation in discourse per 1000 words. Other authors prefer to quantify the tokens per minutes or hours of speech in a sociolinguistic interview. For instance, Rickford and McNair-Knox (1994) examined the effect of the interviewer's race by calculating the tokens of African American syntactic features per hour of speech. Other researchers, on the other hand, favour quantifying the tokens of the target variable per line of transcript, see for example, Vincent and Sankoff (1992) who tabulated the used of punctors (i.e. function words assimilated with the previous phrase) per line of transcription.

Since the informants participating in the current study have been exposed to GPA over a period ranging from one and a half years to twenty-five years, one would expect huge variation in terms of number of words produced in an hour/minute of speech. New speakers are expected to have more pauses and speak slower than those who have spent

---

[2] This hypothesis applies to each of the five morphological features tested in this study, namely agreement, definiteness, pronouns, coordination, and copula.

more than ten years in the Gulf. Calculating per time or per line of transcription would, then, lead to inconsistencies. I, therefore, opted to calculate the tokens per number of words, regardless of the length of the turn or the number of words produced in a minute of speech.

Tabulating the number of tokens produced by every informant per 1000 words enables me to compare members of different language groups (e.g. the relative number of tokens produced by the Bengali sample vs. the relative number of instances produced by the Malayalam respondents) and to compare informants of different lengths of stay (e.g. within the newly settled Punjabi-speaking informants vs. old-staying Punjabi speakers). Tabulating the data also enables me to test the data of members of the same sub-group (e.g. newly-settled Punjabi-speaking informants). Section 4.6 provides a detailed discussion on the quantification of variants.

Overall, the data-base contains 12000 words: 4000 words per substrate language, 2000 of which were from recently settled informants and 2000 words from long-term residents. Distribution of the data is illustrated in figure 1 below. Note that the labels 'new' and 'old' in the figure below do not reflect the actual age of the informants, but their length of stay in the Gulf. Also note that the informants are all males, to increase homogeneity of the sample. Refer to table 1 below for more information on the informants participating in the current study.
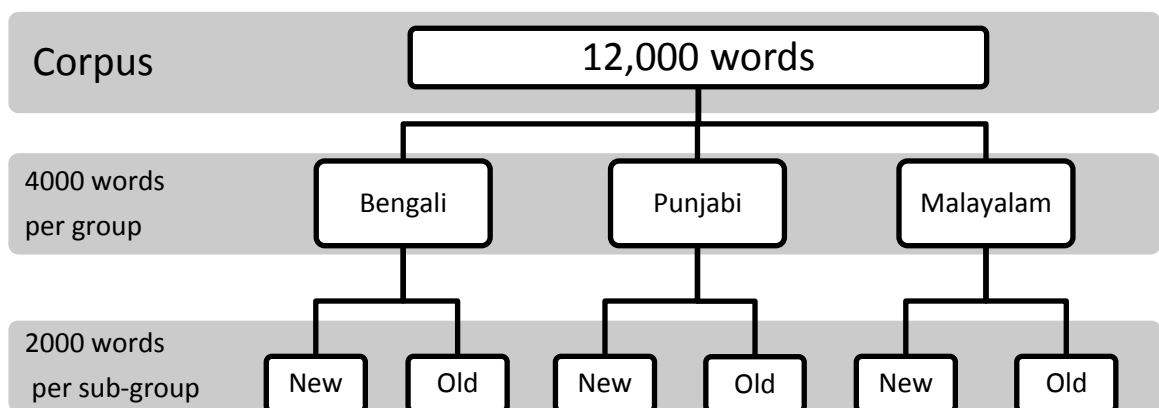

Figure 3: Distribution of the data

The speech of each language sub-group is represented by 2000 words, which are either produced by two speakers – producing 1000 words each – or by more than two speakers whose total number of words equals 2000 words.

4.4 Building the Corpus

As suggested by Schilling-Estes (2007: 165), 'conducting fieldwork to obtain data for sociolinguistic study is at the same time one of the most challenging and most rewarding aspects of sociolinguistic investigation'. Indeed, a researcher is expected to face difficulties when deciding on the speaker sample, collecting the data for his/her project, recording interviews of a reasonable quality, and transcribing the interviews. One common predicament that researchers face while conducting sociolinguistic interviews, for instance, is what Labov (1972) refers to as the *observer's paradox*. Bayley and Preston (1996: 2) describe this phenomenon as: 'the more aware respondents are that speech is being observed, the less natural their performance will be'. Modification of one's speech can be conscious or subconscious and can take place even without the presence of a linguist who monitors the subject's speech (see for example the concepts of divergence and convergence in Le Page and Tabouret-Keller 1985).

Graddol and Swann (1989) and Suleiman (1999) report that the social status of the interviewer can have an effect on the linguistic production of the interviewee. The interviewee may adapt his/ her speech to make it similar to that of the interviewee. Note that the obstacles towards creating accurate data are not only faced when conducting the interviews. Indeed, there can be other hindrances when transcribing the interviews. For example, choosing the appropriate transcription protocol can be problematic and controversial, especially since it greatly influences the questions that can be answered (see Todd 1990, Ammon, Dittmar, and Mattheier 2005, see also the discussion in 1.4 above). The strategies I employed to overcome, or lessen the impact of these problems are discussed in section 4.4.2 below.

In the next sub-section, I report on the first step I followed towards creating the corpus: choosing informants.

4.4.1 *Sampling*

The first element in the preparatory stage is deciding on the sample. Schilling-Estes (2007: 166-67), suggests that sampling depends heavily on the goals of the study. Therefore, she suggests that the following questions might be helpful in deciding on the sample. The first question is: 'what counts as the speech community?'. In other words, the researcher has to either linguistically or socially define the speech community in order to decide on the sample. In the case of GPA, the speech community consists of indigenous people of the Arabian Gulf and South Asian immigrants working in that region (See Smart 1990 and the introduction of this thesis). Since – in many cases – it is

impossible to examine the entire speech community, there are a number of sampling methods which are used in language variation studies such as random, stratified, ethnographic, and network sampling (see Shuy, Wolfram, and Riley 1968, Chambers, Drinkwater, and Boath 2003, Milroy and Gordon 2003, Buchstaller and Khattab, to appear). 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. De Vaus (2001) claims that the best way of avoiding biasedness is by polling from a random sample. Yet, trying to give each individual of the speech community an *equal chance of participation* can be difficult to achieve even by taking random numbers from a telephone directory 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 directory. Besides, it is difficult to obtain the consent of all the people chosen in such a random way of sampling. Ray (1985: 141), for instance, claims that '[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'. Random sampling would have been impossible to accomplish in this study for two reasons: (a) there is no list, or even statistics, of immigrant GPA speakers in Saudi Arabia which I could use to generate random informants (see section 3.1), and (b) there is a high rate of rejection to take part in interviews among immigrant GPA speakers, possibly because some of them are illegal immigrants and fear that the researcher is sent from the local authorities disguised as postgraduate researcher. Refusal to cooperate in sociolinguistic studies seems to be common in a fieldwork study of this kind (see also Schilling-Estes 2007, Patrick 1999). Some potential informants whom I approached did not agree to be interviewed because they did not want to be recorded and linguistically observed. Some others could not participate in the study because they were busy. In many cases, speakers of GPA simply rejected to be interviewed without mentioning reasons for their refusal.

Due to the impossibility to poll from a random sample of the GPA speech community, an alternative approach was employed i.e. snowball sampling.[3] Babbie (2010) describes snowball sampling as a sampling procedure which starts by collecting data from members of the population who are easily accessed, then asking those members to suggest other informants to participate in the study. I started my fieldwork in my hometown (Alkharj, Saudi Arabia) and in Riyadh by going to places where I was

---

[3] Snowball sampling is applied in various ways such as 'social network' and 'friend of a friend' techniques (see Buchstaller and Khattab, to appear).

likely to meet GPA speakers whose first languages are Bengali, Malayalam, or Punjabi.[4]
These places included foreign community centres[5], neighborhoods where GPA speakers
live, and work places of GPA speakers such as barbershops, bookstores, supermarkets,
etc. First I started by asking the GPA speaker about his first language and his length of
stay in the Gulf. If the person I approached met the requirements of this study (i.e. speaks
either Bengali, Malayalam, or Punjabi as his first language and has lived either five years
or less or ten years or more in the Gulf), I introduced myself to him, in GPA, as a
postgraduate student studying at Newcastle University conducting fieldwork on GPA and
asked for their consent to participate in the study.[6] Unfortunately, snowball sampling
worked in very few cases and I had to start the process until I collected a sufficient
amount of data (see figure 1). Another sampling procedure I used was asking the
presidents of foreign community centres in Riyadh to arrange with volunteers who were
willing to participate in the study. This technique is similar to the one above, but differs
from it in that the people I started with are locals who are not potential informants but are
in contact with tens – if not hundreds – of immigrant GPA speakers. I found this
procedure less time consuming and more effective. It seems that GPA speakers are more
confident to participate in the study when the call to participate in this study is made via
their community centre.

  In 4.4.1.1 below, I list the informants participating in this study

4.4.1.1 Informants

  In this section, I provide details of the exact length of each interview and the
social background of every informant polled in this study. In order to control the effect of
the sociolinguistic factors *gender*, *education*, and *social class*, I tried to make my sample
maximally homogenous. Hence, all the informants polled in this study are males who
work in low income jobs. Moreover, all the interviews were conducted in the Saudi
Central Province where Najdi Arabic – a sub-dialect of GA – is spoken. There were
slight inconsistencies, however, as regards the informants' level of education. For
instance, the Punjabi and Bengali speaking informants, with the exception of B3B, have
had not reached post-secondary education (i.e. University level). Malayali informants, on

---

[4] Refer to section 3.1 for a discussion on the pilot study, in which I determined the three largest substrate-language groups.
[5] Foreign community centres are Islamic preaching centres funded by local charity organisations, where members of foreign communities - Muslims and non-Muslims - mostly from the Philippines, Pakistan, India, and Bangladesh, gather and socialise.
[6] In some cases I provided examples of GPA and explained how they are different from Gulf Arabic and told the potential informants that I am interested in studying their Arabic variety.

the other hand, with the exception of M1 and M5, have completed their undergraduate studies (see table 1). It is a common practice in order to achieve anonymity of informants that pseudonyms are used instead of their real names (see Babbie 2010). Hence, the respondents in this study have been labelled with two or three digit and alphabetical codes. The first letter of each label stands for the first language of the informant: B stands for Bengali, M for Malayalam, and P for Punjabi. The number in the label distinguishes members of the same language group (e.g. M1 and M2 are two informants both speaking Malayalam). The third element of the code (only in group interviews) marks for informants participating in the same group interview. Thus, B2A and B2B are two Bengali speaking informants participating in the same group interview. Table 1 below lists the informants, their first languages, their age, years spent in Saudi Arabia, length of the interview, and the place of the interview. Note that I shaded the data of long-term residents.

Interviewee	L1	L2(s)[7]	Education	Age	Years in Saudi Arabia	Length of interview/ focus group	Place of interview
B1	Bengali	None	Primary	39	10	23:55	Riyadh
B2A	Bengali	Urdu	Primary	41	18	16:00	Riyadh
B2B	Bengali	Urdu	Primary	38	15		
B3A	Bengali	None	Secondary	26	3	25:41	Riyadh
B3B	Bengali	None	College	27	2.5		
B3C	Bengali	None	Secondary	23	3		
B4	Bengali	None	Primary	35	5	26:56	Riyadh
M1	Malayalam	Urdu Tamil	Intermediate	43	2.5	22:27	Riyadh
M2	Malayalam	Urdu	College	23	4	20:52	Riyadh
M3	Malayalam	Urdu Tamil English	College	41	18	23:02	Riyadh
M4	Malayalam	Urdu English	College	38	15	20:51	Riyadh
M5	Malayalam	Tamil Urdu	Secondary	24	1.5	18:58	Riyadh
P1	Punjabi	Urdu	None	47	5	22:42	Al-Kharj
P2	Punjabi	Urdu	Primary	30	6	22:59	Al-Kharj
P3	Punjabi	Urdu	Primary	50	20	24:25	Al-Kharj
P4	Punjabi	Urdu	Primary	55	25	22:46	Al-Kharj

Table 1: Informants (all are males)

---

[7] Other than GPA

In the few cases where I interviewed more than an informant at the same time (i.e. as a group interview) or when certain informants produced less than 1000 words in the interview, the two thousand words were culled from more than two persons who belong to the same group (i.e. speak the same language and share a similar number of years of residency in Saudi Arabia). Hence, the code B2 henceforth refers to two informants – B2A and B2B, who both produced a total of one thousand words and whom I interviewed in a group. Therefore, the total number of words produced by long-term resident Bengalis is the required 2000 words; 1000 words by B2A and B2B and 1000 words by B1. Similarly, the code B3 refers to three newly-settled Bengali informants: B3A, B3B, and B3C, who altogether produced a total of one thousand words. B4, who produced 1000 words, complements the data of new Bengalis. Furthermore, since the informant M2 produced less than one thousand words, I extracted 500 words from his speech and I interviewed another newly-settled Malayalam speaker, M5, to elicit the remaining 500 words from. Therefore, there are three Ms in the new Malayalam language group, who have together produced a total of two thousand words: 500 words by M2, 500 words by M5, and 1000 words by M1.

I followed four stages in building the corpus: preparation (deciding on sampling method and preparing an interview schedule and a consent form), conducting the interviews, transcribing them, and extracting the target amount of data. The following section discusses these steps in detail.

4.4.2 *Conducting the interviews*

There were two issues taken into consideration in the interviews. The first is research ethics and the second is the structure of the interviews (e.g. questions to be asked, duration, target data, etc.). These two issues are detailed below.

4.4.2.1 Ethics

As reported by Newman and Ratliff (2001) and Rice (2006) – see also the Ethical Guidelines for Good Research Practice[8] – there are ethics to be taken into consideration in fieldwork studies, which Rice believes to be even more important than the goal of gaining new knowledge via a fieldwork study. For example, a researcher doing fieldwork is obliged to ensure the safety, dignity, and privacy of the informant(s). Moreover, the researcher needs to ensure that informants have given consent to be recorded and

---

[8] Can be retrieved from: http://www.theasa.org/ethics/guidelines.shtml (accessed on 9 January 2012)

interviewed. Bearing these ethical considerations in mind, I adapted a consent form used in the NECTE project (Corrigan 2007). Once the GPA speakers gave oral consent to take part in the interview, I asked them to read the written consent form (see Appendix C), which is written in English. Since most of the informants participating in this study either do not speak English at all or have a little command of it[9], I provided them with an oral translation of the consent form in GPA. There was no point in having a translated version of the consent form into (Gulf) Arabic or GPA because the informants learnt GPA via verbal communication and most of them cannot read Arabic. If the subject agreed with the contents of the consent form, both the interviewer and interviewee signed and dated it.

The Ethical Guidelines for Good Research Practice, composed by the Association of Social Anthropologists of the UK and Commonwealth (1987), state that interviewers are obliged to pay back interviewees for their time and assistance: 'fair return should be made for their help and services'[10], also cf. Grant and Sugarman (2004). Whereas the issue of payment was not discussed before or during interviews in the current study, I decided to pay informants a little amount of money[11] in return for their time and participation once the interview finished.

There were certain techniques which I followed to reduce the impact of the observer's paradox. I am indebted to Huber (1999) and Labov (1972) who suggested most of these measures. For instance, one common way of avoiding the observer's paradox and stimulating the subjects to produce longer turns is by preparing questions that might cause the interviewee(s) – in one way or another – to forget that they are recorded and being linguistically observed. For example, one type of questions commonly used is the danger of death question and other story-telling questions included in Labov's (1972) interview schedule and adapted by many subsequent sociolinguists (such as Poplack 1989, Tagliamonte 2006, see also the network of modules in Milroy and Gordon 2003). It is perhaps needless to say that using GPA myself during the interviews could have contributed towards collecting more natural GPA data. Another technique of reducing the influence of the observer's paradox is to conduct dyadic interviews with two speakers of the contact variety. Huber (1999) reported this method as helpful for him as a foreigner observing Ghanaian Pidgin English. I also used this method in my data collection, but the difficulty to determine 'who said what' when transcribing the data –

---

[9] Only one subject (M3) claimed that he speaks English as a second language (see Table 1 below).
[10] Accessed 9 January 2012 from http://www.theasa.org/ethics/old_ethics.shtml
[11] 20 SAR (approximately equal to 4.5 GBP).

especially when the informants participating in one interview are not easily distinguished by their voice (e.g. similar pitch and speed) – made me decide not to use this method extensively. In fact, I have only used it twice. Huber (ibid) also reports that conducting the interview at the informants' workplace, or where they live, reduces the stress of interviewees, and hence provides the fieldworker with more accurate data. I found this method helpful as it seems to give informants more confidence, especially when their friends or co-workers could attend the interview. Thus, all of the interviews – with the exception of the one made with the informant P4 – were conducted at the informants' workplaces or at their community centre. This method was not free from problems, though. For instance, in my interview with P3 – who works as a tailor – the recording was not clear in some parts of the interview because of the noise of his co-worker's sewing machine. Luckily, P3's colleague did not use the machine for a long period. Otherwise, I would have had to re-interview the subject or search for a different informant. Moreover, in the interviews I conducted in the community centres, other GPA speakers attending the interview took turns in the interview. This had the positive side effect of making the interview more spontaneous. However, I had to make sure that the interviewee remains the main speaker in the interview, since other people taking part in the interview might not meet the selection criteria (the data in the interview are standardised, i.e. collected from people speaking either Bengali, Malayalam, or Punjabi as their first language and their number of years spent in the Gulf should be either five years or less or ten years of more).

Another method I used to reduce the impact of the observer's paradox is having a friendly open chat with the interviewee (e.g. asking them which city they come from and how big this city is compared to where they live in Saudi Arabia, etc.) before starting to record the interview. This method could have helped both the interviewer and the interviewee familiarise themselves with each other before the actual start of the interview. I have also found that asking informants questions about topics of their concern helpful in reducing the effect of the observer's paradox as well as in making them take longer turns. For example, my fieldwork trip was made during the spread of the swine flu pandemic throughout the world. At that time, the disease was a major concern, not just locally but also globally. Hence, I added the question 'are you worried about swine flu?' to the interview schedule. Other questions that were added for the same purpose are 'how was your experience of working abroad?' and 'do you feel homesick?' This brings us to the interview schedule, which I discuss with more detail in the next paragraph.

The interview schedule can be divided into two parts. Questions in the first part – which I will refer to as open-ended questions – were meant to stimulate the informants to answer them with long turns. Questions in the second part, on the other hand, were purposefully formulated to elicit variants of a certain variable.

The combination of these two types of questions (i.e. open-ended and questions eliciting certain linguistic features) is a widely used method in the sociolinguistic field of research; particularly in language variation studies (see Labov 1984, Sridhar 1991, and Llamas 1999) and it worked well for my purposes. Hence, before conducting the interviews I prepared a list of questions in which I adopted some questions from Tagliamonte (2006), including questions about demography, social practices, work life, school days, personal concerns, traditions, and language. Overall, the questions in this subdivision of the interview can be divided into three parts: (a) questions that determine the subjects' demographic backgrounds, (b) their linguistic backgrounds, and (c) questions that stimulate informants to produce long turns in the interviews. A list of the questions I used in in this part of the interview is in Appendix B. In addition to the open-ended questions, I prepared a PowerPoint presentation in which informants were asked to reflect on objects they see in the presentation. This task aimed at eliciting tokens of linguistic phenomena which I expected to be rare in the informant's answers to the open-ended questions such as prepositions and gender and number distinctions in demonstrative pronouns. In the first part of this presentation, subjects were asked to name objects of different quantities and genders located in various distances using a demonstrative pronoun. The purpose of this task was to check the use of GA demonstrative pronouns by GPA speakers. Note that demonstrative pronouns in GA inflect for gender, number, and proximity, while GPA demonstratives are invariant (the close singular masculine GA demonstrative *hatha* 'this.M' is used with all objects, see section 2.1.2 above). In the second task, informants were asked to mention the location of a ball which was positioned in various places in each slide. The purpose of this task is to investigate the use of prepositions by GPA speakers. Slides containing these two direct elicitation tasks are in Appendix B. I normally started the interviews with the general questions part, and then moved to the PowerPoint slides.

The recorded interviews range from 16 to 27 minutes (see table 1) and they were all recorded in MP3 audio format, using a high-quality digital recorder.[12]

---

[12] Make: Genx, Model: GDVR-901.

The next step in building the corpus is transcribing the interviews, which I discuss in detail in the following sub section.

4.4.3 *Transcribing the interviews*

Transcribing the interviews is a crucial means of enabling the researcher – and the readers – to access the information in his/her data easily. Hence, it is not surprising that Kvale, et al. (2008: 178) consider transcriptions to be 'the solid rock-bottom empirical data of an interview project'. It goes without saying that the transcription of audio texts is one of the most tedious tasks a researcher can face. Even a highly skilled typist might take up to five hours to transcribe one hour of speech (see Kvale, et al. ibid). In some cases, e.g. poor recording, large number of speakers in an interview, phonetic transcription rather than etymological transcription (see Powers 2005), it may take up to 24 hours to transcribe one hour. In my case, it took me nearly four hours to transcribe and revise only ten minutes of speech. Arabic transcription/dictation softwares, let alone transcription tools for non-standard Arabic varieties or Arabic-based contact languages, are inaccurate and thus were avoided in transcribing the data for the current project. Therefore, I transcribed the interviews myself, implementing careful procedures, which can be summarised as: (a) listening to the whole interview, (b) listening to the interview again, pausing the audio file every two to three seconds to transcribe that segment of the interview, (c) listening to the interview again and revising the transcribed text.

I concluded the discussion in section 1.4.2 with the suggestion that the best procedure for storing and retrieving the data of pidgin and creole languages might be using the standard spelling of the lexifier language and supplementing that with digital audio recordings. Accordingly, the transcription of the whole interviews and group sessions is in Standard Arabic script (see the excerpt in Appendix A).

Since the transcription of the interviews contains my own turns and other data irrelevant to the number of words used to build the corpus shown in figure 1 above, I also produced a shorter version of the transcribed interviews, which only contains the standardised number of words for the analysis. This process is illustrated in the next sub-section.

4.4.4 *Extracting the required amount of data*

The fourth step in creating the corpus was the extraction and tabulation of the tokens of morpho-syntctic phenomena investigated in this study. Hence, after transcribing all the interviews, I copied them into a different folder named 'interview

extracts' in which I deleted my turns, signs that do not refer to actual words such as the laughter sign @, my comments as a transcriber, and other turns made by people other than interviewee (e.g. the interviewees' co-workers). I only retained the last 1000 words produced by the respondent. The reason for keeping the last 1000 words is that these data are more likely to be naturalistic and free from the effect of the observer's paradox discussed earlier in this chapter, as the interviewee will have had the chance to settle into the situation. Since the interviews are not very long[13], other factors that could affect the accuracy of the collected data, such as the interviewer or interviewee tiredness, boredom, etc. are unlikely to have an effect on the data.

The next steps are labelling, glossing, and counting the tokens.

4.5 Glossing and Counting the Tokens

In order to make the tokens easier to retrieve from the transcribed interviews, I labelled each variant of a variable with a unique code (see the list below). The purpose of implementing this technique is that it allows quick access to the required token using the search facility found in standard word processors. Hence, counting the tokens is as simple as replacing the target code with the code itself, using Microsoft Word's command "replace all", Word will replace all the codes with the same code (e.g. it replaces COP+ with COP+) and will reveal how many replacements were made. The resulting number is the actual number of tokens of the target variant produced by the informant. The codes I used to refer to each variant are illustrated below. Note that these codes can be categorised into tokens for GA features in the GPA data, which I have identified with the sign * next to their meaning, and features typical to GPA only (the rest of codes/features). Examples of the linguistic features of GPA labelled by these codes can be found in section 2.1.2 above.

Code	Meaning
AFF PRO +	Object or possessive pronoun is used as a free morpheme
AFF PRO –	Object or possessive pronoun is used as a bound morpheme*
AFF PRO Ø	Possessive or object pronoun is dropped
DEF +	The definite article is present*
DEF –	The definite article is dropped

---

[13] The longest interview is about 28 minutes.

CONJ +	The conjunction marker is present*
CONJ –	The conjunction marker is dropped
COP +	The copula is used
COP –	The copula is dropped*
AGR+	Verbal agreement is present*
AGR –	The verbal agreement is missing
VØ	The verb is dropped
AGR NP +	Agreement in the noun phrase or in the adjective phrase is present*
AGR NP –	Agreement in the noun phrase or in the adjective phrase is missing

In summary, there are three versions of the transcribed interviews. The first is the transcription of the whole interview (see the example excerpt under [1] below from my interview with the informant M1). The second version is the interview extract which only contains words produced by the informants (see the example in [2]), and the third is the coded version (see [3]).

[1] An example of a transcribed interview:

Mohammad : تاميل هذا وين كلام تاميل

M1: ثاني كيرالا مدراس اردو فيه يمكن موجود اردو لكن بومبي برضة شغل شوية كذا. لكن في صحيح كلمتك العربي نفر الملباري. اول في (انديا). بعدين يجي من هنا كلم واحد اعشرين سنة يجي من جدة. بعدين بعدين هو رجع ممكن حق المدراسي في كيرالا هو في كبير هو في موت ممكن الحين حرمة هو فيه ثلاث حرمة هو في كيرالا حرمة ما يجي مشغول كذا نفر .

[2] An example of an interview extract:

ثاني كيرالا مدراس اردو فيه يمكن موجود اردو لكن بومبي برضة شغل شوية كذا لكن في صحيح كلمتك العربي نفر الملباري اول في (انديا) بعدين يجي من هنا كلم واحد اعشرين سنة يجي من جدة بعدين بعدين هو رجع ممكن حق المدراسي في كيرالا هو في كبير هو في موت ممكن الحين حرمة هو فيه ثلاث حرمة هو في كيرالا حرمة ما يجي مشغول كذا نفر.

[3] An example of a labelled interview extract:

ثاني كيرالا مدراس اردو فيه+ COP يمكن– COP موجود اردو لكن بومبي برضة شغل –AGR شوية كذا لكن في+ COP صحيح – COPكلمتك– AFF PRO العربي+ DEF نفر الملباري+ DEF .. اول في (انديا) بعدين – COP يجي من هنا كلم واحد– CONJ عشرين سنة– COP يجي من جدة بعدين بعدين هو– COP رجع ممكن حق المدراسي+ DEF في كيرالا هو في+ COP كبير هو في+ COP موت– CONJ ممكن الحين حرمة هو + AFF PRO فيه+ COP ثلاث حرمة هو + AFF PRO في كيرالا حرمة ما– COP يجي– AGR مشغول كذا نفر

In this excerpt, I asked (M1) the following question: "How did you learn to speak Tamil?", he replied:

"Schools are multilingual in Kerala, there is Urdu. Maybe Urdu is there. I have also worked in Bombay for some time. Oh yes. the Malabari guy I told you about earlier, who used to speak good Arabic. He came from India to Jeddah and spent twenty years. then returned to Kerala. He is old now. He might be dead. He used to have three wives and did not bring any of them. people (i.e. immigrant workers) are busy here".

Tabulation of the tokens is illustrated in the following section.

4.6 Quantification of tokens

For every linguistic feature chosen in the study (e.g. conjunction), I calculated the percentage of tokens produced of every variant. This was done by dividing the number of tokens of one particular variant (e.g. use of coordination markers) by the total number of tokens of all variants of the variable (e.g. total number of cases were the coordination markers are used plus the total number of instances where the informant dropped the conjunction markers) and multiplying the resulting number by 100. Thus, if – for example – the informant M3 used the copula *fi* 50 times and did not use it in 80 utterances, I divided 50 by 130 and multiplied the resulting by 100 (50 / 130 x 100 = 38.5). This means that M3 used the copula in 38.5% of the times where a copula could have been used. In order to calculate the average use of a variant by members of a sub-group, the resulting percentage (e.g. of dropping the copula by the informant labelled M3) is added to the percentage of the same variant by the other speaker(s) in the sub-group and divided by the number of speakers in the sub-group. Then I compared the average use of the given variant by members of a sub-group with that of other sub-groups (e.g. newly-settled Punjabi speakers vs. long-term Punjabi residents).

The asset of this quantitative method is quite evident: it gives an idea of the use of a linguistic variant as compared to the other linguistic variants produced by a sub-group of speakers in the sample. For example, the variable *definiteness* in GPA has two variants: the prefix *al-* and Ø (i.e. dropping this prefix). Hence, if new Bengalis dropped the definiteness marker *al-* in 90% out of the total number of tokens where they could use the definiteness marker while old Bengalis drop it only in 60% percent, this can be taken as an indication that Bengali speakers shift towards GA as they stay in the Gulf. In terms of possible substrate languages' effect on GPA morpho-syntax, if Punjabi speakers, who lack a copula in their L1 whereas Bengalis have one, are found to produce significantly

less tokens of the GPA copula *fi* than the Bengali speakers, this might be interpreted as a result of substrate influence.

In order to see whether the hypotheses formulated in section 4.2 above can be accepted or not, the data were plotted in tables which list the number of tokens produced by each informant/groups of informants. Chi-square tests were run to establish the significance of the effect of the informants' L1 and years of residency in the Gulf on variation in GPA. Note that this type of statistical tests determines whether a hypothesis can be accepted or rejected via investigating whether distributions of variants differ from each other (see Lilliefors 1967, Satorra and Bentler 2001, Corder and Foreman 2009). The null hypothesis is rejected if the p-value is less than 0.05. The results are displayed in Chapter 5 and discussed in detail in Chapter 6.

4.7 Gold cannot be Pure, and People cannot be Perfect[14]

Although I tried my best to make the data of this project as accurate as possible, there might be pitfalls which I find hard to overcome in the context of Gulf Pidgin Arabic. The first is eliminating the effect of L2(s), which might have an effect on speakers' use of GPA. Ideally, I would have selected informants who only speak their L1 and GPA. This procedure seems easy to do on paper. However, when it comes to real life, it turns out that eliciting data from informants who came to Saudi Arabia as monolinguals and then learned GPA is a difficult, if not impossible, task – for many reasons. The speakers of Punjabi and Malayalam come from highly multilingual areas. It is thus hard to find Punjabi speakers, for instance, who do not speak Urdu as a second language. The same can be said about Malayalam speakers, who mostly speak Urdu and Tamil as second languages. Since Urdu is a lingua franca in the Indian sub-continent (see 3.2.4), I have tried to account for its effect as an L2 in 6.3.3. Yet, Urdu is not the only L2 spoken by my informants. Table 1 shows that English and Tamil are also spoken as L2s by some Malayalam-speaking subjects. It could also be the case that some informants actually speak more languages than indicated in table 1 above, but they did not inform me of that. Hence, the pattern we find in the data could not just result from their L1 and their length of stay in Saudi Arabia – or other GA speaking countries – but could also be influenced by their L2s, L3s, etc. The second factor that I could not control for, but that potentially interferes with the results reported in Chapter 5, is the difference in the speakers' daily exposure to the superstrate language. Although I endeavoured to poll

---

[14] Chinese Proverb

from informants who are in direct contact with GA speakers (e.g. vendors, barbers, mechanics, etc.) and avoided conducting interviews with those who have less or no exposure to GA (such as workers in factories), there might be differences in the amount of exposure to GA among informants. Unfortunately, polling from informants who have had exactly the same amount of exposure to Gulf Arabic, or even Gulf Pidgin Arabic, during their stay in Saudi Arabia seems impossible, especially for long-term residents. Finally, there are a range of uncontrollable personal traits such as openness with strangers, willingness to learn GA, and different language learning abilities. I also need to mention the limited corpus size, which was severely restricted by the fact that I had to conduct the interviews, transcribe them, analyse the data, and present the project in a thesis format within a time frame of three to four years.

Hence, in line with the Chinese proverb I chose as the title of this section, researchers often find themselves in the position of having to make the most of obviously limited data.

In the next chapter, I list the results of this study.

Chapter 5: Results

In this chapter, I present the findings of my fieldwork. As detailed in section 4.2, every language group was split into two groups based on their length of stay in the Gulf (5 years or less or 10 years or more). Therefore, the label new/old in the tables below does not reflect the chronological age of the informants but their length of residency in Saudi Arabia, or any other GA speaking country. Refer to section 4.4.1.1 for the exact age of every informant as well as other demographic details.

As illustrated in section 4.6, the data will be represented as number of GA tokens as opposed to GPA tokens produced by each informant per one thousand words. Thus, in the tables below, the findings will be represented numerically as follows:

Interviewees' group (e.g. New Bengalis)	Variant 1 (GA feature)	Variant 2 (GPA feature)	Total
Informant X	Number of tokens (percentage[1])	Number of tokens (percentage)	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 1: Illustration of the results tables

Comparing the percentages of occurrence of each variable gives me the opportunity to contrast the proportionate use of GA variants as opposed to the proportionate use of the GPA variants by each informant as well as allowing me to compare members of the same group.[2] In addition, the average percentages for each variant (in the grey row) allow me to compare the data of the six sub-groups. Note that in some cases the numbers where very low and results need to be interpreted with caution.

Also be reminded that every sub-group produced a total of 2000 words. For the majority of groups, each informant produced a total of 1000 words (except for cases where informants produced only few words. A normalisation procedure was applied in order to overcome the problems associated with numerical imbalances, refer to section 4.6).

---

[1] The percentage in each cell represents the rate of occurrence of the token out of the total of tokens for each variant

[2] In each group, members speak the same L1 and have lived in Saudi Arabia for a relatively similar amount of time.

The sections below tabulate the observed linguistic variants for each linguistic feature under consideration. I will start by listing the data for each GPA variant in definiteness, then I will do the same for the GPA variants in the use of conjunction markers, the copula, object and possessive pronouns and agreement in the VP and 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 1 above. Each section starts with an explanation of the abbreviations used in the tables and an exemplification of the variants for the linguistic feature under investigation. While this section only tabulates the data and briefly discusses some of the general patterns, sections 6.1, 6.2, and 6.3 of the following chapter will discuss these results in more detail in the light of the hypotheses formulated in section 4.2.

5.1 Variation in Definiteness

GPA speakers variably produce the GA definiteness marker (i.e. the prefix *al*). Tables 2 to 7 tabulate the rates of occurrence of the GA definiteness marker (presence versus absence) by ethnicity and length of stay in Saudi Arabia.

5.1.1 *Bengali informants*

The numbers of tokens where the GA definiteness marker is present and dropped in the data of Bengali informants are presented in tables 2 and 3.

New Bengalis	Def. marker present (GA)	Def. marker missing (GPA)	Total
B3	9 (16.9%)	44 (83.1%)	53
B4	2 (5.4%)	35 (94.6%)	37
Average	5.5 (11.1%)	39.5 (88.9%)	

Table 2: Tokens of the definiteness marker *al-* by new Bengali informants

Old Bengalis	Def. marker present (GA)	Def. marker missing (GPA)	Total
B1	20 (28.6%)	50 (71.4%)	70
B2	13 (21.3%)	48 (78.7%)	61
Average	16.5 (24.9%)	49 (75.1%)	

Table 3: Tokens of the definiteness marker *al-* by old Bengali informants

Generally, Bengali speakers tend not to produce the definiteness marker *al-*. The highest rate of *al-* production is by B1 (29%). Note also that there is a slight difference

between the recently arrived and the long-term resident Bengalis in the sense that production of definiteness marker is higher amongst the two old Bengalis.

5.1.2 *Malayali informants*

The instances of used/dropped GA marker by the Malayali informants are displayed in tables 4 and 5.

New Malayalam	Def. marker present (GA)	Def. marker missing (GPA)	Total
M1	33 (44%)	42 (56%)	75
M2	14 (29.7%)	33 (70.3%)	47
M5	7 (18.4%)	31 (81.6%)	38
Average	18 (30.7%)	35.3 (69.3%)	

Table 4: Tokens of the definiteness marker *al-* by new Malayali informants

Old Malayalam	Def. marker present (GA)	Def. marker missing (GPA)	Total
M3	16 (32.6%)	33 (67.3%)	49
M4	65 (63.7%)	37 (36.3%)	102
Average	40.5 (48.2%)	35 (51.8%)	

Table 5: Tokens of the definiteness marker *al-* by old Malayali informants

The data reveals that there is more variability between speakers of a single group (ranging from 18% to 44% in the case of the recently arrived and between 33% and 64% amongst the longer-term residents) than between groups. However, note that the highest frequencies of the GA definiteness marker are produced by M4, a member of the old Malayali group.

5.1.3 *Punjabi informants*

Tables 6 and 7 depict the use/dropping of the GA definiteness marker by the Punjabi informants:

New Punjabis	Def. marker present (GA)	Def. marker missing (GPA)	Total
P1	6 (10.7%)	50 (89.3%)	56
P2	11 (20.4)	43 (79.6%)	54
Average	8.5 (15.5%)	46.5 (84.4%)	

Table 6: Tokens of the definiteness marker *al-* by new Punjabi informants

Old Punjabis	Def. marker present (GA)	Def. marker missing (GPA)	Total
P3	11 (26.8%)	31 (73.2%)	42
P4	2 (4.9%)	39 (95.1%)	41
Average	6.5 (15.8%)	35 (84.1%)	

Table 7: Tokens of the definiteness marker *al-* by old Punjabi informants

The data of the Punjabi informants also reveal noticeable variation within members of the same group. For example, P3 used the GA definiteness marker in 27% of all cases, while P4 only used it in 5% of the time.

Overall, the data shows that – as regards the definiteness marker – there is observable variation between some members who belong to a single group, both amongst the Malayalam and the Punjabi groups. A clear progression towards use of the definiteness marker is only observable amongst the Bengali sample. Note also that the Malayalam language group seems to use the definiteness marker slightly more than the other language groups. These observations are discussed in more detail in section 6.2.1.1.

5.2 Variation in the Use of Conjunction Markers

This section discusses the use of conjunction markers amongst the GPA speakers in my corpus. Tables 8 to 13 list the instances where informants used GA conjunction markers such as *aw* 'or' and *wa* 'and', compared to the number of cases where they produced asyndetic linkages.

5.2.1 *Bengali informants*

Tables 8 and 9 list the presence versus absence of GA conjunction markers in the two Bengali groups.

New Bengalis	Conj marker present (GA)	Conj marker missing (GPA)	Total
B3	1 (3.6%)	27 (96.4%)	28
B4	0 (0%)	34 (100%)	34
Average	0.5 (1.8%)	30.5 (98.2%)	

Table 8: New Bengalis' use of conjunction markers

Old Bengalis	Conj marker present (GA)	Conj marker missing (GPA)	Total
B1	10 (28.5%)	25 (71.5%)	35
B2	2 (5.7%)	33 (94.3%)	35
Average	6 (17.1%)	29 (82.9%)	

Table 9: Old Bengalis' use of conjunction markers

The data reveals an increase in the use of conjunction markers only by one member of the old speakers, B1. All other Bengali informants produced, if any, very low token numbers of conjunction markers.

5.2.2 *Malayali informants*

Tables 10 and 11 depict the cases where M informants drop/use the GA conjunction markers.

New Malayalam	Conj marker present (GA)	Conj marker missing (GPA)	Total
M1	5 (16.6%)	25 (83.4%)	30
M2	0 (0%)	18 (100%)	18
M5	0 (0%)	10 (100%)	10
Average	1.6 (5.6%)	14.3 (94.4%)	

Table 10: New Malayalam speakers' use of conjunction markers

Old Malayalam	Conj marker present (GA)	Conj marker missing (GPA)	Total
M3	6 (18.1%)	27 (81.9%)	33
M4	3 (17.6%)	14 (82.4%)	17
Average	4.5 (17.8)	20.5 (82.2%)	

Table 11: Old Malayalam speakers' use of conjunction markers

Both M2 and M5 did not use any of the GA conjunction markers, whereas the percentage of M1's GA conjunction marker use is similar to the percentages in the old group. This result is difficult to interpret in the light of my lengh-of-stay hypothesis.

5.2.3 *Punjabi informants*

Tables 12 and 13 tabulate the tokens of dropping/uttering the GA conjunction marker among the Punjabi language group:

New Punjabis	Conj marker present (GA)	Conj marker missing (GPA)	Total
P1	2 (7.4%)	25 (92.6%)	27
P2	11 (23.9%)	35 (76.1%)	46
Average	6.5 (15.6%)	30 (84.4%)	

Table 12: New Punjabis' use of conjunction markers

Old Punjabis	Conj marker present (GA)	Conj marker missing (GPA)	Total
P3	6 (30%)	14 (70%)	20
P4	4 (40%)	6 (60%)	10
Average	5 (35%)	10 (65%)	

Table 13: Old Punjabis' use of conjunction markers

The data in tables 12 and 13 reveal that there is an increase in the use of conjunction markers among the old Punjabi group compared to the newly arrived Punjabis.

In general, the data reveal a possible correlation between the length of stay and the use of conjunction markers in all the language groups. For example, the average use of conjunction markers by the new Bengali informants is 1.8%, while the percentage of old Bengali informants is much higher, at 17%. The same pattern is discernible for the other two groups. The data also show variability between speakers of the same group. This can be clearly seen in the data of the new Malayali group and the old Bengali group. I will discuss this finding in more detail in section 6.2.1.2.

5.3 Variation in the Use of the Copula

As discussed in Chapter 2, 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 14 to 19 list the number of instances where the informants used the copula and compare it to the number of instances where the informants could have used the copula but did not use it.

5.3.1 *Bengali informants*

Tables 14 and 15 show the numbers of the use of/dropping the GPA copula *fi* in the Bengali sample[3]

New Bengalis	Copula dropped (GA)		Copula used (GPA)		Total
	PRS	PST	PRS	PST	
B3	36 (41.3%)	10 (66.6%)	51 (58.6%)	5 (33.3%)	PRS 87 PST 15
B4	100 (82%)	10 (62.5%)	22 (18%)	6 (37.5%)	PRS 122 PST 16
Average	68 (61.6%)	10 (64.5%)	36.5 (38.3%)	5.5 (35.4%)	

Table 14: New Bengalis' use of the copula *fi*

---

[3] Note that I calculated the percentages of the use of the copula for the two tenses (i.e. present and past) separately. For example, the number of tokens of the copula produced by B1 in the present tense (97) is divided by the total number of copulas that could have been produced in the present tense (126) and multiplied by 100 (97 / 126 x 100 = 77).

Old Bengalis	Copula dropped (GA)		Copula used (GPA)		Total
	PRS	PST	PRS	PST	
B1	97 (77%)	20 (87%)	29 (23%)	3 (13%)	PRS 126 PST 23
B2	106 (79%)	16 (72%)	28 (20%)	6 (27.2%)	PRS 134 PST 22
Average	101.5 (78%)	18 (79.5%)	28.5 (21.5%)	4.5 (20.1%)	

Table 15: Old Bengalis' use of the copula *fi*

Both in the present and in the past tenses, the Bengali informants seem to drop the copula more frequently than using it, except for the three informants grouped together under the label B3, who produced more tokens of the GPA copula in the present. The high frequency of the occurrence of the GPA copula in the present tense is characteristic of all members labelled B3.

5.3.2 *Malayali informants*

The instances of dropping and retaining the GPA copula in the Malayalam language group is demonstrated in tables 16 and 17 below:

New Malayalam	Copula dropped (GA)		Copula used (GPA)		Total
	PRS	PST	PRS	PST	
M1	49 (55.6%)	21 (65.6%)	39 (44.3%)	11 (34.4%)	PRS 88 PST 32
M2	39 (75%)	4 (66.6%)	13 (25%)	2 (33.3%)	PRS 52 PST 6
M5	42 (87.5%)	10 (100%)	6 (12.5%)	0 (0%)	PRS 48 PST 10
Average	43.3 (72.7%)	11.6 (77.4%)	19.3 (27.2%)	4.3 (22.5%)	

Table 16: New Malayalam speakers' use of the copula *fi*

Old Malayalam	Copula dropped (GA)		Copula used (GPA)		Total
	PRS	PST	PRS	PST	
M3	70 (68.6%)	24 (64.8%)	32 (31.3%)	13 (35.1%)	PRS 102 PST 37
M4	84 (70.5%)	19 (82.6%)	35 (29.4%)	4 (17.3%)	PRS 119 PST 23
Average	77 (69.5%)	21.5 (73.3%)	33.5 (30.3%)	8.5 (26.2%)	

Table 17: Old Malayalam speakers' use of the copula *fi*

The data of the Malayalam language group show that the members of the old as well as the new group have a preponderance to drop the copula both in the present and in the past tenses.

5.3.3 *Punjabi informants*

Tables 18 and 19 display the tokens of using and dropping the GPA copula by the Punjabi informants.

New Punjabis	Copula dropped (GA)		Copula used (GPA)		Total
	PRS	PST	PRS	PST	
P1	90 (73.7%)	34 (89.5%)	32 (26.2%)	4 (10.5%)	PRS 122 PST 38
P2	102 (67.5%)	13 (62%)	49 (32.4%)	8 (38%)	PRS 151 PST 21
Average	96 (70.4%)	23.5 (75.7%)	40.5 (29.3%)	6 (24.2%)	

Table 18: New Punjabi speakers' use of the copula *fi*

Old Punjabis	Copula dropped (GA)		Copula used (GPA)		Total
	PRS	PST	PRS	PST	
P3	83 (57.2%)	24 (63.1%)	62 (42.7%)	14 (36.8%)	PRS 145 PST 38
P4	98 (71.5%)	24 (75%)	39 (28.4%)	8 (25%)	PRS 137 PST 32
Average	90.5 (64.3%)	24 (69%)	50.5 (35.5%)	11 (31%)	

Table 19: Old Punjabi speakers' use of the copula *fi*

Members of the Punjabi language group drop the copula in more cases than they use it, both in the present and in the past tenses. There seems to be no effect of the length of stay of the informant on the use/dropping of the GPA copula.

Generalising across these three data-sets, tables 14 to19 reveal that all informants drop the copula more often in the present tense, except B3, who uses the copula in 51 cases and drops it in 36 cases. In the past tense, all informants, including B3, 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 copula among GPA speakers. More discussion on these findings can be found in section 6.2.3.1.

5.4 Variation in the Use of the Object and Possessive Pronouns

In this section we look at variation in the use of object and possessive pronouns in GPA. Due to the wealth of agreement in the pronominal system, I have decided to narrow down the realm of possible structures into four patterns which are defined both by the presence or absence of the pronoun as well as by the type of morphology (i.e. bound versus free). The four possible variants are:

(A) **AGR+ Bound**: The agreeing object or possessive pronoun is attached to the verb, noun, or preposition as a suffix, as in GA (e.g. *qalam-i* 'pen-**my**').

(B) **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** asharah* 'you give-**you**' [instead of the GA suffixed pronoun **-ni**: 'me']).

(C) **Free morph**: The subject form of the object or possessive pronoun is used (subject forms are free morphemes), e.g. *Inta kalam **ana*** 'you speech I' (instead of *kallamt-ni* 'speak.PST-1SG.OJB PRO' in GA).

(D) **Dropped**: The object or possessive pronoun is dropped, e.g. *sadig-Ø yiji hina* 'friend-Ø come here'. [instead of *sidig-i yiji hina* 'my friend comes here']

The instances of (A-D) found in my data are tabulated in tables 20 to 25 below. Please be reminded that the percentages represent very low numbers.

5.4.1 *Bengali informants*

Tables 20 and 21 show the instances of the four variants for the GPA possessive and object pronoun in the Bengali language group[4]

New Bengalis	AGR+ Bound (GA)		AGR- Bound pro (GPA)		Free morph. (GPA)		Dropped (GPA)		Total
	POSS	OBJ	POSS	OBJ	POSS	OBJ	POSS	OBJ	
B3	1 (12.5%)	0 (0%)	0 (0%)	0 (0%)	6 (75%)	1 (12.5%)	1(12.5%)	7 (87.5%)	OBJ 8 POSS 8
B4	5 (18.5%)	0 (0%)	0 (0%)	0 (0%)	16 (59.2%)	0 (0%)	6 (22.2%)	3 (100%)	OBJ 3 POSS 27
Average	3 (15.5%)	0 (0%)	0 (0%)	0 (0%)	11 (67.1%)	.5 (6.3%)	3.5 (17.3%)	5 (93.6%)	

Table 20: New Bengalis' 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	
B3	2 (20%)	1 (8.3%)	0 (0%)	1 (8.3%)	8 (80%)	7 (58.3%)	0 (0%)	3 (25%)	OBJ 12 POSS 10
B4	2 (22.2%)	0 (0%)	0 (0%)	1 (33.3%)	2 (22.2%)	1 (33.3%)	5 (55.5%)	1 (33.3%)	OBJ 3 POSS 9
Average	2 (21.1%)	0.5 (4.2%)	0 (0%)	1 (15.2%)	5 (51.1%)	4 (45.8%)	2.5 (2.5%)	2 (29.1%)	

Table 21: Old Bengalis' use of object and possessive pronouns

The general tendency for new Bengalis seems to be pro drop of 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.

---

[4] Note that the percentages of object and possessive pronouns are calculated separately.

5.4.2 *Malayali informants*

Tables 22 and 23 present the results for the Malayalam language group.

New Malayalam	AGR+ Bound (GA)		AGR- Bound pro (GPA)		Free morph. (GPA)		Dropped (GPA)		Total
	POSS	OBJ	POSS	OBJ	POSS	OBJ	POSS	OBJ	
M1	1 (9%)	5 (71%)	0 (0%)	0 (0%)	8 (72.7%)	1 (14.2%)	2(18.1%)	1(14.2%)	OBJ 7 POSS 11
M2	1 (25%)	0 (0%)	0 (0%)	0 (0%)	3 (75%)	0 (0%)	0 (0%)	3(100%)	OBJ 3 POSS 4
M5	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (33.3%)	0 (0%)	2 (66.6%)	2 (100%)	OBJ 2 POSS 3
Average	.6 (11.3%)	1.6 (23.6%)	0 (0%)	0 (0%)	4 (60.3%)	.3 (4.7%)	1.3 (28.2%)	2 (71.4%)	

Table 22: New Malayalam speakers' 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	
M3	4 (50%)	0 (0%)	0 (0%)	0 (0%)	4 (50%)	0 (0%)	0 (0%)	4 (100%)	OBJ 4 POSS 8
M4	3 (33.3%)	2 (50%)	0 (0%)	1 (25%)	6 (66.6%)	0 (0%)	0 (0%)	1 (25%)	OBJ 4 POSS 9
Average	3.5 (41.6%)	1 (25%)	0 (0%)	0.5 (12.5%)	5 (58%)	0 (0%)	0 (0%)	2.5 (62.5%)	

Table 23: Old Malayalam speakers' use of object and possessive pronouns

There is a great variation between members of the same group. For instance, dropped object pronouns in the new group range between 100% and 14%.

5.4.3 *Punjabi informants*

The occurrences of the four variants of the possessive and object pronouns by the Punjabi informants are displayed in tables 24 and 25:

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%)	8 (66.6%)	0 (0%)	4 (33.3%)	6 (100%)	OBJ 6 POSS 12
P2	2(18%)	1(10%)	0 (0%)	0 (0%)	4(36.6%)	3(30%)	5(45.4%)	6(60%)	OBJ 10 POSS 11
Average	1 (9%)	0.5 (5%)	0 (0%)	0 (0%)	6 (51.6%)	1.5 (15%)	4.5 (39.3%)	6 (80%)	

Table 24: New Punjabi speakers' 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	
P3	0 (0%)	0 (0%)	0 (0%)	1 (33.3%)	10(45.4%)	2(66.6%)	12(54%)	0 (0%)	OBJ 3 POSS 22
P4	2 (33.3%)	0 (0%)	0 (0%)	0 (0%)	2 (33.3%)	2 (40%)	2(33.3%)	3(60%)	OBJ 5 POSS 6
Average	1 (16.6%)	0 (0%)	0 (0%)	0.5 (16.6%)	6 (39.3%)	2 (53.3%)	7 (43.6%)	1.5 (30%)	

Table 25: Old Punjabi speakers' use of object and possessive pronouns

Generally, the newly arrived Punjabi informants tend to either use possessive pronouns as free morphemes or drop them and to drop object pronouns, while old Punjabis drop possessive pronouns more than using them either free or bound morphemes. No general pattern can be seen for object pronouns in the old Punjabi informants' data.

The data in tables 20 to 25 reveal that all the informants rarely use pronouns as bound morphemes. Pronouns in GPA are typically used either as free morphemes, or dropped. I will return to these findings in section 6.2.2.1, where I will discuss them in more detail.

The next section lists the tokens of agreement/lack of agreement in the VP and in the NP and ADJP.

5.5 Variation in Agreement

5.5.1 *Verbal agreement*

As discussed in section 2.1.2.1, in GA the verb agrees with the noun in gender, number, and person. The verb also inflects for tense, mood, and voice. In GPA, however, the verb typically does not agree with the noun. Instead, the GA third person singular masculine form of the verb tends to be used with all subjects. Furthermore, tense in GPA is not marked by verbal inflection. Thus, speakers of GPA may use forms like *Ana maalom hatha* 'I know.PST this' (as opposed to the GA form: *Ana aʕarif hatha* 'I know.PRS 1SG this'), or *ana yiji Saudia gabl wahid sanah* 'I come Saudi Arabia last year' (as opposed to the GA form *jiit* 'came-1.SG.PST). GPA speakers might also drop the verb entirely when the information about the action/activity is retrievable from the context. Overall, the verbal agreement variants attested in my data are grouped in this investigation as follows:

(A) **AGR Present**: The inflected verb agrees with the subject in gender, number, and person. This means that the GPA speaker applies the GA TMA verbal inflection. Note that I have excluded from consideration all third person singular masculine tokens where the verb agrees with a third person singular subject. This is because this form constitutes the unmarked GPA verb form and thus does not reflect whether the informant actually applies GA verbal agreement or whether they only use the invariant form, which happens to be the agreeing form.

(B) **AGR Missing**: The verb is inflected but does not agree with the noun in person, number, or gender. Note that this term does not mean that agreement markers are

not used. In fact, this label covers the unmarked form of the GPA verb, where the third person singular masculine prefix *yi-* is used with all nouns.

(C) **Verb Dropped:** The verb is dropped; therefore, no account of agreement can be given

Tables 26 to 31 below list number the tokens for (A), (B), and (C) above:

5.5.1.1 Bengali Informants

Tables 26 and 27 show the number of tokens for verbal agreement, missing agreement, and for verb drop in the Bengali informants' data:

New Bengalis	AGR Present (GA)	AGR Missing (GPA)	Verb Dropped (GPA)	Total
B3	1 (3%)	18 (54.5%)	14 (42.4%)	33
B4	0 (0%)	9 (31%)	20 (69%)	29
Average	.5 (1.5%)	13.5 (42.6%)	17 (55.7%)	31

Table 26: Verbal agreement in the new Bengalis' data

Old Bengalis	AGR Present (GA)	AGR Missing (GPA)	Verb Dropped (GPA)	Total
B1	0 (0%)	27 (87%)	4 (13%)	31
B2	2 (9.6%)	14 (66.6%)	5 (23.8%)	21
Average	1 (4.8%)	20.5 (78.8%)	4.5 (17.3%)	26

Table 27: Verbal agreement in the old Bengalis' data

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 form. Both new and old Bengali informants produce very few tokens of subject-verb agreement.

5.5.1.2 Malayali informants

Tables 28 and 29 demonstrate the use/absence of verbal agreement by the Malayali informants.

New Malayalam	AGR Present (GA)	AGR Missing (GPA)	Verb Dropped (GPA)	Total
M1	4 (11.4%)	29 (82.9%)	2 (5.7%)	35
M2	0 (0%)	11 (55%)	9 (45%)	20
M5	0 (0%)	4 (30.8%)	9 (69.2%)	13
Average	1.3 (3.8%)	14.6 (56.2%)	6.6 (39.9%)	22.6

Table 28: Verbal agreement in the new Malayalam speakers' data

Old Malayalam	AGR Present (GA)	AGR Missing (GPA)	Verb Dropped (GPA)	Total
M3	0 (0%)	23 (76.6%)	7 (23.3)	30
M4	5 (11.6%)	34 (79.1%)	4 (9.3%)	43
Average	2.5 (6.9%)	28.5 (78.1%)	5.5 (15%)	36.5

Table 29: Verbal agreement in the old Malayalam speakers' data

The data show variation within the new group, particularly between M1 and the two other informants. Overall, there is less verb drop in the data of the older group. There are very few cases of subject-verb agreement in both groups.

5.5.1.3 Punjabi Informants

Tables 30 and 31 demonstrate 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
P1	2 (6.7%)	25 (83.3%)	3 (10%)	30
P2	0 (0%)	13 (65%)	7 (35%)	20
Average	1 (4%)	19 (76%)	5 (20%)	25

Table 30: Verbal agreement in the new Punjabis' data

Old Punjabis	AGR Present (GA)	AGR Missing (GPA)	Verb Dropped (GPA)	Total
P3	0 (0%)	12 (75%)	4 (25%)	16
P4	0 (0%)	16 (88.9%)	2 (11.1%)	18
Average	0 (0%)	14 (82.4%)	3 (17.6%)	17

Table 31: Verbal agreement in the old Punjabis' data

The new Punjabi informants seem to drop the verb less frequently than the newly arrived speakers from the other two language groups (i.e. Bengali and Malayalam). They also produce very few tokens of subject-verb agreement. Note that the old Punjabis do not produce any tokens under the AGR Present category. The length of stay seems to have a very limited effect on the Punjabi sample.

Overall, the data in tables 26 to 31 reveal that all informants rarely produce fully inflected verb forms that are marked for TMA and agree with the subject (i.e. the form used in GA). The data also suggest that Bengali and Malayali informants show a length-of-stay related development in the use of verbs: New Malayalam and Bengali informants drop verbs more frequently than their 'old group' counterparts, who seem to use more inflected, but less agreeing forms.

When investigating the agreement data, I noticed that the verb in GPA may take other forms which do not fall under any of the categories in (A), (B), and (C). The high rate of occurrence of the forms (D), (E), and (F) below made them worth discussing a separately.

(D) **PST**: Use of the GA verb root (past form) instead of the present or future form.

(E) **IMP**: Use of the GA imperative form of the verb instead of the GA inflected verb.

(F) **N**: Use of nouns in verbal function

Hence, in addition to the use of forms of the verb which agree with the subject, non agreeing forms of the verb, and dropping of verbs, GPA speakers may also employ the past form of the GA verb to refer to the present or future tenses, the imperative form of the GA verb instead of the indicative form, or a noun for a verbal function. In this section, I merely present the occurrence of the overall token numbers of the verb types (D-F). A detailed discussion of these verbal forms is in section 6.2.2.2.

5.5.1.4 Bengali Informants

Tables 32 and 33 show the frequency of the three strategies D-F produced by the Bengali informants:

New Bengalis	PST	IMP	N	Total
B3	22 (33.8%)	21 (32.3%)	22 (33.8%)	65
B4	11 (14.9%)	24 (31.1%)	42 (54.6%)	77
Average	16.5 (23.2%)	22.5 (31.7%)	32 (45.1%)	71

Table 32: Other verb forms in the new Bengalis' data

Old Bengalis	PST	IMP	N	Total
B1	24 (24.7%)	44 (45.4%)	29 (29.9%)	97
B2	13 (20%)	37 (56.9%)	15 (23.1%)	65
Average	18.5 (22.8%)	40.5 (50%)	22 (27.2%)	81

Table 33: Other verb forms in the old Bengalis' data

In average the new Bengali informants seem to use the noun for verbal function slightly more than the other forms (i.e. IMP and PST), while members of the old group use the imperative form of the GA verb more than the other two forms.

5.5.1.5 Malayali informants

The frequency of occurrence of the verbal strategies detailed in (D-F) produced by the Malayali informants are displayed in tables 34 and 35.

New Malayalam	PST	IMP	N	Total
M1	17 (30.4%)	23 (41.1%)	16 (28.5%)	56
M2	7 (19.4%)	6 (16.7%)	23 (63.9%)	36
M5	9 (32.1%)	8 (28.6%)	11 (39.2)	28
Average	11 (27.5%)	12.4 (31%)	16.6 (41.5%)	40

Table 34: Other verb forms in the new Malayalam speakers' data

Old Malayalam	PST	IMP	N	Total
M3	15 (23.4%)	34 (53.1%)	15 (23.5%)	64
M4	4 (17.4%)	10 (43.5%)	9 (39.1%)	23
Average	9.5 (21.8%)	22 (50.6%)	12 (27.6%)	43.5

Table 35: Other verb forms in the old Malayalam speakers' data

The old Malayalam speakers use the imperative form of the GA verb more than the other two forms. There is variation among the members of the new group. For instance, M1 produces the imperative form 41% of the time while M2 and M5 produce it in 17% and in 29% of the cases, respectively. Moreover, M2 differs from M5 in the sense that the former produces the N form 63% of the time whereas the latter produced it only in 39% out of the total number of tokens.

5.5.1.6 Punjabi informants

Tables 36 and 37 show the use of the three verb forms (D-F) by the Punjabi informants:

New Punjabis	PST	IMP	N	Total
P1	31 (31.3%)	51 (51.5%)	17 (17.8%)	99
P2	24 (36.4%)	30 (45.4%)	12 (18.9%)	66
Average	27.5 (33.3%)	40.5 (49.1%)	14.5 (17.6%)	82.5

Table 36: Other verb forms in the new Punjabis' data

Old Punjabis	PST	IMP	N	Total
P3	21 (23.9%)	48 (54.5%)	19 (21.6%)	88
P4	22 (25.9%)	41 (47.1%)	24 (27.6%)	87
Average	21.5 (24.6%)	44.5 (50.9%)	21.5 (24.6%)	87.5

Table 37: Other verb forms in the old Punjabis' data

Both new and old Punjabi informants show more use of the imperative form than the other two forms of the verb. The informants' length of stay in SA, however, seems to have no effect on the Punjabi informants.

Overall, the data in Tables 32 to 37 show that there is possibly a length of stay development in that the old Bengalis and Malayalam speakers seem to prefer the imperative form of the GA verb more than the two other forms. The Punjabi language group patterns slightly differently in that both old and new Punjabi members use the imperative form of the GA verb more than the two other forms.

Let us now move on to agreement in the NP and in the AP.

5.5.2 *Agreement in the NP and in the ADJP*

The data produced by GPA informants in this study reveal that adjectives in the GPA noun phrase or adjective phrase typically do not agree with their noun in gender and number. Instead, the singular masculine form is used with all nouns. In very few cases, however, some informants produced adjectives which agree with the noun in number and gender. In terms of demonstratives, the unmarked form is the singular masculine, which is usually used with all nouns and adjectives regardless of their gender. A less frequent alternative in the data is the use of a demonstrative which agrees with the noun in number and gender. In order to quantify the GPA system of agreement in the NP and in the ADJP, I have collapsed these strategies into the following two categories:

(A) **Agreement present**: any of the following[5]

    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).

(**B) Agreement missing**: Lack of agreement in gender and/or number in the adjective phrase or in the noun phrase.

Tables 38 to 43 tabulate the number of tokens for agreement/lack of agreement in the NP and in the ADJP for every informant group.

---

[5] Note that instances of agreement of the unmarked form (singular masculine) have not been counted because they do not reflect whether the informants actually apply GA NP/ ADJP agreement or whether they use an invariant form.

5.5.2.1 Bengali informants

The instances of the presence/absence of nominal agreement in the Bengali informants' data are shown in Tables 38 and 39.

New Bengalis	Agreement present (GA)	Agreement missing (GPA)	Total
B3	4 (8.6%)	42 (91.4%)	46
B4	0 (0%)	32 (100%)	32
Average	2 (4.3%)	37 (95.7%)	39

Table 38: Agreement in the NP and in the ADJP, new Bengalis

Old Bengalis	Agreement present (GA)	Agreement missing (GPA)	Total
B1	3 (8.5%)	32 (91.5%)	35
B2	1 (1.7%)	57 (98.3%)	58
Average	2 (5.1%)	44.5 (94.9%)	46.5

Table 39: Agreement in the NP and in the ADJP, old Bengalis

Both old and new Bengali informants use very few cases of agreement in the NP and in the ADJP. There seems to be no development between the new and the old Bengalis.

5.5.2.2 Malayali informants

Tables 40 and 41 present the tokens of nominal agreement/lack of agreement in the Malayalam speakers' data.

New Malayalam	Agreement present (GA)	Agreement missing (GPA)	Total
M1	2 (8.3%)	22 (91.7%)	24
M2	0 (0%)	8 (100%)	8
M5	3 (15.8%)	16 (84.2%)	19
Average	1.6 (8%)	15.4 (92%)	17

Table 40: Agreement in the NP and in the ADJP, new Malayali informants

Old Malayalam	Agreement present (GA)	Agreement missing (GPA)	Total
M3	8 (18.2%)	36 (81.8%)	44
M4	7 (26%)	20 (74%)	27
Average	7.5 (22.1%)	28 (77.9%)	35.5

Table 41: Agreement in the NP and in the ADJP, old Malayali informants

In the Malayalam language group, the predominant form is missing agreement. Yet, old members show some development in the acquisition of the GA nominal agreement system.

5.5.2.3 Punjabi informants

The tokens for agreement in the NP and in the ADJP in the data of Punjabi speakers are displayed in Tables 42 and 43.

New Punjabis	Agreement present (GA)	Agreement missing (GPA)	Total
P1	0 (0%)	20 (100%)	20
P2	3 (13%)	20 (87%)	23
Average	1.5 (6.5%)	20 (93.5%)	21.5

Table 42: Agreement in the NP and in the ADJP, new Punjabi informants

Old Punjabis	Agreement present (GA)	Agreement missing (GPA)	Total
P3	4 (21%)	15 (79%)	19
P4	2 (11.7%)	15 (88.3%)	17
Average	3 (16.5%)	15 (83.6%)	18

Table 43: Agreement in the NP and in the ADJP, old Punjabi informants

Punjabi informants produce very few tokens of agreement between the noun and the number, adjective, or the demonstrative. There seems to be a slight development in the acquisition of GA agreement system by old Punjabi informants.

To sum up the results for agreement in the NP and adjective phrase, the data in Tables 38 to 43 above demonstrate that – while caution is in order due to the very low token numbers – the amount of time the informants have stayed in the Gulf seems to have a very slight positive effect on the occurrence of agreement in the NP and in the ADJP for Malayalam and Punjabi informants, but not for Bengali informants.

After this rather general discussion of the five morphological features under investigation across members of the three L1 groups, let us now move to a more detailed discussion of these findings in the next chapter.

Chapter 6: Discussion

This chapter aims at further scrutinising the data investigated in Chapter 5, focusing on the target language for GPA speakers and the emergence of this Arabic-based variety. In the first section I discuss the input GPA speakers in Saudi Arabia receive and explore how it could have possibly influenced their use of GPA. Then, the GPA speakers' data are analysed in 6.2 in the light of the hypotheses listed in section 4.1.4. In the last section, I provide a theoretical discussion in which I attempt to link the results of this study with potential universal and substratal factors which have led to the emergence of the patterns manifest in my data.

6.1 What is it that GPA Speakers Acquire?

One vital question to ask at the beginning of this discussion is the following: what is the target language of GPA speakers once they arrive at the Gulf? It seems that there are two possible scenarios for language learning in the context of GPA, one has GA as target and the other has GPA as target. These two scenarios are dependent on the quantity and quality of input which GPA speakers receive during their stay in the Gulf. For a very limited number of immigrant workers in the Gulf, the target language seems to be GA. This is the case in the scarce instances of inter-marriages between locals and expatriates from the Indian sub-continent (see Bakkir 2010). GA can also be the target language in some cases where female maids live with a local family who mostly use GA when communicating with them. The second possible scenario for newly-arrived GPA speakers – which appears to be the case for the vast majority of GPA speakers including all the informants polled in this study – is receiving a GPA input with a very limited amount of GA input.

In research for my MA Dissertation (Almoaily 2008), I asked 77 Saudi respondents about their opinion on the following statement: 'I don't mind using GPA with speakers who are not fluent in GA'. This question revealed that the issue of using an altered form of Arabic when speaking to foreigners seems to be controversial for Saudis. Half of the respondents did not mind using GPA with non-Arabic speaking foreigners and the other half were either disagreeing or strongly disagreeing with this statement. 40% of the respondents answered with: *agree*, 4% answered with: *strongly agree*, 6.5% with: *I do not know,* 28.5% with: *disagree,* and 20.7% with *strongly disagree*. In reality, however, the use of GPA by locals when speaking to GPA speakers is likely to be higher than 50%, especially among the younger generation of locals. Indeed, this claim can be supported by the answers of the same Saudi respondents to an experiment, which I asked

them to undertake once they completed the questionnaire. They were put in the following

situation: *An Indian vendor asked you about the meaning of the following cartoon in*

GPA. The cartoon does not include any Arabic linguistic data that the Saudi respondents

could use in the description of the cartoon, apart from the two words لا تعليق 'no

comment' (see Appendix E). Thus, local informants had to create linguistic expressions

in GPA from scratch, and they readily did. Hence, there is evidence that immigrant

workers get GPA input from locals as well as expat workers. Note that the Saudi

informants only produced 619 words. In order to supplement their experimental data with

GPA data produced by a GA speaker, I analysed my own turns in my interviews with

GPA speakers.

In the remainder of this sub-section I will numerically investigate the data

produced by the Saudi respondents as well as the data I produced, assuming that they

represent the input GPA speakers receive from indigenous people of the Gulf. Hence, in

table 1 below, I investigate the use of GPA variants produced by 40 local speakers[1] per

1000 words. Note that the codes V1, V2, and V3 mean: variant 1, variant 2, and Variant

3. For example, there are two variants for definiteness in GPA, the GA-like variant *use of*

the definiteness marker al- (V1) and *zero definiteness markers* (V2). A short explanation

for V1, V2, and V3 if there is any is provided beneath every variable.

Feature \ Variants	GA	GPA	
	(V1)	(V2)	(V3)
Definiteness (V1=*al-*, V2 = Ø)	16 (13.2 %)	105 (86.7%)	
Conjunction (V1=*wa,aw,ya*,etc, V2=Ø)	26 (72.2%)	10 (27.7%)	
Copula (V1= Ø, V2 = *fi*)	32 (58.1%)	23 (41.8%)	
Nominal Agreement (V1=AGR, V2 = AGR-)	1 (2.3%)	42 (97.6%)	
S-V Agreement (V1= AGR, V2= AGR-, V3= Ø)	3 (7.1%)	38 (90.4%)	1 (2.3%)
Object and possessive pronouns (V1= bound, V2= free, V3= Ø)	1 (16.6%)	3 (50%)	2 (33.3)

Table 1: GPA data produced by GA speakers (per 1000 words)

If we assume that the data in table 1 are indicative of the input immigrants in

Saudi Arabia receive, the patterns produced by GA speakers match the data GPA

speakers produce – with the exception of the use of conjunctions. Indeed, both groups

(i.e. locals and expatriates) show a similar preference for using GPA variants over the

---

[1] Me and 39 Saudi participants in the questionnaire.

GA variants (compare the data in table 1 with the tables in Chapter 5, see also table 2 below). For example, the GA definiteness marker was used by locals (i.e. the 39 Saudi respondents and me) in 13.2% of the cases where it could have been used. New GPA speakers (i.e. who spent five years or less in the Gulf) used *al-* in 19.1%, while old speakers (i.e. who stayed for more than ten years in the Gulf) used the GA definiteness marker in 29.6% (see section 5.1). Similarly, nominal agreement was present in only 2.3% in the data of locals, while it was used in 6.8% in the data of newly-arrived GPA speakers and in 15.5% in the data of those who stayed longer in the Gulf (see 5.5.2). The GA fully inflected verb was used in 7.1% by locals. It was used in 3.1% by new speakers and in 3.8% by old speakers (see 5.5.1). Interestingly, locals were even more distant from the superstrate language, GA, than the expatriates with respect to the two features: *definiteness* and *nominal agreement.* This clearly suggests that the input GPA speakers receive is definitely not GA. Note that Wiswall (2002) reports a similar phenomenon, namely that locals produce less GA tokens than immigrants when speaking GPA (consider section 1.5 where I discuss potential causes of this phenomenon).

Conjunction is the only feature where there is a statistically significant shift to GA among the long-term residents, see 6.2.2 below. This could be explained by the finding that conjunction was the only feature where locals use its GA variant more than its GPA variant (72.2% versus 27.8%). Below I will suggest that this finding could be an argument in favour of the imperfect L2 learning hypothesis (see the discussion in 6.3.3).

Although table 1 suggests that GPA speakers receive a very limited input in GA, we should be careful with the findings listed in the table. These data were collected from a group of locals aged between 18 and 35. Local children, adolescents, and elders are not represented. It is my impression that GPA speakers receive more GA input from older locals, but I do not have empirical data to support this hypothesis. Second, the data are not fully representative: because GPA is a spoken medium of communication, asking informants to produce a written form of it might lead to the production of slightly different data than oral GPA. Despite these concerns, the data in table 1 still provide an indication that GA speakers mainly use GPA when addressing GPA speakers, with the exception of conjunction markers (see table 2 below). In the remaining variables, the locals' use of GPA features was relatively similar to that of the GPA speakers.

Table 2 below attempts to answer the question whether there is a noticeable shift towards a target, either GA or GPA, among the speakers investigated here. It seems that for some features, the speakers acquire the GPA patterns quickly, start shifting to GA and

fossilise[2] at some point. For other features, the speakers shift towards a variety different from GA. Note that I discuss the significance of these shifts in section 6.2.

Features	Percentage: GA feature[3] + (other variants)	
	New speakers	Old speakers
The definiteness marker *al*	19.1	29.6
Use of conjunction markers	7.6	23.3
No copula (in the present tense)	68.2	70.6
OBJ and POSS pronouns	bound 11.5	22.8
	free 41.8	44.8
Nominal agreement	6.8	15.5
S-V Agreement	3.1 (AGR- 58.2, dropped 38.6)	3.8 (AGR- 79.7, dropped 16.4)
word order	SVO 71 (SOV 8.7, VOS 1.6, VSO 1.8, OSV 10.6, OVS 5.9)	SVO 68.1 (SOV 12.3, VOS 1.2, VSO 4, OSV 6.8, OVS 7.3)

Table 2: What do GPA speakers shift to?

Comparison between the data of the new speakers with that of the long-term ones reveals that the following features fall under the category *slight shift to GA*: definiteness (19% vs. 29%), conjunction (7.6% vs. 23.3%), and nominal agreement (6.8% vs. 15.5%). But this development is slight and the GPA speakers interviewed in this study are still far from the target (see 6.3.3 for more discussion on the emergence of GPA from a language acquisition point of view). For other features, the development seems to be towards a variety different from GA. This seems to be the case for *object* and *possessive pronouns* and *subject-verb agreement*. Although there is an increase in bound pronouns (from 11.5% to 22.8%), this development is not to the detriment of free pronouns but of null pronouns. Free pronouns were used in the majority of cases by both old (44.8%) and new (41.8%) speakers. This strongly suggests that the free pronouns are a GPA feature (i.e. a property of a target language rather than a transitional stage or a case of failed acquisition of GA). Please be reminded that object and possessive pronouns in GA are represented as bound morphemes and that GA has SVO word order (see section 2.1.1.3). As regards *word order,* SVO was predominant in both the data of new and old speakers, but new speakers are slightly closer to GA norms than the old speakers (71% vs. 68.1%). This decrease in the rate of SVO among the old speakers is caused by an increasing use of

---

[2] We have to be careful with the term fossilisation as there is a minor development towards GA for some features. I have used this term because old speakers are still far from the GA target despite this minor shift.
[3] If there is only one variant in the cell, this means that the other variant is dropping the features (e.g. long-term speakers used the definiteness marker in 29.6%, which implies that they dropped it in 70.4%).

SOV word order (8.7% vs. 12.3%).[4] Finally, in the feature *subject-verb agreement,* the target is clearly not GA. The development is from dropping the verb (38.6% vs. 16.4%) to the use of AGR- verbs (58.2% vs. 79.7%).[5]

Given the different targets for the features investigated in this study, we cannot assert with certainty which target are the speakers orienting to (whether it is GA or GPA). In general, however, GPA speakers do not seem to be aware that GPA is distinct from GA at the syntactic, phonological, or morphological levels. As I show in the following quotes, the respondents claim that they know some Arabic and seem to be satisfied with the fact that they are able to communicate with locals in Arabic (GPA). For example, in my interview with the informant labelled B1, I asked him: Do you speak Arabic? He replied: *fi shwayyah Arabi* 'there is (i.e. I know) little Arabic'. I also asked M2 the same question, his answer was: *hina Arabi nus* 'here Arabic half' (i.e. I know some Arabic). M1's answer to the question: Did you learn Arabic when you moved to Saudi Arabia? was*: shwayyah shwayyah malom lakin alhen ziyadah* 'little little I know, but now better'. Perhaps the clearest statement I have in my data which shows that GPA speakers look at GA and GPA as one variety (i.e. Arabic) is made by M3 'people here speak quickly'. This statement suggests that GPA speakers, at least in the case of M3, conceive of GA as different from GPA only in terms of speed of delivery. Hence, it would be implausible to say that GPA speakers would have GA as their target language while they assume that the variety they speak/are acquiring (GPA) is not very different from the language which locals speak (GA). The generalisation that GA is not the target language for GPA speakers is supported by the fact that, as discussed above, even in the features which show a potential shift to GA, GPA speakers are still far from the GA target.

I discuss the informants' shift in more detail in the following section.

6.2 How Significant is Language Variation between GPA Speakers?

In Chapter 5, I introduced some general trends in the GPA data. In this section, I provide a detailed discussion of the factors *length of stay in the Gulf* and *L1* as potential factors for language variation in GPA. Following the investigation in the previous

---

[4] Note that non-target-like use of SOV order is something we would expect among the new speakers, given that their mother tongue, whether it is Punjabi, Bengali, or Malayalam, has SOV as basic word order. The fact that there is an increase (albeit only a slight one) in the use of SOV is unexpected, and hard to explain, see the discussion on GPA word order in section 6.3.

[5] As for the copula, no noticeable shift to GA or to a different variety can be seen. Compare the percentage of copulas in the present tense by new speakers (68.2%), with that of copulas used in the present tense be old speakers (68%). Please note that GA uses a null copular system in the present tense.

section, I examine the features which show a potential shift towards GA (i.e. definiteness, conjunction, and nominal agreement) in the first sub-section. Then I discuss the features which show a possible internal GPA development (i.e. pronouns and subject verb agreement). The copula is investigated in the third sub-section.

A theoretical discussion on the genesis of GPA – based on the findings of this study – is provided in section 6.3. In the reminder of this section, I compare speakers coming from different language groups (i.e. Bengali, Malayalam, and Punjabi) and new versus old speakers for every feature.

6.2.1 *Features showing a potential shift to GA*
6.2.1.1 Definiteness

As demonstrated in Chapter 3, Bengali is the only substrate language investigated in the current study which has a marker for definiteness. It was thus surprising – from a substratal point of view – to find that the Malayalam speakers use the definiteness marker more than the other language groups, despite the absence of definiteness markers in their L1. Compare the rate of occurrence of the GA definiteness marker, *al-*, in the data of Malayalam speakers (39.4%), with that produced by the Bengali sample (18%), and Punjabi-speaking subjects (15.6%). The possible effect from Hindi/Urdu as a second language for Malayalam and Punjabi speakers can be eliminated here due to the fact that Urdu does not have definiteness markers. Thus, the hypothesis that Bengali speakers use GPA definiteness markers more than their Malayalam and Punjabi-speaking counterparts is rejected.

The data has also revealed that there is considerable variation between speakers who belong to a single group, especially amongst the Malayalam and the Punjabi speakers. Compare, for example, the rate of occurrence of the GA definiteness marker in the data of P1 (10%) with that of P2 (20%), both of which are recently-arrived Punjabi speakers (see section 5.1.3). Also compare the overall percentage of the use of definiteness markers in the data of the old Malayalam speakers: M3 used it in 32.6% while M4 used it in 63.7% of the relevant NPs (see section 5.1.2). In section 4.5 I attempted to explain this variation among members of the same group, arguing that the choice among GPA variants might be affected by other factors which are hard to test such as the amount of exposure to GA, the informants' different aptitudes in language acquisition, and their dissimilar attitudes towards learning GA. The weak correlation between the informants' L1s and their choice of GPA definiteness variants might thus be

used as an argument against substratal effects on the emergence of GPA, but one has to consider the other morpho-syntactic features before making this claim.

As for the production of definiteness markers by the newcomers versus long-term residents, we notice a shift towards GA in using the definiteness marker *al-* averaged over the three linguistic backgrounds: the newly-arrived members produced the definiteness marker with an average of 19.1%, whereas those who stayed ten years or more produced it in 29.6% of the time. Although the difference is not statistically significant (p-value = 0.08), and even though there is obviously a vast amount of variation within the groups, there seems to be a trend towards the acquisition of GA norms. This slight 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 morpho-syntactic features that are easiest to learn as it only involves adding the prefix *al* − or one of its allophones − to the target noun. This assumption is supported by Sedlatschek's (2009: 49) hypothesis that: 'areas that are relatively difficult to acquire for learners of English may be particularly likely to undergo change'. For example, Agnihotri, et al. (1988) argue that the scarcity of the past perfect tense in Indian English student writing is due to the difficulty of the past perfect as compared to the present perfect. See the discussion on the genesis of GPA from an L2 acquisition view in section 6.3.3 below.

6.2.1.2 Conjunction markers

The description of the substrate languages in Chapter 3 above shows that the use of conjunction markers is optional in Bengali, whereas it is obligatory in Malayalam and Punjabi. Hence, in section 4.2.4 I hypothesised that Bengali speakers might drop conjunction markers more often than Punjabi and Malayalam speakers when speaking GPA.

The data in section 5.2 above have indeed revealed that the Bengali language group produced fewer tokens of the conjunction markers (only 9.4%) as compared to the Punjabi group (25.3%) and the Malayalam speakers (11.7%). Thus, the hypothesis that Bengali speakers drop conjunction markers more than the two other language groups seems valid. The chi-square test reveals that the difference between speakers of the three groups is significant at a p-value of 0.003. It is also clear from these figures that the Punjabi sample produces a noticeably higher number of conjunction markers as compared to the speakers of the remaining two languages. This could be due to the fact

that in Punjabi, just like in GA, conjunction markers are free morphemes, whereas they are suffixes attached to the noun in Malayalam and optional in Bengali.

The data also reveal a possible link between the years spent in GA speaking countries and the use of conjunction markers. This effect is attested in all three language groups. For instance, new Malayalam speaking informants used conjunction markers at an average of 5.6%, while the use of conjunction markers is much higher in the data of Malayali informants who spent a longer time in the Gulf (with 17.8%). Similarly, the newly-settled Bengali speakers used conjunction markers in 1.8%, while their old counterparts used them in 17%. In the Punjabi sample, the newcomers produced GA conjunctions in 15.6%, while long-term Punjabis produced tokens of conjunction markers in 35%. Overall, the newly-arrived speakers produced conjunction markers in 7.6% of the cases, whereas the old informants produced them in 23.3%. 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 0.002. Again, the potential explanation of this shift is that GA conjunction markers are not hard to learn. They are free morphemes and most of them are one-syllable words (e.g. *aw* and *ya* 'or' and *wa* 'and'). Another possible reason for this noticeable shift to GA among long-term residents is the input received from locals. As I have shown in table 1 above, unlike other features investigated in this study, locals seem to choose the GA variant for conjunction markers more than the GPA variants.

6.2.1.3 Nominal Agreement

In GA, the adjective agrees with the noun in gender and number and the demonstrative inflects for gender, number, and proximity (see section 2.1.1). In GPA, however, the bare adjective (i.e. singular masculine) is the unmarked form and GPA speakers add gender and number markers to the adjective only on very few occasions (see section 5.5.2). Similarly, only one demonstrative, *hatha* 'this.M', tends to be used with all objects, regardless of their gender, number, and distance. Thus, if long-term residents are found to produce more tokens of noun-adjective agreement in number and gender and more tokens of feminine or plural demonstratives this might be an indication of a shift towards GA. Note that, as stated in 5.5.2, tokens of agreement between the unmarked GPA form (i.e. singular masculine) and the adjective are not included into this account because their use by old speakers is not necessarily a result of a shift towards GA.

As for the substrate-language based comparison, Malayalam and Punjabi speakers are expected to have more cases of noun-adjective agreement. This is due to the fact that in Malayalam, the predicative adjective agrees with its head noun in person, number, and gender (Asher and Kumari 1997) and in Punjabi adjectives agree with their head in number and gender, except for loan words (Bhatia 1993). In Bengali, on the other hand, adjectives do not inflect for number or gender. Instead, the singular masculine form is used with all nouns (Milne 1993).

The results in Chapter 5 revealed that none of the informants polled in this study produced more than 10 tokens of nominal agreement out of 1000 words. In fact, the highest number of tokens is produced by the two old Malayalam speakers M3 and M4 (with 8 and 7 instances respectively). The remaining informants produced less than five tokens. Thus, we should be cautious when dealing with these small numbers. Despite this reservation, patterns can be noticed – especially when comparing the newly-settled informants with the long-term residents.

The Malayalam language group shows a slightly better performance in acquiring the GA nominal agreement system. They produced instances of nominal agreement in 15.1% of the total number of tokens. Compare this result with the percentage of Punjabi-speaking sample (11.5) and that of the Bengalis (4.7). Hence, the data is in parallel with our hypothesis that Bengali speakers have less agreement in the nominal system. This difference just fails to reach significance (p-value = 0.051), but a substratal trend is nevertheless noticeable.

As for the years of stay comparison, the data show that the number of years the informants have stayed in the Gulf seems to have a very slight positive effect on the occurrence of nominal agreement for Malayalam and Punjabi informants, but not for Bengali informants. Indeed, newly-arrived Bengali speakers produced nominal agreement in 4.3% of the potential NPs, while members of the old Bengali group produced instances of agreement in the NP in 5.1% of the time. The percentage of nominal agreement by newly-settled Malayalam speakers is 8%, while the old Malayali group produced 22%. Similarly, members of the new Punjabi group used instances of nominal agreement in 6.5%, while their old counterparts had nominal agreement in 16% of the cases. On average, the new informants produced AGR+ tokens in 6.6% of the total number of cases, while their long-term counterparts produced it in 14.6% of the total number of cases. Although the difference just misses statistical significance (p-value = 0.054), an indication of development can be seen, especially when comparing the data

of Malayalam and Punjabi speakers. More data, and hence bigger numbers, might achieve significant results in this area.

6.2.2 *Features showing development towards GPA norms*
6.2.2.1 Pronouns

In section 2.1.2, I have shown that there are four variants for object and possessive pronouns in GPA: (a) use of a GA bound pronoun which agrees with the noun it refers to in person, number, and gender, (b) use of a GA bound pronoun which does not agree with the noun, (c) use of a free pronoun (the same set of subject pronouns is used as object or possessive pronouns), and (d) dropping the object or possessive pronoun. The description of the three substrate languages in Chapter 3 reveals that object and possessive pronouns are free in all the three substrate languages polled in the current study. Hence, based on their L1s, informants are expected to have only minor differences in their choice among the four GPA variants for object and possessive pronouns, despite the fact that they come from three different linguistic backgrounds. In other words, on the basis of transfer effects, speakers of the three languages are expected to behave similarly in terms of producing high rates of 'free pronouns' as compared to their production of bound pronouns (refer to section 4.2.3).

The results presented in section 5.4 reveal that, on average, instances of bound morphemes are far fewer than the tokens of dropped or free pronouns. For example the total frequency of the two variants *agreeing* and *non agreeing bound pronouns* in the data of Bengali speakers is only 16%, while free pronouns are chosen in 47% and the pronouns are dropped in 36%. Similarly, the Punjabi informants used bound possessive and object pronouns (both agreeing and non agreeing) in a total of only 9.1%. Malayali informants – with the exception of M5 – seem to do better than their Punjabi and Bengali counterparts in using the GA bound pronouns. On average, Malayali informants used bound pronouns in 27.7% of the times they could use them. The difference between the three language groups is significant at a p-value of 0.002. This effect might be considered surprising since none of the substrate languages under investigation has bound pronouns. Note, however, that we should be careful with these percentages as they often represent small numbers of tokens. For example, the total number of bound object and possessive pronouns altogether in the data of the Bengali sample (4000 words) only amounts to 13. Similarly, the number of bound object and possessive pronouns in the data of the Punjabi sample is only 6. However, despite the small number of produced bound pronouns, there seems to be a development in the acquisition of bound pronouns: On average, newly-

settled informants in all three language groups produced bound object and possessive pronouns in 11.5%, while the long-term residents produced them in 22.78%. The difference is significant at a p-value of 0.03. Importantly, the overall shift is clearly not towards GA, as the use of free pronoun forms is predominant in the data of both new (41.8%) and old speakers (44.8%). 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 not GA (bound pronouns) but GPA (free pronouns). The fact that this GPA feature is learnt relatively early could be due to the fact that this feature is also found in the informants' L1s. Section 6.3.1 provides a more detailed theoretical discussion on the substratal influence on the emergence of GPA.

As can be seen in the descriptive account of GPA pronouns in section 2.1.2, the same set of pronouns is used as object and possessive pronouns. Moreover, the variants are the same (i.e. bound pronouns, free pronouns, and dropped pronouns). Yet, there are differences across the informants polled in this study as regards their choices among available variants for object and possessive pronouns. For instance, new Bengali speakers dropped object pronouns in an average of 93.7%, 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 6.3%. Possessive pronouns, however, were dropped in 17.3% and were used as free morphemes in 67% (see table 20 in Chapter 5). Different choices are also attested in the data of the old Malayali group, who used possessive pronouns as free morphemes in 58% of the total number of tokens of possessive pronouns and never dropped them where they could have been used. Object pronouns, on the other hand, were dropped in 62.5% and were not used as free morphemes by either of the two long-term Malayalam speakers (see table 23 in Chapter 5). Differences in the use of object and possessive pronouns are also found in the data of old Punjabis. They dropped object pronouns in 13.6% and possessive pronouns in 33%. These different choices across almost all language groups make it rather hard to generalise whether informants prefer dropping object and using possessive pronouns as free morphemes or vice versa. Looking for differences in the use of object and possessive pronouns across speakers of the three languages, however, is not relevant to the purpose of this study for two main reasons. First, both object and possessive pronouns are free in Bengali, Malayalam, and Punjabi. Hence, any difference in their use (e.g. using object pronouns as free morphemes and dropping possessive pronouns) cannot be explained by transfer of the pronominal systems of the substrate languages. The second reason is that both object and possessive pronouns are bound morphemes in GA. Hence, I have not seen any reason

to formulate a hypothesis predicting differences in the use of the two types of pronoun across the three groups. Overall, there is a tendency to drop object and use possessive pronouns as free morphemes[6], which must be a GPA internal development that cannot be explained with recourse to substrate or superstrate languages of GPA. The fact that pro-drop is common in the early stages of second language acquisition (see Eubank 1991, Towell and Hawkins 1994, and Montrul 2004) could support the potential role of language acquisition in the genesis of GPA. Yet, we have to remember that we are dealing with small numbers and that possessive pronouns are mostly used as free pronouns. Section 6.3 will revisit the emergence of GPA from a language acquisition perspective.

6.2.2.2 Subject-verb agreement

As detailed in Chapter 3, Malayalam is the only substrate language of the three which lacks subject-verb agreement (Asher and Kumari 1997). In Bengali, the verb agrees with the subject in person (Ray et al. 1966), and in Punjabi the verb agrees with the subject in number, gender, and person (Bhatia 1993). Hence, in section 4.2.1 I hypothesised that the Malayalam speakers would have less subject-verb agreement as compared to the Bengali and Punjabi speakers participating in this project.

At this stage I need to remind the reader that in GPA, the unmarked form of the verb is the GA third person singular masculine form. The data revealed that there are many other variants such as (1) the GA fully inflected verb, which agrees with the noun in gender, number, and person, (2) dropping the verb, (3) using the noun for verbal function, (4) using the imperative form of the GA verb, or (5) the verb root. Since our focus here is on agreement in the VP, I will start by discussing the results of variants relevant to agreement, namely agreeing GA verb, non-agreeing GA verb, and dropped verb.

Overall, the data in tables 26-31 in Chapter 5 reveal that all of the informants rarely produced the form used in GA (i.e. fully inflected verb forms that are marked for TMA and agree with the subject in number, gender, and person). The data also suggest that informants show a length-of-stay related development in the use of verbs: Members of the new group drop verbs more frequently (38.6%) than their old group counterparts (16.4%). We thus witness a move from dropping the verb to the use of the GA third person singular masculine form of the verb, which was used in 58.3% in the data of the

---

[6] Old Punjabi speakers and the informant labelled B2 do not follow this general pattern.

new informants and in 79.7% in the data of the old informants. The rate of dropping the verb is significantly higher in the data of new informants at a p-value of 0.0004. However, there seems to be no development in the acquisition of agreement in the GA verbal system. On average, new informants produced a fully inflected GA verb only in 3.1% of the total number of tokens, while old informants produced it in 3.8%.

The language groups do not display great differences in terms of choosing among the three variants. In all of the three substrate language groups, the variant *AGR-* (i.e. agreement markers are used, but the verb does not agree with the noun) is used in more than half of total number of tokens. Compare the overall percentage of *AGR Present* tokens (i.e. the verb agrees with the noun in gender, number, and person) in the data of Bengali informants (3.1%), with that of 5.3% in the Malayali sample, and 2% in the Punjabi data. Thus the hypothesis that Malayalam speakers use less S−V agreement can be rejected, as the data revealed that they use the GA fully inflected verb forms slightly more than members of the Bengali and Punjabi language groups.

As stated earlier in this sub-section, there are a number of variants of the GPA verb which cannot be linked to the acquisition of GA verbal agreement. These forms are: the use of the GA noun for verbal function, the use of the GA imperative form, and the use of the GA verb root (i.e. singular masculine past form of the GA verb). Results of the informants' choice among these forms are discussed below.

Comparing the percentages of each of the three variants reveals slight differences between language groups. Figures show that the Bengali sample used the noun for verbal function in 36%, the verb root in 23%, and the imperative form of the verb in 41%. Likewise, the Malayali group used the noun form as a replacement for the verb in 34.6%, the verb root in 24.6%, and the imperative form in 40.7%. The Punjabi language group was slightly different from the two other groups. They used the noun form in 21%, the verb root in 29%, and the imperative form in 50%.

A close look at the data, however, shows length-of-stay related patterns across members of the same language groups, see section 5.5.1. Indeed, there seems to be a general trend in shifting towards the imperative form across speakers of all the three language groups. For instance, the predominant variant in the data of the new Bengalis is the noun form, which was used in 45%, while old Bengalis prefer the imperative, with 50%. Similarly, Malayalam speakers shift from the noun form (41.5% in the data of newly-settled speakers) to the imperative form (50.5% by old speakers). M1, a newly-settled Malayalam speaker seems to have acquired the imperative form slightly earlier than other members of his subgroup, at 41.7%). A slight trend towards the use of the

imperative is even found in the data of the Punjabi language group, which uses this variant as a majority form (with 49.1% by the new speakers and 50.8% by the old speakers). The general shift towards the imperative might be due to the fact that this verbal form is the one GPA speakers hear most frequently in their contact with GA speakers. Hence, this development could be specific to the socio-pragmatic situation in which the pidgin is formed (cf. Bakker's 2011 review of Pidgin Madame).[7] As far as I am aware, these verbal forms (i.e. noun form, imperative, and the verb root) are not used in the substrate languages' verbal systems as alternatives to the fully-fledged verbs. Thus, variation in their usage cannot be interpreted as a result of substrate influence. In summary, the movement from using a noun for a verbal function to using the imperative form can be considered an internal GPA development. Moreover, the fact that inflection is absent in these three forms (i.e. noun, verb root, and the GA imperative verb) can be linked with the general features of pidgins and creoles discussed in section 1.3 above, particularly given that pidgins are characterised by a lack of inflection.

6.2.3 *Features that do not display a noticeable development*

In this sub-section, I discuss the potential substratal influence on the development of the copula, which did not show a development, neither towards GPA nor GA (as shown in section 6.1).

6.2.3.1 The copula *fi*

I have demonstrated in Chapter 3 that all of the three substrate languages have a copula. However, there are differences in the use of the copula across the substrate languages. In Bengali, for instance, the copula is only used with stage-level (i.e. temporary) predicates (see Finch 2001). In Punjabi, the copula is used with positive sentences only (see Bhatia 1993). In Malayalam, on the other hand, it is used with all predicates (see Asher and Kumari 1997). These differences are expected to have an effect on the informants' use of the GPA copula, *fi*. Thus, in section 4.2.5 I hypothesised that Malayalam speakers would produce more tokens of the copula *fi* than the Punjabi and the Bengali sample. Note that there is no copula in the superstrate language, GA, in the present tense. Thus, the discussion here will only take into consideration the use of the copula *fi* in the present tense in GPA. If long-term residents are found to drop the copula

---

[7] In this respect, my turns in the interviews do not fully represent the input GPA speakers in Saudi Arabia receive as I do not give any instructions, orders, or commands to the GPA speakers during the interview.

more than the newcomers, this might be an indication of a shift towards GA. If not, this might be an indication that GPA speakers shift to a form of GPA different from GA.

The data in section 5.3 above show that, on average, speakers in the Bengali sample dropped the copula *fi* in 69.8% of the total number of cases where a copula could have been used in the present tense. Similarly, Punjabi speakers dropped the GPA copula in 67.3%. Malayali informants, on the other hand, are found to drop the copula in the present tense in 71.1% of the total number of cases where a copula could have been used. Hence, the hypothesis that Malayalam speakers might produce more tokens of the copula than the speakers of the other two language groups cannot be accepted as the difference between the three groups is negligible.

The number of years of residency seems to have a slight negative effect on the use of the GPA copula *fi*[8] in the data of the Punjabi speakers. The new Punjabi-speaking informants dropped the copula with an average of 70.4%, whereas old Punjabis dropped it in 64.3%, which can be interpreted as a shift to a GPA-internal norm. The difference among the Malayalam speakers is negligible. Newly-settled Malayalam speakers dropped the copula in 72.7% and long-term Malayalam residents in the Gulf produced null-copula utterances in 69.5% of the times where a copula could have been used in the present tense. Contrary to these two language groups, the correlation between the years of stay and the shift towards GA seems to be slightly positive in the data of the Bengali sample. Compare the percentage of dropping the copula by newly-settled Bengali speakers (61.6%) with that of the old Bengali sample (78%). Although long-term Bengali speakers seem to have made a minor shift towards GA – unlike the two other groups – the rate of dropping the copula in their data is still relatively similar to that of members of all the three language groups. Since a copula is used in all of the three substrate languages, no substratal explanation for this difference between the Punjabi informants on one hand and the Bengali and Malayalam speakers on the other can be provided. Overall, except for possibly the Bengali speakers, there is no reportable shift towards Gulf Arabic in the data of speakers participating in this study regarding the use of a copula, as new speakers dropped it in an average of 68.2% and old speakers dropped it in 70.6%.

6.2.4 *Summary of the discussion on the significance of L1 and length of stay effects*

I will first discuss the results in table 3, which summarises the hypotheses related to substrate-based effect and indicates whether the data of this study support them or not.

---

[8] Please be reminded that the more informants drop the copula the closer they are to the superstrate language, GA

Table 4 displays the features investigated in this study and shows whether there is a shift towards the superstrate language, GA. Table 5 lists GPA internal developments.

The data revealed that the following hypotheses can be	accepted	rejected
Bengali speakers use the definiteness marker more than the Punjabi and Malayalam speakers		✓
Bengali speakers drop conjunction markers more than the two other language groups.	✓	
Malayalam speakers use the copula *fi* more frequently than the Punjabi and Bengali speakers		✓
No difference in using the object and possessive pronouns, as all the three substrate languages use free pronouns		✓
Malayalam speakers have less subject-verb agreement compared to Bengali and Punjabi speakers		✓
Bengali speakers use less nominal agreement than the speakers of Punjabi and Malayalam		✓ (T[9])

Table 3: Summary of substrate language-based hypotheses

Is there a significant shift towards GA?	yes	No
Acquiring the GA definiteness marker *al*		✓ (T)
Acquiring the GA conjunction markers	✓	
Dropping the GPA copula *fi*		✓
Acquiring the GA S-V agreement		✓
Acquiring the GA nominal agreement		✓
Using SVO word order		✓

Table 4: Summary of informants' shift towards GA

As regards the length of stay, we can also ask whether there has been a significant GPA internal shift	yes	no
Pronominal system (from dropping to free OBJ and POSS pronouns)	✓	
S–V agreement (from dropping the verb to AGR- form)	✓	

Table 5: Summary of informants' GPA internal shift

Thus, with the exception of conjunction markers, there seems to be at best a weak correlation between the morpho-syntactic properties of the speakers' substrate languages and the choice that native speakers of these languages make among the available variants of a variable. This weak substrate effect might be understood in terms of theories which assume that contact languages emerge as a result of universal cognitive processes, rather than being influenced by the morpho-syntactic systems of the superstrate, substrate or

---

[9] (T) Means that there is a trend which fails to reach significance.

adstrate languages (see Ferguson 1971, Todd 1974, Bickerton 1981, Muysken and Veenstra 1995, and Singh 2000). Yet, there are negative correlations between the substrate languages and variation in GPA which I find difficult to explain, even when resorting to Universalist theories. For example, despite the fact that Bengali is the only substrate language investigated that has a definiteness marker, Bengali speakers produced fewer instances of definiteness markers (18%) than Malayali informants (39.4%).

The potential substratal and universal factors leading to the genesis of GPA are further investigated in the subsequent section.

6.3 What can the Results of this Study Tell us about the Emergence of GPA?

In section 1.1.1, I reviewed a number of theories on the genesis of pidgins and creoles such as substratal influence and Universalist theories including *imperfect second language acquisition* and the *Language Bioprogram Hypothesis* (LBH). Since the current theories of pidgin and creole genesis are mainly based on the investigation of Indo-European pidgins and creoles and since these theories are still subject to controversy, I have suggested that more evidence should be sought using the data of lesser studied pidgins and creoles such as the Arabic-based pidgin in this thesis. If the morpho-syntactic systems of the tested substrate languages of GPA did not show a significant effect on the informants' choice among GPA variants, this could be used as an argument against substratist theories. By the same token, if GPA were to display linguistic features similar to Indo-European pidgins and creoles despite its different superstrate and substrate languages, this could be an argument in favour of Universalist theories. At this stage, I need to point out that these assumptions have failed to capture the complex nature of GPA. Indeed, as detailed below, it is difficult to eliminate any of the two potential factors (i.e. substrate influence and universal influence) leading to GPA evolvement. For instance, the coexistence of *serial verbs* in GPA and in the substrate languages could support the potential substrate role in the emergence of GPA. At the same time, Universalists claim that serial verbs are a property of contact languages generally. Similarly, the fact that GPA shares many features with Indo-European pidgins and creoles (such as pre-verbal TMA particles and analytic morphology) could be conceived of as evidence in favour of universal factors leading to the emergence of pidgins. But at the same time, the fact that old GPA speakers produce slightly more GA tokens that the new ones in definiteness and nominal agreement can be explained in terms of theories of adult second language acquisition, namely that there are constraints on second language

acquisition, leading to fossilisation in the acquisition of the target language. For example, Arends and Veenstra (1995: 129) write 'A quite well-developed view holds that creoles are really the result of gradual stabilisation and expansion of jargons by second language learners'. Language acquisition, substrate, and universal factors on GPA genesis are discussed with more detail in sections 6.3.1, 6.3.2, and 6.3.3 below.

This difficulty of eliminating any of the two potential factors leading to the emergence of GPA (i.e. universal and substratal factors) is not surprising in view of the continuous debate about the genesis of pidgin and creole languages discussed in section 1.1.1 Indeed, as can be seen in the discussion below, proponents of all the three theories can use the data of this study to support their view. It thus seems that a theory that appeals to more than one factor for the emergence of contact languages is best placed to capture the complexity of the situation. 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.

Attributing more than one factor to the genesis of contact languages (especially substratal and superstratal) has often been referred to as 'the Cafeteria Principle', randomly selecting features found in the substrate languages and in some dialects of the superstrate languages and attempting to link these features with the linguistic systems of the pidgin and creole languages (see Dillard 1970). Mufwene (2001: 78), however, defends the complementary hypothesis by suggesting that it 'has been hurriedly dismissed by misinvoking the Cafeteria Principle'. What differentiates the complementary hypothesis is that it tries to propose a 'set of principles that account for how competing forms and constructions have been selected into the new vernaculars' (Mufwene, ibid: 78). Indeed, although the complementary hypotheses might be regarded as eclectic, it 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. In the remainder of this section I will discuss the potential universal, substratal, and language acquisition factors leading to the emergence of GPA, based on the results I reported in Chapter 5 and discussed in the two sections above.

6.3.1 *Substrate influence on GPA*

The discussion in 6.1 reveals that a significant correlation between the informants' L1 and their choice among GPA variants can only be found in their use of conjunction markers. The participants' L1 also seems to have an effect on their GPA speech with respect to *nominal agreement.* Whereas this correlation does not reach the arbitrary significance level (i.e. a p-value of 0.05), there is nevertheless an observable trend (p-value = 0.051). There is no relation at all between the substrate languages and GPA variants concerning two linguistic features: *definiteness* and *verbal agreement.* Finally, while we expected no difference between speakers of the three substrate languages with respect to using object and possessive pronouns, speakers of Malayalam produced a significantly higher number of bound pronouns. This difference cannot be linked directly to the structural properties of the pronominal system of Malayalam, which has free object and possessive pronouns. Indeed, this finding invalidates the hypothesis that the pronominal systems of the substrate languages will influence pronoun use among GPA speakers participating in this study. In summary, a significant relation between the informants' L1s was found in one feature only (i.e. conjunction), and a trend was found in one feature (i.e. nominal agreement). No relation is found in three features (i.e. definiteness, copula, and verbal agreement), Furthermore, a significant difference (p value = 0.03) was found between the speakers of the three language groups in terms of their choice between variants in the GPA pronominal system, despite the relatively similar systems of their L1s. In table 6, I list the GPA features which show a potential substratal influence, and show their deviation from the cut-off significance point (0.05).

Feature	Conjunction	nominal agreement
p-value	0.003	0.051

Table 6: Potential substrate influence on language variation in GPA

In sum, although substrate influence is statistically evident in only one morpho-syntactic feature, it cannot be entirely ruled out, for a number of reasons. First, there is a clear substratal trend in the feature *nominal agreement*. This is especially pertinent since the results are based on a very low number of tokens. A larger corpus might provide a clearer view of the substrate influence on GPA variation. Moreover, there are linguistic features of GPA – other than the six features above – which may be the result of substrate influence. I will discuss these potential substratal features (i.e. serial verbs and verb-final word orders) in the remainder of this sub-section.

As pointed out above, serial verbs have been proposed to be a universal feature of creole languages (see Bickerton's 1981 LBP Hypothesis). This claim, however, is

questioned in the literature (see Seuren 1984 and Siegel 2008a). Importantly, serial verbs are a common feature found in the syntax of all substrate languages investigated in this study (see Paul 2003 for Bengali, Steever 1987 for Malayalam, Bhatia 1993 for Punjabi, and Schmidt 1999 for Urdu). Hence, Smart (1990) and Bakir (2010)[10] suggest that the existence of serial verbs in GPA (such as: *fi* + verb, *sawwi* 'make' + verb, and *ruuh* 'go' + verb, see examples 1-3 below) could be a result of substrate influence. I also find these constructions to be common in my data. The fact that these three verbal constructions do not exist in the lexifier language can thus be taken as evidence of substrate influence on the verbal system of GPA (Smart 1990, Bakir 2010). The examples 1-3 below are produced by the informants participating in the current study:

(1) Examples of *fi* + verb:

A. Jumah	subh	ma	fi	yi-shtgil (B1)
Friday	morning	no	COP	3.SGM-work

'I do not work on Friday morning'

B. kurah	fi	shuf	Saudi	kurah (M3)
Ball	COP	see	Saudi	ball

'I watch Saudi football' (Saudi Football League)'.

(2) Examples of *sawwi* 'make/do' + verb:

C. Ruuh	inshallah	sawwi	zawaj (B4)
Go	godwilling	make	wedding[11]

'God willing, I will get married when I go (home)'.

D. Ay	nafar	yi-ji	huwa	sawwi	taleem (P2)
Any	person	3SGM-come	he	make	learning

'Any person who registers in the school, they teach him'.

(3) Examples of *ruuh* 'go' + verb:

E. Ruuh	sajjal	maktab (M1)
Go	register	office

'(Employees) register in the office'.

F. Ana	ruuh	safar (P3)
I	go	travel

'I (want to) travel back home'

---

[10] Both authors refer to these verbal constructions as *compound verbs*. I call them *serial verbs* because they are known by this name in the literature on Urdu (see Schmidt 1999) and on creole languages (see Muysken and Jansen 1978, Mühlhäusler 1986, *inter alia*).

[11] *Wedding* is a noun which is used here for a verbal function (B4 expresses his plans to get married, not his intention to make a wedding party).

In table 7 below, I summarise the use of serial verbs by the speakers polled in this study:

		sawwi + verb	fi + verb	ruuh + verb	Total
New Bengalis	B3	0 (0%)	18 (100%)	0 (0%)	18
	B4	0 (0%)	7 (87.5%)	1 (12.5%)	8
Old Bengalis	B1	3 (16.6%)	13 (72.2%)	2 (11.1%)	18
	B2	2 (14.2%)	10 (71.4%)	2 (14.2%)	14
New Malayalam	M1	2 (7.4%)	22 (81.4%)	3 (11.1%)	27
	M2+M5	0 (0%)	9 (100%)	0 (0%)	9
Old Malayalam	M3	1 (6.6%)	13 (86.6%)	1 (6.6%)	15
	M4	2 (20%)	8 (80%)	0 (0%)	10
New Punjabis	P1	1 (3.3%)	26 (86.6%)	3 (10%)	30
	P2	3 (9.3%)	29 (90.6%)	0 (0%)	32
Old Punjabis	P3	0 (0%)	38 (97.4%)	1 (2.5%)	39
	P4	0 (0%)	32 (94.1%)	2 (5.8%)	34

Table 7: Use of serial verbs by GPA speakers

It is noticeable in the table above that the most common first verbal component in the data of all the three language groups is *fi*. On average, it was used in 82% of the total number of serial verbs produced by the Bengali sample, in 87% by the Malayali sample, and in 92.2% by the Punjabi sample. A possible explanation for this high use of the *fi* plus *verb* serialisation is that *fi*, unlike *sawwi* and *ruuh*, is often used as a habitual marker (see the examples 1A and 1B above). In other words, GPA speakers use *fi* + *verb* more often because of their need to express the meaning of performing an action habitually. In another visible trend, the speakers in the Punjabi sample use the serial verbs significantly more than the two other groups (p-value = 0.003). Compare the average number of serial verbs used by the Punjabi-speaking informants (33.7) with the Bengali (14.5) and the Malayalam (15.2) speakers. I cannot find an explanation for this phenomenon as the 'be + verb' serial verbs are not only found in Punjabi (see Bhatia 1999). They are also used in Malayalam (see Frohnmeyer 1989) and Bengali (see Basu and Wilbur 2010). Indeed, if the preponderant use of *fi* + *verb* by the Punjabi sample was due to substratal effects, then their use would be relatively similar to that of the Malayalam and the Bengali speakers because this type of serial verbs also exists in Malayalam.

Another potential substratal effect on GPA is the use of SOV and OSV word orders. Note that all the substrate languages of GPA investigated in the current study have SOV as their basic word order (see Rasinger 2007, Bhatia 1993, Bhatia and Koul 2000, and Dayal and Mahajan 2004). GA, on the other hand, is robustly a VO language, where SOV order is exceedingly rare. Thus, the GA sentence in (4) sounds ungrammatical.

(4) *Ali   it-tuffaha   kal

   Ali   DEF-apple   ate

SVO is by far the most frequent word order in transitive sentences in GPA, reflecting the order in the lexifier language GA. But other word orders, including SOV, VOS, VSO, OSV, and OVS, are also found in the informants' data. The following examples are taken from participants in my study:

(5)   ana   ya-kil   kabsah (B1)

        I   3SGM-eat   kabsah (SVO)

        'I eat kabsah' (a Saudi dish)

(6)   ana   inta   kalam (P1)

        I   you   speech (SOV)

        'I spoke to you'

(7)   yi-shtiri   sayyarah   ana (M4)

        3SGM-buy   car   I   (VOS)

        'I bought a car'

(8)   kitab   ana   ma   yi-rif (M1)

        book   I   not   3SGM-know (OSV)

        'I cannot read'

This flexibility in GPA word order seems to be a function of substratal effects. Given that the basic order of the speakers' L1 is verb-final, we expect to see a significantly higher rate of verb-final orders (i.e. SOV and OSV) than other orders, not counting SVO, as this is the predominant order in the lexifier language. We also do not expect to see any significant differences among the three language groups in this regard. Table 8 displays the percentage of the informants' use of GPA word orders:

	SVO	SOV	VOS	VSO	OSV	OVS
Bengali	71.6%	14.1%	3%	4.7%	3.3%	3%
Malayalam	65.5%	10.6%	0%	1.3%	11.9%	10.6%
Punjabi	71.6%	6.8%	1.3%	2.6%	11%	6.3%

Table 8: Average use of GPA word orders

The data in table 8 above show that speakers of all the three language groups use the GA word order (SVO) in more than two-thirds of their GPA output. Nevertheless, if we examine the remaining variants in the table above (i.e. other than SVO) we find that the verb-final word orders (SOV and OSV) are used more often than the non-verb-final orders. This could be due to a minor substratal effect. Compare, for instance, the use of verb-final word orders in the data of Bengali speakers (17.4%) with that of VOS, VSO,

and OVS (10.7% altogether). Similarly, the Malayali sample use the SOV and OSV word orders in 22.5%, while their combined use of VOS, VSO, and OVS is only 11.9%. Finally, the Punjabi language group use the verb-final word orders in 17.8%. Compare this percentage with their use of object and subject-final word orders, other than SVO. There are some exceptions to this norm, however. For instance, the subject-final word order OVS is used as frequently as the SOV word order by the Malayalam speakers (10.6%). Moreover, the Bengali speakers' use of the OSV word order is low. [12]

Although the findings of this study do not straightforwardly support substrate influence on GPA variation, the existence of serial verbs and SOV word order in GPA makes it rather hard to discount the potential substratal effect on the emergence of GPA. It could be noted, though, that in at least one of the serial verb constructions considered above, the non-lexical item is an aspect marker. Free morphemes as aspect markers are a characteristic creole feature as discussed in 6.3.2 below.

6.3.2 *The emergence of GPA from a Universalist point of view*

As investigated in 1.1.1 above, Universalist theories of pidgin and creole genesis attempt to explain the similarities found in the grammars of pidgin and creole languages by relating them to 'universal aspects of the human linguistic capacity' (Muysken and Veenstra 1995: 121). The descriptions of GPA (see Smart 1990, Naess 2008, Almoaily 2008, Alshammari 2010, Bakir 2010, and section 2.1.2 of this thesis) reveal that its linguistic system manifests some of the proposed general features of pidgin languages listed in section 1.3. Indeed, GPA is characterised by features which are also found in Indo-European language-based pidgin and creole languages. Some of these traits are typical of both pidgins and creoles such as reduced inflectional and derivational morphology, a reduced lexicon of content words as well as function words. Other features such as TMA adverbs, reduplication, and serial verbs are thought to be typical of creoles. Finally, GPA has a relatively free word order. This feature is thought to be typical of pidgins (see Holm 1988, Romaine 1988, Sebba 1997, and the discussion in section 1.3).

Sebba (1997) suggests that there are three universal principles involved in the process of pidginisation. In short, these universal principles are: constraints on adult

---

[12] Word order in GPA is promising for interesting insights and more research should be done on it in the future. For example, we may also choose to generalise over SOV, OSV and OVS, distinguishing the class of OV languages (see Dryer 2011). We may also investigate the high use of OVS word order by the Malayalam speakers, which was used as often as the SOV word order.

language learning, semantic transparency, and language simplification. The first universal is discussed in more detail in sub-section 6.3.3. The discussion in the remainder of this sub-section will mainly focus on the second and third principles.

a) Semantic transparency

The term *semantic transparency* is used to refer to compounds whose meaning is easily guessed by the meanings its constituents (see Baayen and Schreuder 2003 and the discussion in 1.3). GPA seems to abide with this universal principle of pidgin languages. For instance, the GA words *sˤɪɣɪːr* 'small' and *kɪbɪːr*[13] 'big' are combined with other words to create new GPA meanings. Thus a GPA speaker can use the compound *omur kabiir* (B1) 'age big' to form the meaning for the word *ʕədʒuːz* 'elderly'. Similarly, the compound *baggalah kabiir* (M4) 'shop big' is used to create the meaning for the Arabic term *suːg* 'supermarket'. In another example of transparency in GPA, informant P2 used the compound *shajaraat saghiir* 'trees small' to refer to *shatlaat* 'sapling'.

b) Language simplification

Siegel (2004: 140) lists some examples of simplicity in pidgin and creole languages. He writes '[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.' Indeed, as detailed below, GPA is characterised by reduced inflection, a reduced lexicon compared to the contributing languages, allowing many word orders, and reduplication. Hence, these features could exist in GPA due to the universal principle *language simplification*. Below I discuss some GPA features displaying simplicity:

Reduced inflection:

As shown in section 2.1.2, the morpho-syntax of GPA displays a heavily reduced verbal and nominal system as compared to the superstrate language. An example of this reduction in the verbal system is the generalisation of the third person singular masculine prefix *yi-* at the expense of other verbal inflections such as *ti-* (3SGF), *a-* (1SG), *-iin* (2SGF), and *-uun* (2PL). Another example of reduced inflection in GPA is the generalisation of the singular masculine form (bare stem) for adjectives to also cover

---

[13] Often pronounced as /səgɪːr/ and /kəbɪːr/ by GPA speakers.

singular feminine, and plurals, see 2.1.2 for a wider discussion of reduction in GPA as compared to GA.

Reduced lexicon

Mühlhäusler (1997) and Siegel (2008), define simplicity in pidgin and creole languages quantitatively: the small size of a pidgin lexicon is a feature indicating simplicity in GPA. There are various examples of reduction in the lexical system of GPA as compared with GA. For instance, one can hardly find synonyms in GPA, whereas synonymy is common in GA. For example the word 'hurry up' can be translated into GA as *bsɪrʕah*, *ɪχlɪsʕ*, *ɪstəʕdʒɪl*, and *yəllah*; whereas in GPA it can only be translated as *sʊrah*. An example of polysemy in GPA is the word *hurmah* which means 'woman' in GA, but is also used in GPA as 'wife', in addition to its GA meaning. Similarly, the word *baba* means 'dad' in GA, but it also means 'employer', in GPA. Moreover, the word *mama* 'mom' has other meanings in GPA such as 'female employer'. It can also be used as an honorary title given to elderly women. The GA existential *fi(h)* has also gained a new meaning in its copula function in GPA and as a marker of habitual aspect.

Small inventory of function words

In section 1.3, I have argued that the traditional view of pidgin and creole languages is that they are characterised by a lack of copulas, definite and indefinite articles, and by reduced pronominal systems. Apart from use of a copula, GPA seems to be in line with these assumptions. For example, the number of personal pronouns in GPA is five, while in GA it is eighteen. I have also shown in section 2.1.2 that in GPA, the GA demonstrative *hatha*[14] 'this.SGM.PROX' is used with all objects in GPA, regardless of their distance, number, and gender. The remaining five GA demonstratives[15] are hardly – if at all – used in GPA.

c) Free Word-order

Bakker (1995) claims that pidgins are not confined to a certain word order. In GPA, the most common word order is SVO, but other word orders such as SOV, VOS, VSO, OVS, and OSV are also found across the data of GPA speakers (see 6.3.1).

---

[14] Often pronounced by GPA speakers as /hæzə/.
[15] *hathi* (close SGF), *hatholi* (close PL), *hathak* (far SGM) *hathiik* (far SGF), and *hatholiik* (far PL)

d) Other typical creole features found in GPA

In addition to the features above, GPA exhibits some features claimed to be typical of creoles only such as serial verbs (Bickerton 1981, discussed in 1.1 above), TMA adverbials, and reduplication (see Bakker 1995). Since serial verbs are also found in the substrate languages, their existence could be considered supporting evidence for the substrate theory of pidgin and creole genesis (see 6.3.1). Indeed, the fact that GPA has TMA adverbials – despite being a contact language spoken by adults only – could be considered as evidence against the LBH, which assumes that this feature, among others, exists in creole languages as a result of children's use of their innate language capacities to transform the pidgin input to a creole language (see section 1.1).

Adverbs to mark for TMA[16]

Section 2.1.2.1, has demonstrated that, even though one can find examples of verbs in the past or present form in the GPA data, these forms are not always used with the same temporal reference, as in the lexifier (e.g. reference to the present in GPA can be expressed with a verb of a past form and vice versa). Hence, I have argued that – similar to other contact languages – tense in GPA is marked by the use of adverbials such as *awwal* 'first', *alhen* 'now', and *baaden* 'later' if the reference to the tense is not clear from the context.

Reduplication

According to Bakker (1995), reduplication is rare in pidgins and common in creoles. Note that there are plenty of examples of reduplication in GPA. For instance, the word *nʊs* 'half' is repeated (i.e. *nʊs nʊs*) to convey the meanings: 'not so good' and 'not complete'. Similarly the word *miy:ah* 'one hundred' is reduplicated (i.e. *miy:ah miy:ah*) to create the meaning 'perfect'. Moreover, the word *sawa* 'together' is repeated (i.e. *sawa sawa*) to form a new meaning 'two or more similar things'.

The fact that GPA shares these features with most Indo-European language-based contact languages can be taken as evidence in favour of Universalist theories of pidgin and creole genesis. In other words, the striking similarities between the structure of GPA and the structures of Indo-European language-based pidgins and creoles, despite the different contributing languages, could be the result of similar universal processes leading to the emergence of pidgins. Indeed – with the possible exception of serial verbs

---

[16] Please note that aspect can also be expressed via the existential marker *fi*, as a habitual marker. The discussion here is on adverbial makers only.

and the relatively high proportion of verb-final order – the features above cannot be linked to the linguistic systems of the substrate languages. Nominal and verbal inflections, for instance, are common in the morpho-syntactic systems of Bengali, Malayalam, Urdu, and Punjabi (see Chapter 3). Similarly, the four languages above have rich inventories of pronouns. I have shown in Chapter 3, for example, that possessive pronouns in Malayalam inflect for person, case, number, gender, politeness, and proximity. Similarly, possessive pronouns in Urdu inflect for person, number, politeness, and case. I have also shown in Chapter 3 that – with the exception of Malayalam – all the substrate languages of GPA have a form of subject-verb agreement. Yet, the data of this study demonstrate that all informants participating in this study rarely produce the GA fully inflected verb. Hence, the fact that GPA has developed widespread features of contact languages that cannot be linked to its substrate languages could support Universalists claims that contact languages emerge as a result of universal parameters, and not due to substratal influence. Yet, as discussed at the end of the previous section, the potential substratal role on the emergence of GPA cannot be totally eliminated. Furthermore, we need to consider in the next sub-section the emergence of GPA from a language acquisition point of view.

6.3.3 *The potential role of imperfect L2 acquisition in the genesis of GPA*

The imperfect L2 acquisition theory of pidgin and creole genesis can be grouped with the Universalist theories of genesis (see Muysken and Veenstra 1995, Mufwene 1986, Sebba 1997). Nevertheless, since my thesis aims at probing the possibility that language variation in GPA results from substrate influence and 'incomplete' L2 learning, I have opted to discuss this theory in a separate section. I have reported in 1.1 that some researchers such as Mufwene (1990) and Siegel (2008b), link pidgin and creole genesis with imperfect second language acquisition. Indeed, the findings of the current study can be of interest to creolists for three reasons. First, speakers of GPA are acquiring and using this contact variety during a period ranging from just a few months to more than twenty-five years. Thus, observing the progress of this contact language would be of help to creolists involved in the gradual vs. abrupt creole emergence debate (see section 1.1.1). Second, GPA is a living, and possibly evolving, pidgin. Thus, collecting accurate data on this pidgin is possible (see the discussion in 1.4 on the limited/scarce data of dead pidgins and creoles as a limitation in the field of pidginisation and creolisation). Finally, this pidgin is only acquired as a second language by adults. Hence, there are no chances for creolisation for this pidgin from an LBH point of view. At the same time, it is

possible for GPA to become a creole if we follow competing views to the LBH discussed in 1.1

Although the idea that contact languages are the result of imperfect L2 acquisition dates back to the 1880s (see section 1.1.1), researchers seem to have failed to reach an agreement as to the exact role of the second language acquisition processes leading to pidgin and creole genesis. Siegel (2008a: 208), for example, 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'. By the same token, Klein and Perdue (1997: 340) believe that *pidgins* and *the basic* variety, which they define as: 'adult second language learners (outside the classroom) universally develop a well-structured, efficient, and simple form of language' (1997: 301), are two different things. Despite their view that they are distinct, the authors conclude that 'there are certainly similarities, but it is quite unclear how far-reaching these are.' (Klein and Perdue (1997: 340). What is challenging in the attempt to link the genesis of pidgin and creole languages with L2 acquisition is the different nature of acquisition in the two cases: indeed, treating the superstrate language as an L2, or L3, L4, etc. in complex multilingual context might be misguided for two reasons: (i) pidgin speakers might not have a similar input from the superstrate language compared to that of second language learners (see Andersen 1983, also see table 1 above), and (ii) pidgin speakers might not have similar attitudes and motivations towards learning the target language as those of the second language learners. With these two observations in hand, we can still draw a parallel between the genesis of contact languages and some existing theories in the field of second language acquisition. For example, the results of the current project presented in Chapter 5 and analysed in section 6.2 seem to be in line with Klein and Perdue's *basic variety* theory on adult L2 acquisition. This theory assumes that – with the exception of vocabulary learning – adults acquiring a second language stop progressing after they learn the 'basic' language necessary for communication. As discussed in section 6.2, long-term GPA speakers made a significant shift to GA in only one feature, conjunction. Table 9 also shows the trends in the acquisition of GA in two features (i.e. definiteness and nominal agreement, consider also table 4 above). Note that the significance level is 0.05.

Feature	Definiteness	Conjunction	Nominal agreement
p-value	0.08	0.002	0.054

Table 9: Informants' shift towards GA after spending 10 years or more in the Gulf

There seems to be a slight development in the acquisition of GA definiteness and nominal agreement, but this difference is hardly reportable for two reasons. First, it is not statistically significant[17] (p-value= 0.08 for definiteness and 0.054 for nominal agreement). Second, even with the existence of a slight development, the long-term residents are still far from the target language (see the discussion on fossilisation below). For instance – on average – GA nominal agreement is used only in 14.5% by the speakers who spent ten years or more in the Gulf. Similarly, members of the old groups produced the GA definiteness marker in only 29.6% of the total number of times where a definiteness marker could have been used. As for subject-verb agreement, all of the informants participating in this study produced a very limited number of GA subject-verb agreement patterns and no development in the acquisition of GA verbal agreement could be detected.

Also, the newly arrived GPA speakers show a slightly better performance – compared to the members of the old group – in terms of the use of GA-like null-copula constructions. In addition to the lack of development in the acquisition of the target language (if we assume here that GA is the target language), GPA shares other features reported to be typical to Klein and Perdue's *basic variety* such as reduced inflectional morphology, lack of complex constructions, and preference for SVO word order. However, as table 10 reveals, it is hard to classify GPA as a basic variety as it shows a significant development internal to GPA in two features. This was discussed in section 6.2, see table 9 below.

Feature	OBJ and POSS pros (null to free)	Null verbs to AGR-
p-value	0.03	0.0004

Table 10: Informants' GPA internal shift after spending 10 years or more in the Gulf

Hence, it might be more reasonable to interpret the data as a movement towards a non-GA variety, namely a pidgin-internal target.

Other terms referring to the 'failure' (Han 2004: 5) of adults to fully acquire the second language are *fossilisation* (Selinker 1972) and *ultimate attainment* (Birdsong 1999). Han (2004: 28-36) reports a number of possible explanations for fossilisation in adult second language acquisition, some of which are: (a) 'absence of corrective feedback', (b) low 'quality of input', (c) 'automatisation of faulty knowledge' (i.e. the production of incorrect language due to lack of information), (d) 'lack of sensitivity to input' (i.e. difficulty in learning target language features), (e) 'change in emotional state',

---

[17] Please note that these results are based on very low token numbers.

(f) 'satisfaction of communicative needs', (g) 'avoidance' (i.e. fossilisation is the result of avoidance rather than false language acquisition), and (h) 'will to maintain identity'. The impact of these possible factors is sometimes even stronger in the case of pidgins than it is in the case of second language acquisition. For example, pidgin speakers can be expected to have less corrective feedback than language learners who are more integrated in the target language community, and certainly less than language learners in the classroom. Moreover, the quality of input is different in the two cases. As discussed in section 6.1, input to pidgin speakers by speakers of the superstrate language – as in the case of GPA – is mostly in the pidgin rather than the superstrate language. This is not the case, certainly not to the same extent, in contexts where the L1 learner is operating in the target language community. Similarly, the impact of communicative needs seems to be stronger in the case of contact languages than it is in L2 acquisition.[18]

The lack of development in the acquisition of the superstrate language even after spending more than ten years in the Gulf can also be explained by the Fundamental Difference Hypothesis (see Schachter 1988, Bley-Vroman 1989), which argues that first and second language acquisition are two different processes. Gass and Selinker (2008: 164) describe this hypothesis as follows: '[i]n second language acquisition (at least in adult second language acquisition), not only is "complete" knowledge not always attained, it is rarely, if ever, attained'. Thus, unlike children acquiring their first language, adult second language learners – such the GPA speakers interviewed in this study – are not expected fully acquire the superstrate language, GA.

6.3.4 *Concluding remarks*

As shown in the three subsections above, both competing theories of the genesis of contact languages, i.e. the substratist and the Universalist (including the imperfect L2 acquisition) are supported by the data of this study and can help us explain different aspects of the emergence of GPA. The contribution, though, seems to be unequal in the case of this contact language. For example, substrate influence is minimal as it significantly influenced the speech of GPA speakers in one feature only (i.e. conjunction). Yet, substrate influence cannot be totally ruled out as it appears to have caused minor substratal effect in some features such as nominal agreement and word order. It should be noted that the three languages tested in this study (i.e. Bengali,

---

[18] The fact that GPA speakers, especially the newly-settled ones, produce verbless clauses (see table 2 above) could be considered a further argument for a potential imperfect language acquisition role on the emergence of this contact variety. The production of verbless statements is common in the speech of second language learners (see Selinker 1996, Schachter 1988).

Malayalam, and Punjabi) are typologically relatively similar. Punjabi and Bengali in particular are quite closely related as they belong to the Indo-Aryan language family (see Chapter 3). This may be the reason why almost no substrate effects have been found by comparing them. If 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. As for the Universalist theories, the Bickerton line of Universalist theory has problems, too, given that GPA has a mix of pidgin and creole features despite the fact that it is spoken/acquired by adults only. But a less strict Universalist theory gets ample support by my findings, as shown in sections 6.3.2 and 6.3.3 (e.g. semantic transparency, simplification, and fossilisation in L2 acquisition). The joint application of substratal (though their effect seems to be minor) and universal factors may hold in the emergence of contact languages more generally, which may explain in part the ongoing debate on the evolution of pidgin and creole languages. Certainly, proponents of both views can equally support their theories using data of pidgin and creole languages. One good example in the case of GPA is serial verbs. While they are proposed to be a universal feature of contact languages, (see Bickerton 1981, Bakker 1995) serial verbs can be claimed to be existing in GPA as a result of substrate influence, as they are existent in the syntactic systems of Bengali, Urdu, Malayalam, Punjabi, as well as other Dravidian and Indo-Arian languages (see the discussion in 6.3.1). This 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. As such, eclectic or more encompassing approaches to pidgin and creole genesis such as the complementary hypothesis (see Baker and Corne 1986, Hancock 1986, and Mufwene 2001, 2006, and the discussion at the beginning of this section), seem to be providing a more convincing explanation for the emergence of pidgin and creole languages.

Conclusion

The main aim of this thesis was to investigate language variation in GPA resulting from the morpho-syntactic differences in the speakers' L1s and from their length of stay in the Gulf. The thesis also aimed at examining supporting evidence for the competing theories of pidgin and creole genesis. Hence, this study contributes to the literature of less-described non-Indo European pidgins and creoles. It further provides useful insights for researchers interested in language variation and change in general and the evolution of modern Arabic in specific.

Compiling the corpus for this investigation was the most challenging part. First, I had to collect GPA data from speakers who meet certain criteria (i.e. have spent either five years or less or more than ten years in the Gulf and speak Bengali, Malayalam, or Punjabi as their first language). When finding GPA speakers who meet these requirements I had to overcome an even bigger obstacle, which is convincing the GPA speakers to participate in the interview. Many simply refused to be interviewed and many others could not make it because they were too busy. Also, transcription and extraction of tokens was not an easy task. It took me three and a half to four hours to transcribe only ten minutes of speech, as the transcription was done in three stages: listening to the whole interview, transcribing it, and revising my own transcription. Once I finished transcribing the interviews, I compiled a corpus of 12,000 words out of these interviews (see figure 1 in Chapter 4). In order to make the tokens easier to access and retrieve, I devised a list of glosses for every variant, e.g. COP + (for *copula used*) and COP − (for *copula not used*), refer to section 4.4.4 for a full list of these glosses.

In the first chapter I briefly discussed some common theories on the emergence of pidgins and creoles. Then, I provided a historical overview on the development of the field of contact languages. I also discussed the general linguistic features of pidgins and creoles and reviewed the literature of GPA and other Arabic-based contact languages. It seems that the literature of pidgin and creole languages suffers from fallacies resulting from an Indo-European centric view and lack of consensus in defining different forms of language contact. For example, the definitions in the literature for the terms jargon, pidgin, and creole make it difficult to classify GPA into one of these categories. Hence, I redefined the term *pidgincreole,* which was first introduced by Bakker (2008), to fit contact languages carrying features typical to pidgins and to creoles at the same time (see 1.4.3).

Since it was rather impractical to analyse linguistic variation in GPA without providing an illustration of the target features (i.e. definiteness, conjunction, copulas,

object and possessive pronouns, and verbal and nominal agreement) in GPA, GA, and the substrate languages, namely Bengali, Malayalam, and Punjabi, Chapters 2 and 3 aimed at describing the contributing languages in the emergence of GPA. Proposing a description of the Bengali, Malayalam, Punjabi, and Urdu – despite the fact that I do not speak any of these substrate languages – was a challenging task because not all available descriptive accounts are accurate. For example, Müller-Gotama (1994) claims that Malayalam has free word order, while it is mainly a verb-final language (see Asher and Kumari 1993). Hence, I had to verify some of these data with linguists speaking the substrate languages.[1] Contrasts between the morpho-syntactic systems of the substrate languages were used to formulate the hypotheses of this study in Chapter 4. For example, all the three substrate languages use conjunction markers, but Bengali is the only language where the use of conjunction markers is optional. Hence, I expected Bengali speakers to use GPA conjunction markers less frequently than the speakers of Malayalam and Punjabi. As regards the length of stay, I expected long-term speakers to produce more GA tokens than the newly-settled GPA speakers (see section 4.2.6). It should be noted that some factors such as exposure to GA, willingness to learn it, and different language learning abilities, were impossible to control in this study. Effects from these factors could have had an effect on the informants' choice between the selected features' variants (see 4.7 for a detailed discussion on this limitation).

An analysis of the data revealed rather complex results. Hence, I divided Chapter 6 into three sections. In the first section I attempted to discover the target language for long-term GPA speakers. It seems that GPA speakers shift to GA in three linguistic features: definiteness, conjunction, and nominal agreement. However, this shift is significant in only one feature, namely conjunction (p-value=0.002). Interestingly, my research suggests that GA speakers use GPA when speaking to GPA speakers in all the features investigated in the current study, apart from this very feature. This finding suggests important avenues for future research regarding the role of input in pidgin/creole formation. As regards the effect of the informants' L1 on their GPA output, a significant relation was, again, only found in one feature (conjunction, p-value = 0.003). Other features either show a substratal trend, such nominal agreement (p-value = 0.051), or do not show substratal effects at all, such as nominal and verbal agreement, copula, and object and possessive pronouns. Despite the weak substratal role in the emergence of GPA, however, substrate theories of genesis still find some evidence in the

---

[1] Many thanks to Jay Jayaceelan (Malayalam), Wim van der Wurff (Bengali), and Nadeem Bukhari (Punjabi).

data of this study such as the existence of serial verbs and verb-final word orders. Yet, more evidence can be found in favour of Universalist theories of pidgin and creole genesis such as reduction and language transparency. Hence, I suggested that my data support Mufwene's (1993) complementary theory of genesis, which claims that universal as well as substratal factors can contribute to the emergence of contact languages.

I conclude this project with a set of recommendations for future research on this pidgin language. First, conducting a substrate-language based phonological variation analysis of GPA might reveal more substratal differences between its speakers than an analysis that concentrates on morpho-syntactic variation. Indeed, one can easily trace phonological features between a given pidgin language speakers to their L1s. Morpho-syntactic phenomena, on the other hand, can be more complicated. For example, TMA markers are considered to be a universal feature by some researchers and a substrate feature by others, which make it rather hard to explain the existence of TMA markers in GPA. Second, more substratal differences might appear when comparing linguistically dissimilar substrate languages to the Indo-Aryan or the Dravidian languages investigated in this study. It might be interesting, for instance, to compare the GPA production of Malayalam speakers with speakers of Tagalog or Indonesian. Third, comparing the GPA production of male speakers with that of female speakers might reveal gender-variation in this pidgin. Finally, I suggest considering the impact of this pidgin on GA. Indeed, there are potential pieces of evidence for lexical as well as morphological effects of GPA on GA that are worth investigating. For example, the word *si:dah* 'across the street, straight', which is widely used by GPA as well as GA speakers, is borrowed to GA from Urdu. Similarly, the word *nafar* is used in Classical Arabic for 'a group of people'. In GPA and GA, however, it is used for 'one person'. This different meaning of the word *nafar* could, again, have been borrowed to GA from Urdu. As for potential morphological effects, one finds some plural forms in GA that are similar to those found in GPA. For instance, the plural form of *riyal* 'the Saudi currency' is hardly ever pronounced as *riyal-at*. Instead, the GPA plural form (i.e. no agreement between the noun and the numeral in number) is used in GA (e.g. *θalaθah riyal* 'three riyal', instead of *θalaθat riyal-at* 'three riyals'). The reason why the GPA-like pluralisation is used in GA with the word *riyal* but not with other GA words is possibly because GA speakers mostly have daily monetary transactions (e.g. buying groceries) with GPA speakers. The extensive use of this GPA-like pluralisation of *riyal* could have resulted in GA speakers shifting from the GA-pluralisation to the GPA-pluralisation. Yet, to date we lack quantitative evidence to support this claim.

Appendix A. Transcription of Interviews

(One example interview per language group: B1, M1, and P1)

Transcription codes:

Lengthening	:
Non Arabic words	( )
Pause	-
Quick turns	=
Overlap	[ ]
Raised intonation	?
Falling intonation	.
Laughter	@
Transcriber's comment	" "
Tanscription of non-participant data	X

Interviewee: B1

Participants:
• Interviewer: Mohammad Al-Moaily, age 29, male, lived 26 years in Saudi Arabia, 3 years in the UK. Parents: from Saudi Arabia. 1st language: Arabic.
• Interviewee: B1, age 39, lived 29 years in Bangladesh and ten years in Saudi Arabia. Parents: From: Bangladesh. 1st language: Bengali.
Recording:
23 minutes and 55 seconds. In Batha community Centre: Riyadh, Saudi Arabia.
Interview:
The participants did not know each other before the interview. The interviewee showed confidence and willingness to take long turns and therefore provided excellent data. The interview went so smoothly that interviewer did not ask most of the interview schedule questions.
Punctuation:

Mohammad: طيب.. بسم الله.. اا من وين.. انتا من وين؟

B1: انا من بنقلاديش

Mohammad: من بنقلاديش.. وين اي مكان في بنقلاديش

B1: بنقلاديش هذا.. دكا كوته..

Mohammad: يعني.. دكا

B1: دكا فوته.. دكا خلاص هذا شوف.. بعيد.. قريب. هذا..

Mohammad: طيب.. في بنقلاديش ايش لغة.

B1: بنقلا.. هذا

Mohammad: بنقلا؟

B1: نعم

Mohammad: كلو بنقلاديش كلام بنقلا

B1: بنقلا

Mohammad: مافي لغة ثاني

B1:لا مافي لغة ثاني بس شوي اختلاف مافي لغة ثاني شويه اختلاف هذا (سيم سيم) جيزان؟ (سيم سيم) تبوك

Mohammad: ايوه

B1: (سيم سيم) دمام اختلاف خفيف بس هذا اختلاف

Mohammad: لغة يعني (سيم سيم)

B1: بنقلا

Mohammad: بنقلا.. يعني الهند في لغة ثاني

B1: هند.. لغات مليان.. يمكن تسعطعش.. تسعة عشرين لغة

Mohammad: ايه كثير.. لكن بنقلاديش بس بنقلا

B1: حمد الله واحد لغة

Mohammad: ايه.. طيب انتا في بنقلاديش ايش شغل

B1: انا بنقلاديش شغل بس ادرس. ادرس خلاص؟ يجي سعودية

Mohammad: يعني بس خلص دراسة بعدين يجي.. قبل كم سنة يجي؟

B1: انا يجي هنا قبل هذا تسعة سنة..

Mohammad: تسعة سنة.. طيب هنا ايش شغل

B1: انا شغل الحين الكترونيات

Mohammad: معرض الكترونيات=

B1: نعم..

Mohammad: طيب=

B1: في هنا

Mohammad: طيب بنقلاديش كيف دراسة يعني

B1: بنقلاديش دراسة.. مليان هذا شسمه.في اختلاف.. في مدرسة في (سكول) هذا (سكول) اسم هذا (سكول) هذا هنا ادرس.. شوي عربي.. انجليزي وهذا شسمه انتا ركب حسابات.. وتاني بنقلا.. بس هذا كثير كذا بعدين شسمه هذا (ساين).. (بوقولو).. هذا كذا اكتب.. مدرسة عربي.. مدرسة عربي

Mohammad: في عربي

B1: دين.. نعم.. مدرسة عربي؟ سنة ادرس؟ في ناس. حفظ القران.. في حفظ الحديث.. في مفسر قران.. في اذان.. كلو شي عربي.. يعني رسول صلى الله عليه وسلم.... اي: طريقة.. هو هنا يدرس. (سكول)؟ في شويه عربي الحمد لله انا برضه كويس.. انا في شوية (كلاس فايف).. روح.. (فايف فنش).. روح هذا (سيكس) على طول لازم هو يدرس عربي شوية كلو.. بنقلاديش الحمد لله بنقلاديش هذا طريقة كذا.. بزروة ممكن اثنين سنة ثلاثة سنة.. هو صبح؟ صبح بعد فجر صبح بعد فجر الحمد لله هو روح مسجد.. بابا سوى سوى.. جيب اختي جيب اخو هو جيب مسجد.. مسجد في؟ .. اثنين ساعة هو تعليم سورة.. قراءة.. حديث؟ كيف سوي صلاة.. كيف سوي وضوء.. كيف سنة.. كيف فرض.. كيف سن.. اا واجب.. كل شي هو هذا تعليم.. صغير

Mohammad: تعليم بالبنقلا ولا بالعربي

B1: لا بنقلا في هذا؟ هو سوي لغة بنقلا لكن كلو عربي

Mohammad: ايه

B1: كلو عربي حديث.. عربي.. قران.. عربي لكن شوي كلام هو؟ يعني هو فاهمني؟ بس.. بنقلا

Mohammad: ايوه

Appendix A

B1: كلو عربي

Mohammad: ايوه.. طيب ماشالله يعني انتا في.. في تعليم في بنقلاديش في تعليم عربي

B1: حمد لله

Mohammad: طيب=

B1: كلو ناس.. لحظة يا شيخ.. كلو ناس.. ممكن انا في بزروة.. سنة.. هو سوي شوي شوي هو كلم.. انا هو تعليم كلام.. السلام عليكم هو روح انا سوى سوى اكل.. انا قول قول بسم الله.. بس.. شوية شوية تعليم.. هو يجي كبير.. خمسة سنة ستة سنة مافي مشكلة صغير هو عمر روح مسجد.. سوى سوى بابا سوى سوى اخو.. سوى سوى عم.. روح صلي كذا كلو مس؟لم؟ الحمد لله

Mohammad: الحمد لله

B1: الحمد لله

Mohammad: طيب انتا يجي للسعودية في مشكلة عربي؟

B1: بس انا شوية شوية فاهمني.. اول مرة انا مافي فهم.. كثير عربي.. حمدو لله.. الحين شوية شوية فهم الحمد لله

Mohammad: يعني انتا في كذا بنقلاديش اتعلم عربي.. يجي هنا.. في مشكلة فهم عربي

B1: بس.. فهم ايش عربي يا استاذ.. بنقلاديش عربي لغة.. شوف.. انا هنا بنقلاديش ادرس ماعون.. انا هنا يجي حصل موية

Mohammad: ايه

B1: هذا اختلاف

Mohammad: ايوه.. صح

B1: بنقلاديش انا حصل هذا.. تاجون.. انا.. يجي هنا حصل طاقية..

Mohammad: طاقية صح..

B1: هنا لغة خفيف شوية.. عشان هذا لغة عربي.. مافي هذا كتب لغة هذا لغة عربي هو كلم.. فهم شوية شوية بس كلو ناس فهم

Mohammad: طيب انتا في زواج

B1: حمدو لله

Mohammad: كم (بيبي)

B1: انا اثنين (بيبي)

Mohammad: كلو ولد ولا=

B1: لا.. الحمد لله واحد بنت وواحد ولد

Mohammad: كم عمر (بيبي)

B1: انا.. بنتي عمر.. عمر تسعة سنة

Mohammad: ماشاء الله

B1: ماشاء الله

Mohammad: يعني زواج انتا زمان.. طيب.. انتا كم عمر.. كم عمرك

B1: انا عمر.. تسعة ثلاثين

Mohammad: تسعة ثلاثين.. ماشاء الله.. طيب في هنا نفر قريب في السعودية

187

B1: بس انا جيت هنا سعودية.. انا مدام اخو.. هو الحين انا مافي موجود.. هو روح بنقلاديش.. هو اول شغل هذا شركة الجميح.. هذا شغل خلاص خلاص؟ هو كلم انا ما ابغى شغل انا روح.. بعدين جميح كلم خلاص انتا روح؟ خلاص

Mohammad: يعني زوجة انتا في بنقلاديش مافي هنا.. مافي اخو ما في

B1: في انا اخو اثنين اخو كبير.. انا صغير

Mohammad: كلو في بنقلاديش

B1: نعم.. كلو في بنقلاديش

Mohammad: مافي فكر يجي للسعودية

B1: والله فكر لكن يا استاذ هذا الحين ما حصل (فيزا.. فيزا) غالي

Mohammad: ايه

B1: ما حصل..عشان انا مسكين.. انا في اهل.. انا حصل راتب انا لازم جيب اهل.. انا ماما الحين.. شيبة الحمد لله في موجود سعودية شوية تعبان لازم انا علاج انا (بيبي) اثنين هو روح مدرسة انا مدام.. بس.. الحمد لله سعودية انا يجي.. والله سعودي.. والله انا شوف سعودية بركة بركة مية مية.. اي ناس امشي سيدة.. والله يا شيخ انا شوف.. انا امتحان انا شوف كثير اي ناس يمشي سيدة؟ الله جيبو خير بركة كلو شي.. عشان ايش؟ هذا سعودية شوف حمد لله كلو شي حصل

Mohammad: الحمد لله

B1: سعودي؟ .. مافي نفر كثير.. لكن الحمد لله سبحانه وتعالى.. جيب كلو شي سعودية

Mohammad: كلو.. كلو اكل موجود

B1: كلو اكل موجود.. وبرضة ثاني ايش.. هذا بركة ايش؟ حمدو لله رسول صلى الله عليه وسلم هنا اول موجود.. كلو صحابة موجود.. حديث هنا موجود.. هذا هذا قران نزل هنا.. حمدو لله.. هذا سعودية بركة.. خير وبركة

Mohammad: الحمد لله.. طيب انتا في اول شغل في مكان ثاني

B1: انا شغل اول مكان ثاني..

Mohammad: وين

B1: شغل اول بقالة في سويدي

Mohammad: لا لا انا انا كلام يعني مثلا دمام او جدة

B1: لا لا

Mohammad: بس رياض.. مافي مدينة ثاني

B1: لا لا

Mohammad: مافي امارات دبي

B1: لا لا لا

Mohammad: بس بنقلاديش يجي رياض

B1: بنقلاديش انا يجي رياض الحمدو لله الحين برضو موجود في رياض

Mohammad: طيب في اختلاف رياض انتا قبل تسعة سنة عشرة سنة يمكن او..

B1: والله الحمد لله اختلف انا شوف اول في مافي انا.. مافي عمارة الحين الحمد لله في كثير كلو شي.. في موجود.. وبرضه الحين ملك عبدالله رحمه الله.. حفظة الله كلو شي موجود مسجد.. هذا كلو شي موجود.. الحمدو لله كلو ناس سعادة.. والله العظيم هذا ملك عبدالله الله جيب هو طويل عمر.. بنقلاديش هو كثير.. كثير كثير مساعدة

Appendix A

Mohammad: ايوه

B1: كثير كثير مساعدة.. الحمد لله.. الله جيب هو جنة الكبير.. جنة الفردوس ان شاء الله

Mohammad: ان شاء الله.. ان شاء الله.. طيب بنقلاديش في.. في تغيير

B1: ايش؟

Mohammad: يعني قبل.. يعني قبل ثلاثين سنة بنقلاديش (سيم سيم) الحين ولا في اختلاف

B1: لا والله الحين الحمد لله بنقلاديش.. كثير دولة

Mohammad: الحمد لله

B1: هذا ناس يجي سعودية عمارة كلو شي ناس ودي فلوس شوية شوية الحمد لله الحين.. كويس

Mohammad: حمدو لله

B1: الحمد لله

Mohammad: طيب بنقلاديش مين في.. مين في.. شخص ناس مشهور يعني ناس معروف كلو عالم معلوم

B1: مافي معلوم

Mohammad: يعني (famous people in Bangladesh)

B1: ايش

Mohammad: في.. في واحد.. يعني مثلا عالم في بنقلاديش او

B1: انا حبو.. مولانا.."voice not clear" عشان ايش.. معلوم هو مية انا ما دخل.. الحمد لله سبحانه وتعالى هو جيب.. قران ترجمة.. مية مية الحمد لله.. الله جيب هو خير وبركة والله طويل العمر هو

Mohammad: ترجمة الى بنقلا

B1: الى بنقلا.. انجلش.. هندي.. عربي..

Mohammad: ماشاء الله

B1: اردو.. فارسي.. سبحان الله.. الله جيبو خير وبركة.. الله جيبو طويل عمر

Mohammad: ما شاء الله.. طيب انتا كيف كلم اهل في بنقلاديش

B1: انا كلم تلفون

Mohammad: تلفون.. جوال

B1: جوال

Mohammad: كم.. كم مرة في الاسبوع

B1: والله يا شيخ في.. هو ابغى شي هو جيب انا (مسد كول) انا كلم.. مافي كلام كيذا لا.. هو جيب ابغى انا كلم اي شي انا كلم.. هو ابغى انا مدام انا بزورة انا ماما سوي جيب نغمة (مسد كول). انا كلم.. روح (كبينة) الحين (كبينة) رخيص الحمد لله.. خمسين هللة.. وبرضة موبايلي جيب تخفيض.. سوى جيب تخفيض.. حمد لله

Mohammad: حصل..

B1: حمد لله كويس

Mohammad: طيب انتا.. طيب.. انتا كلام اول شي كلام.. بنقلا فيه شوية اختلاف.. انتا كلام ايش في اختلاف عن

B1: بس.. في سعودية.. فيه ناس كلام قبل.. واحد ناس كلام هذا لف واحد.. واحد ناس كلام لفة.. واحد ناس كلام لفة.. هذا كذا شوية اختلاف مافيه زيادة

Mohammad: مافيه اختلاف كبير يعني

Appendix A

B1: لا لا مافي اختلاف

Mohammad: طيب انتا كم كلم عربي في اليوم.. كلم كثير ولا شوي

B1: بس حمدو لله شوية شوية

Mohammad: لا. يعني كم ساعة يتكلم عربي في اليوم

B1: والله. انا ما يدري ممكن انتا سؤال انا شوية شوية كلم.. فاهمني انا كلم لكن=

Mohammad: لا اذا انتا في. في شغل انتا في محل الكترونيات

B1: نعم نعم

Mohammad: في يجي عربي كوستمر

B1: حمدو لله يجي كثير

Mohammad: يجي كثير. انتا كلم..

B1: عربي.. في ناس موجود هذا سعودي.. في ناس موجود كلم انجليزي.. هو سعودي الحمد لله.. في.. كثير ش هذا.. كثير كثير كلام انجلش ما شاء الله.. كويس هذا.. في ناس كلام عربي

Mohammad: ايه

B1: في اختلاف

Mohammad: طيب انتا في يتعلم عربي سرعة سرعة ولا.. شوية

B1: والله.. في ناس كلم ممكن انا فهم هذا لغات قران.. كذا.. انا فهم شوية شوية.. سرعة سرعة.. لكن.. لغات هذا سعودية

Mohammad: مم

B1: شوية شوية مافي كثير

Mohammad: ايه.. يعني انتا اول ما تجي.. اول ما يجي بنقلاديش.. في كلام عربي

B1: لا مافي

Mohammad: مافي..

B1: مافي

Mohammad: مافي معلوم مرة؟ بعدين هنا يتعلم.

B1: لا لا

Mohammad: بعدين هنا يتعلم.

B1: هنا يتعلم

Mohammad: كم شهر يعني عشان يتعلم

B1: والله.. ممكن انا.. يجي انا.. بس انا شوف اسمع.. انا شوف نفر كلام.. جيب ورقة

Mohammad: مم

B1: انا شوف هو ايش كلام؟ شوف؟ اسمع.

Mohammad: مم بعدين=

B1: اثنين. انا شوف؟ برضه يسمع بس انا شوف هو قول ورقة.. ورقة كيف هذا ابيض كذا. انا ماشي.. بس كذا

Mohammad: طيب كم شهر يعني كم شهر بعدين انتا معلوم عربي

B1: بس.. مافي زيادة ممكن ستة شهر الحمدو لله

190

Mohammad: مافي كثير يعني؟

B1: لا

Mohammad: ستة شهر بعدين معلوم.. طيب انتا شوف تلفزيون عربي؟

B1: حمدو لله انا شوف كثير

Mohammad: شوف تلفزيون عربي

B1: شوف

Mohammad: اا.. مافي شوف بنقلا

B1: بس في خبر في وقت موجود.. انا.. في يسمع خبر بعدين.. مافي وقت؟ مافي..

Mohammad: بس اكثر شي شوف تلفزيون عربي ولا=

B1: بس.. اكثر انا شوف ايش.. معلوم انا دوام موجود هنا.. في تلفزيون.. في انا مدير شوف اخبار

Mohammad: ايه.. طيب انتا في فايدة عربي.. معلوم يعني في فايدة انتا شوف تلفزيون عربي

B1: الحمد لله=

Mohammad: فهم؟

B1: مافي فهم.. ممكن هو كلام هذا.. هذا فلبين كذا كذا.. شوية شوية انا معلوم

Mohammad: ايه.. بس مافي معلوم كلو كلام سعودي

B1: لا.. والله مافي معلوم

Mohammad: ايه.. طيب راديو انتا يسمع

B1: والله ما يسمع

Mohammad: ما يسمع راديو... طيب انتا يجي للسعودية كيف يعني شوف سعودية كيف.. في اختلاف

B1: والله الحمد لله انا شوف.. تلفزيون بنقلاديش.. هذا.. بلد ثاني.. اي بلاد.. احسن من سعودية.. عشان ايش معلوم.. هنا سعودية ناس احترام مية مية.. صغير كبير لا.. مافي اختلاف.. هو شوف؟ السلام عليكم.. وثاني؟ اي وقت.. صلاة.. الله اكبر.. كلو سكر

Mohammad: كلو نفر=

B1: بس.. الله كبير... هو جيب شهادة؟ الله كبير كلو سكر لازم روح صلي.. هذا خير ونعمه

Mohammad: الحمد لله

B1: وثاني.. اي مشكلة انا شوف شرطة السلام عليكم احترام انا كلم ايش في مشكلة تعال هو سوي مساعدة.. اي سعودي انا شوف.. في مشكلة انا.. حمدو لله مافي مشكلة الحين.. في انا كلام السلام عليكم ورحمة الله تعالى في كذا سوي مساعدة.. انا شوف واحد يوم.. انا هنا مشكلة.. في واحد سعودي يجي.. هو يجي؟ كلام السلام عليكم ايش في مشكلة يا اخي؟ انا كلم هذا شوية مشكلة هذا سيارة.. هو سوي دف.. هو سوي دف.. هو كبير عمر.. ممكن انا زيادة عمر.. سيم سيم انا بابا

Mohammad: مم

B1: انا كلام يا بابا.. شكرا.. قول لا انتا انا ولد.. كذا هو سوي.. انا شوف كثير مساعدة كثير

Mohammad: الحمد لله

B1: حمدو لله

Mohammad: طيب.. ايش اكل في بنقلاديش

B1: والله انا بنقلاديش؟ في اكل زيادة رز.. سمك.. خضار

Mohammad: يعني كلو يوم كذا.. رز سمك

B1: نعم

Mohammad: مافي غير

B1: ما في غير.. في غير ناس في خبز.. خضرة.. لحم كذا.. في فواكه.. في.. في ناس كلو بنقلاديش.. ناس كثير.. بيت سوى سوى موجود مزرعة.. ليمون.. مانقو.. هذا رز.. في خضرة شوية.. في كثير ناس.. سوي كذا..

Mohammad: طيب سمك من وين يجيب.. في نهر؟ ولا بحر

B1: لا في.. موجود هذا كثير بنقلاديش بحر.. حمدو لله.. في موسم.. ثلاثة اربعة شهر.. يجي موية كثير.. اي مكان في موية موجود حصل سمك

Mohammad: ايوه

B1: في ثاني.. في كلو ناس شوية شوية سوي هو تراب طلع.. ممكن عشرين متر عشرين متر كذا

Mohammad: ايوه

B1: هو تراب طلع.. سوي كذا؟ سوي هنا ركب موية سوي هذا صغير (بيبي) سمك..

Mohammad: ايوه زراعة

B1: زراعة. سوي زراعة.. اشتري..بيع.. حمدو لله هذا فايدة كثير هو بيع.. برضة اكل.. في ناس هدية

Mohammad: ايوه

B1: كذا

Mohammad: فيه كثير كذا؟ كثير نفر سوي كذا

B1: لا هذا شوف مدينة؟ ما يقدر.. عشان مدينة عمارة جنبو جنبو عمارة.. لكن بعد مدينة

Mohammad: في قرية

B1: في قرية.. موجود تحت

Mohammad: طيب انتا في هنا في اكل سمك رز خبز (سيم سيم) بنقلاديش ولا

B1: لا اختلاف انا اكل

Mohammad: ايه

B1: انا يجي هنا الحمد لله انا اول يجي.. ما يقدر اكل كبسة.. عشان كيف ما يقدر.. لكن الحمد لله نفسو اكل سعودية.. مافي مشكلة

Mohammad: يعني في بنقلاديش مافي دجاج؟

B1: بنقلاديش في دجاج انا اكل ممكن في واحد مرة.. كذا

Mohammad: كيف صلح دجاج.. مشوي ولا برياني ولا

B1: لا لا انا دجاج صلح كيف؟.. (سيم سيم) ايش اسمه هذا ايدام سوى سوى

Mohammad: ايه

B1: سوي كذا

Mohammad: طيب.. هنا.. كم ساعة شغل

B1: انا شغل هنا.. رمضان شغل؟ بعد ضهر ساعة خمسة.. سكر؟ بعد فطور تراويح خلاص؟ ساعة ثمنية ونص و ساعة واحد.. واحد ربع.. في زبون انا وقف مافي زبون مدير كلام سكر روح

Appendix A

Mohammad: طيب انتا في.. في نفر ثاني سكن؟

B1: في انا سوى سوى

Mohammad: فيه سكن سوى سوى نفر ثاني

B1: اربعة نفر

Mohammad: ايوه.. طيب هو كلو (سيم سيم) يعني كلو[بنقلا ولا]

B1:   [في ناس]

Mohammad: في هندي

B1: لا لا بنقلا كلو.. نعم

Mohammad: طيب.. انتا مافي يجلس بعدين فكر انا يبغى روح بنقلاديش

B1: والله يا شيخ انا فكر لكن.. ان شاء الله انا روح لكن الحين ما حصل فلوس انا ما يقدر.. لازم فلوس.. تذكرة.. بعدين انا لازم اشري شوية هدية اوما

Mohammad: ايه

B1: هدية اخو.. اختي.. ولد.. كلو ناس في قرية.. في ناس كلام جيب شوية انا واحد كذا انا ركب.. هذا شمسه دواء.. في ناس كلام جيب انا شوية واحد قماش.. في ناس كلام جيب واحد ثوب.. لازم شوية شوية شيل

Mohammad: يعني انتا مافي مدينة كبير انتا في قرية

B1: لا انا مدينة سوى سوى لكن مافي كبير مدينة

Mohammad: مافي كبير.. يعني (سيم سيم).. اصغر من رياض.. رياض كبير

B1: اكثر اكثر.. انا دكا خلاص انا موجود.. دكا

X: سلام عليكم

B1: وعليكم السلام.. الله اعطيك العافية.. شكرا.. دكا (ير فيفتو).. انا بيت ممكن ساعتين.. ساعة.. مافي ساعتين.. ساعة

Mohammad: طيب انتا في.. في كل يوم روح دوام ولا في

B1: حمدو لله كلو يوم دوام

Mohammad: كل يوم دوام مافي اجازة

B1: مافي اجازة بس عشان يجي زبون لازم.. زبون بعدين زعل

Mohammad: مم

B1: انا اجازة جمعة صبح مافي يشتغل.. بعد رمضان كلو يوم يفتح ساعة تسعة.. ساعة واحد ونص سكر.. وبالليل ساعة اربعة.. بعد العصر ساعة اربعة الى ساعة عشرة ونص سكر.. لكن يوم الجمعة.. انا كفيل مدير قول.. يوم الجمعة صبح مافي سكر... اا.. مافي افتح لا عشان لازم جمعة روح.. صلي مبسوط بعدين تعال

Mohammad: طيب انتا كيف روح دوام.. يعني شغل كيف روح شغل.. يمشي ولا سيكل ولا..

B1: لا.. بس انا دوام هنا خمسة عمارة يمشي.. رجل

Mohammad: ايه.. مافي سيكل

B1: لا مافي

Mohammad: مافي يشتري طيب سيكل عشان =

B1: لا مافي=

193

Appendix A

Mohammad: يروح بعيد يروح بعيد يروح كذا

B1: بس انا ابغى ممكن بعيد انا روح ليموزين روح ريالين باص

Mohammad: ايه...

B1: كذا

Mohammad: في الباص

B1: الحمد لله.. الحمد لله رخيص

Mohammad: باص رخيص ريالين مافي كثير.. طيب انتا معلوم هذا (سواين فلو) انفلونزا خنازير

B1: انا اسمع

Mohammad: اسمع...فيه خوف انتا ولا مافي

B1: بس انا خوف ايش معلوم..سبحانه وتعالى هوا جيب؟ انا معلوم هوا مافيه جيب انا معلوم.. وبرضو ايش معلوم انا مسلم.. انا مااكل هذا..

Mohammad: امم.. مااكل هذا

B1: حرام انا مافي اكل..وبرضو تاني ايش معلوم..هذا كلو انا هنا موجود..قليل ..هذا كافر

Mohammad: اممم

B1: قليل.. كافر.. حمدو لله مسلم زيادة ممكن تسعة مية نفر.. خمسة تسعة مسلم.. عشان الحمد لله انا..

Mohammad: الحمد لله.. طيب انتا كيف.. يعني كيف.. يعني كل وقف مرض هذا.. مافي ينتشر..

B1: بس انا وقف هذا كذا.. انا شوف ناس.. هذا شسمه (كرستيان).. يهود.. لازم شوية بعيد.. هو يجي انا شوفو لازم شوية بعيد.. انا مافي معلوم هو فيه ولا مافيه هو مرض.. لا صاحي.. الله سبحانه وتعالى معلوم.. بس.. انا مافي خوف كذا انا بس.. شغل؟ شغل حلال.. ممكن انا روح امشي سيدة الله مافي جيب انا مرض.. بس انا فكر هذا كذا

Mohammad: طيب العيد في بنقلاديش كيف.. كيف عيد

B1: لا الحمد لله هذا عيد بنقلاديش ممكن انا.. واحد يوم عيد.. كلو ناس.. سوى سوى اكل.. انا بيت.. ثاني بيت.. ثاني ناس بيت.. سوي حلوى.. طبخ حلوى.. سوي لحم.. سوي رز.. كثير أكل.. الحمد لله كلو سوى ناس مبسوط.. هذا عيد.. صلي.. كلو ناس روح لازم صلي دعاء سوى سوى.. هذا ناس موت هو سوي دعاء كلو دعاء بعدين يجي بيت انا سلم ماما.. سلم بابا..

Mohammad: بعد صلاة العيد

B1: ايه.. بعد صلاة العيد

Mohammad: ايوه=

B1: انا صلي خلاص؟ يجي.. ماما بيت.. انا سلم ماما.. سلم اخو كبير سلم ولد كلو.. انا سلم حق كبير اخوان.. ماما بابا هو جيب انا فلوس انا مبسوط

Mohammad: ايوه

B1: عشان انا صلي هو؟ هو جيب انا فلوس=

Mohammad: ايه

B1: اي ناس كذا

Mohammad: انت متى اخر مرة روح بنقلاديش

B1: انا روح اخر مرة.. الفين ستة.. واحد وثلاثين ديسمبر

194

Appendix A

Mohammad: كم يجلس شهر

B1: انا يجلس.. قريب ستة شهر

Mohammad: يعني ستة شهر بعدين يرجع

B1: بعدين يرجع

Mohammad: كم مرة روح.. انتا قبل تسعة سنة يجي

B1: مرتين

Mohammad مرتين

B1: نعم

Mohammad: يروح ستة شهر بعدين يرجع

B1: يرجع

Mohammad: طيب.. الحمد لله

B1: حمدو لله

Mohammad: طيب انتا فكر بعد واحد سنة عشرة سنة كذا.. يرجع بنقلاديش.. خلاص ما يرجع سعودية.. ايش سوي هناك

B1: والله انا فكر ممكن والله جيب انا.. بعدين انا سوي شوية واحد محل.. تجارة صغير.. بس انا يجلس اهل سوى سوى

Mohammad: ايه

B1: الحمد لله

Mohammad: الحمد لله.. طيب.. خلاص.. (الانترفيو) خلاص.. شكرا.. بس في انا كلام واحد.. لو انتا كلم كلام نفر ثاني.. كلام انتا

B1: صح

Mohammad: انتا.. طيب لو اثنين.. كيف

B1: والله.. كيف هو مافي حصل الحين

Mohammad: لا لا مافي حصل بس لو في حصل.. ثنين انتا كلام انتا ولا

B1: انا بس.. ممكن انا يجي انا كلم.. ثاني لغة بنقلا انتا يبغى

Mohammad: لا لا عربي

B1: ثاني ناس انا كلم ثاني ناس بس

Mohammad: بس طيب.. لو فيه واحد حرمة انتا كلام انتي ولا انتا

B1: انا كلم.. هو

Mohammad: هو.. طيب خلاص مية مية

B1: يعني ثلاثة نفر

Mohammad: انتا كلام اول يبغى كلام.. يبغى قول شي

B1: بس انا شوية كلم.. عشان ناس معلوم

Mohammad: ايه

B1: حمدو لله انتا يدرس كبير الله جيب انتا خير وبركة.. انا كلام هذا كذا.. سعودية بنقلاديش.. كثير مساعدة

195

Mohammad: ايه

B1: والله الحمد لله في ناس ممكن شوية غلطان.. في ناس كثير غلطان.. الحين هذا سعودية.. حكومة قول.. هذا.. بنقلا ما في نقل كفالة.. سكر

Mohammad: مم

B1: ممكن نقل كفالة يفتح.. لازم كثير ناس احسن.. عشان ايش معلوم؟ في كثير ناس كفيل اول في شغل كويس.. مية مية.. الحين هو ما حصل فلوس.. روح وين روح؟ صح

Mohammad: صح

B1: هو مسكين.. في كثير ناس برضة كفيل قول قول ياله روح انتا نقل كفالة مافي قول نقل كفالة انا ايش سوي حكومة مشكلة لازم امسك سيدة.. هو الحمد لله هذا سعودي كلم كويس.. عشان حصل انتا شغل كويس روح.. انا شغل قليل.. انا ما اقدر الحين كثير عمال.. عمال جيب راتب حقه.. لازم انتا شوف روح.. انتا كفيل كويس راتب كويس دور انتا.. لكن الحين ما في حصل نقل كفالة.. ممكن افتح نقل كفالة.. انا يفتح.. بعدين..ان شاء الله انتا مساعدة كل ناس لازم مساعدة.. ممكن حكومة.. ان شاء الله ان شاء الله يفتح ان شاء الله

Mohammad: ان شاء الله

B1: ان شاء الله.. وثاني ايش معلوم.. الحين انا سعودية.. الحين انا سعودية.. كلو ناس يجي حج عمرة.. انا يسمع خبر.. بنقلاديش.. ما ادري والله ايش هذا.. في كثير عمر زيادة.. يعطيك (ناشر) ناقص.. مايقدر يجي..

Mohammad: ايه

B1: هذا بنقلاديش ناس اول مافي فلوس كثير.. الحين الحمد لله شغل عمارة شغل في فلوس.. في ولد اثنين ولد موجود.. بابا شيبة.. هو ابغى بابا حج سوي حج لكن ممكن هذا عمر كبير ما يقدر يجي

Mohammad: ما يقدر يجي

B1: ممكن يجي الحمد لله كويس

Mohammad: الحمد لله.. ان شاء الله

B1: ان شاء الله لازم يجي..

Mohammad: ان شاء الله..

Interviewee: M1

Participants:
• Interviewer: Mohammad Al-Moaily, age 29, male, lived 26 years in Saudi Arabia, 3 years in the UK. Parents: from Saudi Arabia. 1st language: Arabic.
• Interviewee: M1, age 43, lived 40 years in India and two and a half years in Saudi Arabia. Parents: From: India. 1st language: Malayalam.
Recording:
22 minutes and 27 seconds, in Batha community Centre: Riyadh, Saudi Arabia.
Interview:
The participants did not know each other before the interview. The interviewee took very short turns in the first five minutes then started taking longer turns as he gained more confidence. The interview went on so smoothly.

Mohammad:طيب.. اا انتا من وين عبدالله؟

M1: انا.. كيرالا..

Appendix A

Mohammad: من كيرالا

M1: ماليبوروم

Mohammad: ماليبوروم.. اا طيب ماليبوروم هذا كبير ولا صغير

M1: كبير.. ان شاء الله

Mohammad: كبير=

M1:مدينة كبير

Mohammad: طيب كيف يعني (سيم سيم) رياض ولا صغير

M1: أكثر من رياض ان شاء الله

Mohammad: أكثر من رياض؟ .. كم (ببيوليشن).. يعني كم عدد سكان

M1: كيف

Mohammad:في.. كيرالا

M1: سكن؟

Mohammad: لا.. عدد سكان يعني عدد

''someone intervenes to explain the meaning in Malayalam''

M1: في والله كم نفر انا ما يعرف

Mohammad: انتا مافي معلوم.. كثير يعني

M1: كثير.. أكثر من سعودية

Mohammad: أكثر ايه.. في كيرالا؟ أكثر من سعودية؟ كثير هذا

M1: أكثر من سعودية يعني... ثلاثه مليار

Mohammad: ممكن ايه.. كثير.. طيب.. انتا في الهند ايش وضيفة

M1: شغل ولا كيف

Mohammad: شغل ايه

M1: سواق سيارة

Mohammad: سواق.. طيب سيارة كبير ولا صغير

M1: سيارة كفر ثلاثة

Mohammad: ثلاثة كفر.. مافي موجود في السعودية هذا

M1: في موجود.. في مكة موجود هنا اول في موجود مكة=

Mohammad: طيب هو سيارة ولا قدام دباب بعدين ركب

M1: ركب (سيم سيم) سيارة

Mohammad: ايوه

M1: ثلاثة كفر في.. ركب اربعة نفر.. ثلاثة نفر.. ركب

Mohammad: طيب فيه سيارات كثير في الهند ولا

M1: فيه كثير ان شاء الله

Mohammad: فيه كثير زحمة

M1: فيه زحمة موجود

197

Mohammad: طيب انتا متى يتعلم يسوق سيارة.. متى كم عمر انتا يعني يتعلم

M1: الحين اربعين.. ممكن خمسة ثلاثين سنة كذا.. يتعلم سيارة

Mohammad: يعني انتا عمرك عشرة سنوات انتا معلوم يسوق هذا سيارة

M1: لا ممكن خمسة عشرين سنة.. عشرين سنة فوق

Mohammad: ايوة عشرين سنة انتا معلوم.. ايه.. اا مافي وظيفة ثاني بس هذا

M1: شغل ثاني والله عايدي

Mohammad: طيب فيه فلوس كويس ولا شوي

M1: شوي زين مافي فلوس كثير

Mohammad: مافيه زياده

M1: مافيه زيادة

Mohammad: طيب في=

M1: في شغل.. في يوم في اكل مافي @

Mohammad: طيب فيه بنزين غالي؟

M1: غالي (سيم سيم)

Mohammad: كم ريال واحد لتر

M1: ريال ممكن.. ثلاثة.. ممكن خمسة ثلاثين ريال

Mohammad: خمسة ثلاثين ريال

M1: خمسة ثلاثين ريال.. ثلاثين ريال. لا في ثلاثة ريال.. اربعة ريال.. اربعة ريال اربعة نص كذا

Mohammad: يعني انتا مافي فايدة

M1: مافي فايدة

Mohammad: سواق مافي فايدة يعني طيب ياخذ هذا يشرب بنزين كثير يعني سيارة هذا ثلاثة كفر

M1: سيارة ثلاثة كفر يعني هذا ثلاثة.. ثلاثة ثلاثين.. لتر ثلاثة ثلاثين كيلو

Mohammad: مم.. واحد لتر=

M1: في يوم ممكن بنزين.. بنزين ثاني كذا خلاص ممكن ثلاثمية ميتين ثلاثمية زيادة باقي ان شاء الله

Mohammad: طيب انتا يروح بعيد؟ ولا.. يعني=

M1: انا يمشي بعيد ممكن خمسين كيلو ستين كيلو فوق مية كيلو برضه يمشي كذا

Mohammad: طيب انتا يسوق سيارة ثاني ولا بس هذا ثلاثة كفر

M1: بس هذا ثلاثة كفر

Mohammad: طيب هنا يسوق سيارة ولا

M1: لا هنا مافي يسوق.. هنا بس في دباب

Mohammad: دباب.. موجود دباب

M1: موجود ان شاء الله

Mohammad: مافي موجود سيكل

M1: لا سيكل مافي

Mohammad: مافي موجود.. طيب بعدين ممكن يشتري سيارة؟

198

Appendix A

M1: انا مافكر الحين

Mohammad: ما فكر الحين

M1: سواق ما في معلوم مية لكن شوية معلوم لكن سواق دباب يختلف حق يختلف ثاني في يختلف

Mohammad: طيب في ال.. في الهند سيارة وين دركسون هنا ولا هنا

M1: في يسار.. يمين..

Mohammad: يمين.. يعني سيم سيم (انجلاند)

M1: سيم سيم (انجلاند)

Mohammad: ايه.. يعني فيه هنا سواقه مختلف كثير؟

M1: ايه كثير زيادة .. لكن من هنا يمين يسار في فرق

Mohammad: انا سيم سيم فرق انا يروح بريطانيا (انجلاند) هذا شوية صعب.. بس معلوم.. الحين انا معلوم

M1: @ معلوم زيادة

Mohammad: طيب.. انتا في الهند,, ايش يدرس.. يدرس جامعة ولا

M1: يدرس في كلية

Mohammad: كلية.. ايش فرق يعني كلية.. كم سنة

M1: فيه يعني سبعة سنة

Mohammad: انتا يدرس سبعة سنة في كلية

M1: سبعة سنة

Mohammad: كلو كلية سبعة سنة

M1: كلو كلية.. مافي كلية "interviewee chats with someone in Malayalam" متوسط مافي كلية متوسط

Mohammad: يعني انتا خلص متوسط

M1: خلص متوسط

Mohammad: مافي روح ثانوي؟ (سكندري)؟

M1: لا ثاني مافي

Mohammad: (سكندري) مافي روح.. طيب انتا يدرس عربي؟ .. هناك في=

M1: يدرس عربي هناك لكن في يروح (سكول) روح عشرة.. لكن مدرسة روح ثمانية.. عربية.. لا عندنا في مدرسة.. مدرسة العربية.. ثمانية سنة مدرسة ثاني لكن مجموع.. اقل كلو اقل يعني..

Mohammad: متى يدرس.. يعني مدرسة في العصر؟ ولا=

M1: مدرسة صبح.. ساعة تسعة ونص الى عشرة..

Mohammad: (سيم سيم.. سيم سيم ببلك اديوكيشن)

M1: ايوه.. ساعة سبعة نص الى عشرة.. فيه ساعة اكل الى عشرة ونص ثاني كذا

Mohammad: ايه ثاني هذا يتعلم عربي يتعلم

M1: ثاني هذا يتعلم عربي فيه قران موجود.. في قران

Mohammad: موجود.. فقه موجود.. كيف يسوي صلاة كيف يسوي زكاة..

M1: موجود ان شاء الله

Mohammad: طيب انتا كلام قبل كم سنة؟ ثنين ونص سنة

199

M1: ثنين ونص سنة

Mohammad: ثنين ونص سنة.. يجي سعودية ايه. طيب كم عمر انته؟

M1: انا واحد اربعين

Mohammad: واحد واربعين

M1: لا.. ثلاثة اربعين

Mohammad: ثلاثة اربعين.. يعني يجي عمرك واحد اربعين تقريبا.. مافي اخر؟ مافي تاخير كثير.. أكثر نفر يجي جديد عمره عشرين سنة كذا

M1: كذا زيادة نفر ثاني لكن انا.. فلوس مافي في مشاكل.. هنا..

Mohammad: طيب انتا يدفع فلوس كثير قبل ما يجي

M1: ممكن.. زيادة ممكن مية..فوق مية عشرين مية ثلاثين الاف روبية

Mohammad: يعني كم ريال

M1:يمكن حدعشر الف ريال.. ثنعشر الف ريال كذا

Mohammad: كيف انتا حصل حدعشر الف بيع شي..

M1: سيارة بيع.. حرمة كلم بيع ذهب بيع

Mohammad: ااااه

M1:كذا @

Mohammad: طيب فيه حصل فايدة ولا ما=

M1: الحمد لله فيه كويس كفيل كويس كله مية مية

Mohammad: الحمد لله

M1: كفيل مية مية تمام

Mohammad: طيب انتا فيه نفر قريب هنا؟

M1: هنا نسيب موجود انا

Mohammad: نسيب يعني زوج اخت انتا

M1: زوج اخت

Mohammad: اا اخت موجود ولا مافي موجود

M1: موجود في صناعية

Mohammad: هو موجود نسيب انتا مع اختك

M1: كيف

Mohammad: يعني نسيب انتا موجود زوجة موجود هنا ولا

M1:لا زوجة مافي موجود

Mohammad: مافي موجود

M1: قليل موجود قليل نفر موجود..

Mohammad: كثير نفر معلوم

M1: كثير نفر

Mohammad: طيب سكن لحال ولا سكن مع..

M1: كلو لحال.. انا في.. انا نسيب ثاني بس

Mohammad: انتا سكن وين لحال؟

M1: صناعية انا نسيب انا سكن سوى سوى حق دكان فوق

Mohammad: انتا فوق يعني انتا لحالك مافي نفر ثاني

M1: لا ما في ثاني

Mohammad: طيب لو حصل نفر ثاني يمكن اا كويس لو انتا حصل نفر ثاني فيه كلام فيه قرقر سوى سوى

M1: نفر ثاني اول جاي اول فيه ممكن ثلاثة اربعة نفر موجود ثلاثة نفر موجود ادفع ايجار اكل كلام كثير انتا يبغى اي واحد شيل ما مشكلة لكن نفر كويس يبغى حساب كويس (سيم) فلوس (سيم سيم) لكن في موجود اول فيه ثنين نفر ثلاثة نفر موجود اول كلو ما يدفع فلوس اول مافيه روح سجل مكتب ولا ايش هو موجود ممكن.. واحد سافر.. بعدين هو ما يجي.. نفر كويس يا سلام يبغى.

Mohammad: ايه طيب انتا.. فيه مو.. زواج انتا زواج؟

M1: ايوه زواج

Mohammad: كم (بيبي)

M1:اربعة

Mohammad: اربعة ايش اسم (بيبي)

M1: واحد كبير صالح.. ثاني سحر..

Mohammad: سحر؟

M1: سحر ايه @ هذا اسم لكن معنى ما يعرف

Mohammad: ايه

M1: ثاني.. اخت. بنت هذا هذا سلوى ثاني واحد صغير ان شاء الله..

Mohammad: ما شاء الله. كم عمر صغير اخر واحد؟

M1: اخر واحد الحين ثنين ونص.. انا يجي من.. ثلاث سنة ممكن زيادة

Mohammad: ايه.. ما شاء الله.. طيب ايش وظيفة هنا في السعودية ايش شغل

M1: هنا قطع غيار سيارات ورشة

Mohammad: ايه.. انتا معلوم عربي كويس.. اثنين سنة حصل=

M1: فيه ثاني. اول انا شغل (انديا) في ثاني زين موجود في بومبي شغل.. انا اول موجود واحد جدة ملباري محمد ملباري سعودي موجود (سوى سوى) انديا.. هو (سوى سوى) شغل انا اربعة سنة..يعني في مدرسة وقف.. اترك شغل انا سعودية في موجود مزرعة هندية.. فيه روح انا امي.. امي اختي. (سوى سوى) مزرعة موجود.. لكن كفيل @

Mohammad: ايوه يعني انتا قبل ما يجي في معلوم عربي

M1:.. ثاني برضو انا في موجود مدرسة عربي كويس ان شاء الله اول كلمتو

Mohammad: هذا نفر.. نفر قبل انته في اربعه سنة مافي عربي كويس يعني ما يفهم كويس

M1: ايوه.. بعدين فيه روح مدرسة لكن زيادة في تعليم.. في انا قول لازم تعليم لكن فلوس مافي بابا يبغى شغل.. انا (سوى سوى) واحد يوم ثلاثة اربعة روبية. في اكل شراب ملبس @ مافي تعليم كثير

Mohammad: طيب يعني في السعودية معلوم عربي

M1: شوية شوية معلوم.. لكن الحين زيادة

Mohammad: لا لا.. قبل سنتين ونص انتا يجي سعودية جديد.. اول.. معلوم عربي؟

M1: شوية شوية معلوم.. لكن مدرسة=

Mohammad: يتعلم عربي في مدرسة.

M1: في مدرسة لكن في.. مسجد في كذا

Mohammad:ماشاء الله.. والله كويس انتا ما شاء الله معلوم عربي كويس.. اا طيب انتا في روح مدينة ثاني في السعودية ولا بس شغل رياض بس

M1: رياض بس

Mohammad: بس شغل رياض. مافي شغل برى في دبي

M1: لا لا مافي محل

Mohammad:بس الهند سعودية

M1: (انديا) سعودية بس

Mohammad: مافي سفر دولة ثاني

M1: مافي سفر.. واحد عمرة..في مكة مدينة ثلاث مرة روح

Mohammad: بس مافي طول.. عمرة واحد يوم ثنين يوم بعدين يرجع

M1: كذا بس يجي

Mohammad: ايه طيب كم سوي عمرة انت في ال..

M1: ممكن ثلاثة مرة انا سوي عمرة.. واحد مرة سوي حج

Mohammad: حج سوي؟

M1: ايوه سوي

Mohammad: ما شاء الله

M1: يجي قبل سنتين سوي عمرة بعدين سوي حج.. الحمد لله كفيل مية مية.. لكن ما يسوي اي مشاكل انا

Mohammad: ايوه كويس.. طيب انتا ثنين ونص روح (انديا) ولا مافي روح

M1: واحد مرة روح واحد سنة في روح

Mohammad: كم يجلس هناك

M1: ثلاثة شهر يجي

Mohammad: طيب متى يروح مرة ثاني؟

M1: مرة ثاني ان شاء الله.. كفيل كلمتو ممكن ريح لكن انا لسة مافي ريح انا شغل بيت ثلاثين خمسين انا كلو شي انا لسة ما خلاص شغل انا لسى كفيل كلام انتا سوي شغل بيت بعدين هو شوية شوية رجع الحين انتا خلي شغل بيت.. بعدين شوية شوية جيب فلوس

Mohammad: طيب انتا كلم اهل في=

M1: هاا؟

Mohammad: كيف كلم اهل في الهند.. كلم جوال؟

M1: ايه جوال

Mohammad: كلم اهل.. كلم (بيبي) كلم=

Appendix A

M1: كلو جوال بس

Mohammad: طيب بابا ماما موجود ولا

M1: موجود ان شاء الله الحمد لله في موجود

Mohammad: كلم انتا

M1: كلم كلو نفر

Mohammad: كم مرة في الاسبوع

M1: يمكن.. يومين @

Mohammad: يومين.. مرتين في الاسبوع

M1: اسبوع مرتين كذا كذا مافي كثير كلام.. ممكن ثنين دقيقة ثلاثة دقيقة.. لكن في خمسين ريال اسبوع تلفون

Mohammad: انتا شغل ورشة؟

M1: قطع غيار سيارات

Mohammad: قطع غيار.. طيب فيه فيه حصل فايدة عربي؟ في شغل يعني في شغل انتا في نفر عربي كثير يجي انتا كلم

M1: كثير يجي هذا

Mohammad: انتا يتلعم عربي

M1: كلو عربي.. بس ما يجي نفر ثاني.. عربي؟ اردو

Mohammad: عربي اردو

M1: ايه

Mohammad: طيب انتا.. ايش كلم لغة ثاني غير عربي غير؟

M1: عربي؟ اردو؟ تاميل.. ثلاثة لغات

Mohammad: تاميل هذا وين كلام تاميل

M1: ثاني كيرالا مدراس.. اردو فيه يمكن موجود اردو لكن بومبي برضة شغل شوية كذا.. لكن في صحيح كلمتك العربي نفر الملباري.. اول في (انديا).. بعدين يجي من هنا كلم واحد اعشرين سنة يجي من جدة.. بعدين بعدين هو رجع ممكن حق ال..مدراسي في كيرالا هو في كبير هو في موت ممكن الحين حرمة هو فيه ثلاث حرمة هو في كيرالا.. حرمة ما يجي مشغول.. كذا نفر.

Mohammad: ايه طيب انتا كلم بابا ماما.. كلم ماليياري؟ ولا كلم معهم اردو؟

M1: ماليياري

Mohammad: بس.. بابا وماما معلوم اردو؟

M1: لا ما يعرف.. اخت معلوم

Mohammad: بابا وماما ايش لغة يتكلم؟

M1: ماليالم

Mohammad: بس

M1: بس ما يعرف

Mohammad: مافي انجليزي ما في

M1: لا لا

203

Mohammad: عربي

M1: لا.. ام مافي يتعلم مرة.. امو ممكن قران برضة ما يعرف يقرى قران.. في معلوم كذا بس لكن قران برضه ما يعرف.. بابا الحمد لله قران معلوم كلو شي معلوم.. (امو) اول مافي مدرسة ولا اي شي.. مرة ما في يعرف ومافي كتابة برضة ما يعرف (امو)

Mohammad: طيب ماليالم هذا فيه اختلاف.. ماليالم؟ فيه شوية اختلاف مثلا.. يعني هو (سيم سيم) ماليالم

M1: هاه

Mohammad: بس فيه اختلاف شرق مثلا عن غرب.. يعني (سيم سيم) سعوديه.. فيه اختلاف سعودي مصري مافيه يتكلم عربي (سيم سيم) سعودي.. سعودي مافي يتكلم (سيم سيم) سوري؟ فيه شويه اختلاف.. ماليالم فيه (سيم سيم) اختلاف؟

M1: ماليالم.. تاميل.. ترناراقا.. كيذا يختلف شوي شوي (سيم سيم) عربي

Mohammad: ايه.. بس انتا يفهم؟

M1: فهم.. زيادة كذا.. لكن.. تاميل.. تاميل في زيادة فهم.. تاميل شويه فرق اقل ترناراقا..

Mohammad: مم

M1: هذا فيه.. فيه ثاني لغة. شوية شوية.. فرق زيادة

Mohammad: طيب في مدرسة.. لو لو مدرس حصل مثلا هذا.. يتكلم تاميل.. هذا نفر يتكلم.. مالياري.. هذا نفر يتكلم اردو.. كيف هو يتكلم؟

M1: هو مدرسة عندنا.. في مالباري بس

Mohammad: ايه

M1: مالياري عندنا اقل. لكن مدرسة العربية.. حق الابتدائي او (البريمري سكول) فيه روح يتعلم هنا؟ اي واحد هنا حق مليبوروم.. الى جامعة (اربك كولج) موجود عندنا.. جامعة اصول الدين موجود.. لكن فيه روح يتعلم كويس امشي هذا.. حق (سيم سيم) المدينة جامعة موجود

Mohammad: ايوه

M1: موجود ثاني (اسلامك سكول).. انا في درس حق (اسلامك سكول) كيرالا.. الحمد لله @

Mohammad: طيب انتا قبل كم سنة يتعلم عربي يعني انتا اول ما يبدا يتعلم عربي كم عمر انتا؟

M1: عمر ممكن انا... زيادة يتعلم كلام هذا من هنا

Mohammad: من هنا

M1: من هنا لكن عندنا ما يجي هندي ولا كيرالا ثاني ما يجي.. كلو جاي عربية بس..

Mohammad: ايه

M1: كذا زيادة معلوم لكن في هنا اول برضه في اربي سجل هنا معلوم.. هندي ما يعرف.. عربي كتاب انا معلوم انجلش كتاب انا ما يعرف

Mohammad: مافي معلوم ما يعرف

M1: ما يعرف.. مرة.. حق انا درس.. حق العربي

Mohammad: ايوه

M1: لكن زيادة تعلم؟

Mohammad: هنا

Appendix A

M1: هنا تعلم لكن زياده تعلم فلوس مافي @

Mohammad: ايه

M1: كذا بس.. في روح انا برضه

Mohammad: يعني يعني هنا في فايدة

M1: في هنا في فايدة.. امي وبابا في سوى سوى في حج سنة روح كذا

Mohammad: طيب.. انتا فيه حصل شي سيم سيم ماليباري عربي

M1: كيف

Mohammad: يعني كلمات موجوده في ماليباري موجوده في العربي.. يعني مثلا بابا ماما (ابو) (امو) فيه كلمات ثاني؟

M1: لا لا لا بابا (اموه).. فيه... امه بابا..

Mohammad: يعني مافيه حصل شي (سيم سيم).. يعني مثلا كلمه سيم سيم موجود ماليباري موجود عربي

M1: مافيه

Mohammad: مافيه كثير

M1: ما فهمت انا هذا كلام

Mohammad: ما فهمت السؤال

M1: ما فهمت

Mohammad: يعني مثلا كلمة موجوده.. في ماليباري موجوده في عربي (سيم سيم) كلمة

M1: اه (سيم سيم)

Mohammad: مرادف يعني

M1: مرادف..

Mohammad: مافي معلوم

M1: مافهم

Mohammad: خلاص مافي مشكلة

M1: @

Mohammad: طيب.. انتا شوف تلفزيون؟

M1: الحين.. قبل رمضان انا شوف اخبار شوف ان شاء الله.. اخبار حق.. اربي مافي شوف.. اربي قليل.. لكن

Mohammad: تلفزيون عربي قليل شوف

M1: في موجود لكن.. مافهم مزبوط مية مية.. كذا ما في شوف.. لكن تلفزيون حق المليباري حق الاخبار في شوف في ثاني يوم الاثنين انا في روح جاليات البطحا.. بطحا جاليات حلو.. "voice not clear".. في ثاني يوم الجمعة.. من هنا يجي هنا.. فاضي يوم انا يجي كذا يمشي بس ما في طريقة..

Mohammad: مافي شوف تلفزيون

M1: اخبار شوف في الليل بس بعد صلاة العشاء نص ساعة بعد نص ساعة في سي دي هذا في شوف بس.. مافي...

Mohammad: طيب (راديو)

M1: (راديو) ماعند انا

Mohammad: ما.. ما.. طيب انتا يجي سعودية.. كيف حصل سعودية

205

M1: كلو ما يعرف زين اول جاي من.. كلو..@

Mohammad: كلو شي جديد

M1: كلو شي جديد..

Mohammad: طيب=

M1: الرجال هذا مرة في (انديا)؟ اغراض كلو سيارة صغير كلوانا ناس صغير انا فكر في انديا كلو ناس صغير.. هنا كلو كبير نفر كذا

Mohammad: ايه... طيب.. ايش اكل موجود في الهند؟

M1: رز.. رز موجود.. في ثاني في مزرعة انا موجود في اي شي موجود في.... في.. موز.. في.. في ثاني اكل خضار.. في موجود رز.. على طول اكل رز.. رز هذا اكل موجود مزرعة انا

Mohammad: طيب

M1: اكل ايجار.. في سوي رز اكل موجود الحمد لله.. في ثاني سوي خبز بر.. في ثاني حق الدقيق حق الرز برضة.. سوي بيض.. مرة اكل.. مرة واحد ممكن روح مطعم.. قليل بس.. مافي.. ثاني كلو في البيت

Mohammad: طيب لو واحد يبغى يروح سوي سياحة في كيرالا

M1: كيف

Mohammad: يعني واحد يروح؟ يروح كيرالا يبغى يشوف ايش كيرالا كذا.. وين يروح؟

M1: روح مكان.. في مكان كثير الحمد لله انا ما يروح بعيد.. ممكن ميتين ثلاث مية متر مافي بعيد

Mohammad: لا لا يعني مثلا في تاج محل.. وين؟ ممكن في بومبي

M1: تاج محل في دلهي

Mohammad: روح انتا تاج محل قبل؟

M1: لا مافي روح

Mohammad: مافي روح.. طيب ايش في غير مكان في..

M1: تاج محل في بومبي.. بومبي تاج محل في بومبي انا في روح.. انا في شغل بومبي اول

Mohammad: طيب ايش فيه غير.. ايش فيه غير في الهند غير تاج محل

M1: ثاني كثير انا ما يعرف في حق الكيرالا في موجود كوشنسان.. مكان حق كوشنسان موجود.. في ثاني كذا حاجة بعد في مكان في كيرالا.. اخر كيرالا موجود في بارد في.. مكان بارد ثاني.. كلو مية مية مزبوط

Mohammad: طيب.. انتا شغل فيه.. فيه اجازة يوم جمعة

M1: يوم الجمعة في بعد العصر

Mohammad: بعد العصر شغل

M1: بعد العصرالى صلاة المغرب انا سكر بعد ممكن ساعة سبعة

Mohammad: يعني كل يوم في شغل

M1: كل يوم شغل

Mohammad: مافي حصل اجازة.

M1: لا

Mohammad: ايه طيب انتا كيف روح دوام.. كيف روح ورشة

M1: كيف

206

Mohammad: يعني روح ورشة يمشي

M1: انا دكان

Mohammad: اوو صح انتا سكن=

M1: دكان.. تحت دكان فوق انا

Mohammad: ايوه.. طيب مافي روح ثاني.. يبغى يجي هنا بطحاء مثلا

M1: دباب

Mohammad: دباب

M1: دباب موجود

Mohammad: ايه.. طيب كم ساعة.. يعني نص ساعة من الصناعية الى هنا.. كم.. كم يعني

M1: انا هنا.. في هذا.. لا ما نص ساعة.. اقل عشرة دقيقة ممكن اربعة كذا ممكن

Mohammad: مافي كثير

M1: لا لا

Mohammad: طيب انتا معلوم (سواين فلو).. هذا انفلونزا خنازير.. اتش ون ان ون..

M1: في معلوم زيادة

Mohammad: في خوف انتا؟

M1: لازم خوف..

Mohammad: لازم خوف

M1: لازم خوف

Mohammad: خوف كثير؟

M1: خوف لازم كيف.. ممكن.. موت مافي خوف لكن موت انا في بعدين بعد الموت انا مافي سوي اي شي.. هذا انا خوف بس.. لكن انا

Mohammad: ايه

M1: حق موت ممكن صبح حق بعد الموت ممكن (زيرو) مافيه اي شي.. فكر زين كويس لكن الله أعلم كذا بس.. لكن موت كويس سوي بس

Mohammad: ايوه.. طيب انتا يعني بعدين يجلس كم سنة عشرة سنة خمسطعش سنة بعدين روح هند ولا كم

M1: الله اعلم

Mohammad: مافي معلوم

M1: الله أعلم هذا انا مافي معلوم زياده كلو @

Mohammad: ايوه مافي معلوم.. طيب روح انتا خلص روح ايش سوي

M1: برضه ما في

Mohammad: مافي فكر انتا سوي (بلان)

M1: مافكر يسوي اي شي

Mohammad: ما في فكر

M1: هذا كلو الله معلوم

Mohammad: طيب رمضان كيف فرق رمضان في الهند عن السعودية

Appendix A

M1: رمضان.. هنا في.. انا في شوية شغل رمضان اول هنا.. في.. عندنا مافي مشاكل اي شي لكن هنا في.. شغل صباح

Mohammad: مم

M1: شوية تعب.. نوم شوية تعب عندنا في رمضان.. نوم زي الحين.. بالليل في بعد صلاة التراويح في شغل.. بعدين صلاة العصر.. واحد ثنين ساعة شغل.. ثاني برضة قوم ساعة ثمانية=

Mohammad: يعني تعب كثير

M1: كذا شوية تعب لكن رمضان الحمد لله كويس مبسوط من هنا

Mohammad: الحمد لله

M1: هنا في دعوة كلو شي.. احسن من دعوة هنا مبسوط ان شاء الله

Mohammad: طيب هنا فيه حر كثير يعني فرق عن..ال=

M1: هنا حر زيادة

Mohammad: حر زيادة.. هنديا مافي حر كثير

M1: هنا مافيه.. فيه هنا.. في حار موجود.. لكن في حار برضه ريح واحد ساعة تحت "voice not clear".. احسن من مكيفات @

Mohammad: ايه.. الحمد لله طيب خلاص.. شكرا

Interviewee: P1

Participants:
• Interviewer:
Mohammad Al-Moaily, age 29, male, lived 26 years in Saudi Arabia, 3 years in the UK. Parents: from Saudi Arabia. 1st language: Arabic.
• Interviewee:
P1, age 47, lived 42 years in Pakistan and five years in Saudi Arabia. Parents: From: Pakistan. 1st language: Punjabi.
Recording: 22 minutes and 42 seconds (in interviewee's shop).
Interview:
The interviewer did not know interviewee before the interview. The interviewee showed confidence in the interviewer before and during the interview. However he took relatively short turns throughout the interview.

Mohammad: طيب ياخي.. انتا من. من وين. من اي مدينة

P1: اسلام اباد

Mohammad: اسلام اباد

P1: ايه

Mohammad: اسلام اباد ايش في لغة.. ايش في لغة في اسلام اباد

P1: ايش في؟

Mohammad: لغة يعني في اسلام اباد لغة لغة ايش

P1: لغة؟

Mohammad: ايه. لغة يعني عربي في=

208

Appendix A

P1: اردو..

Mohammad: اردو=

P1: ايه

Mohammad: ايش ثاني غير اردو

P1: بنجابي

Mohammad: ايش ثاني

P1: بس

Mohammad: يعني اسلام اباد بس اردو ب=

P1: ايه

Mohammad: طيب مكان ثاني في باكستان ايش لغة

P1: في سراكي.. ملتون.. سراكي

Mohammad: ايوه

P1: بعد سراج.. بشتو

Mohammad: ايوه

P1: بعد.. باتاري

Mohammad: ايوه

P1: ايه

Mohammad: سندي فيه

P1: سندي.. بلوشي..

Mohammad: بلوشي.. ايش اكثر.. اكثر شي ايش

P1: اكثر شي اردو

Mohammad: ايه.. وغير.. وبعد اردو ايش..

P1: اا.. بس.. وقف مدينة.. هو كلام بلوشي سندي بشتو

Mohammad: طيب انتا من اسلام اباد.. يعني قبل ما يجي للسعودية قبل=

P1: لا قبل ما يجي

Mohammad: لحظة قبل ما يجي انتا معلوم عربي ولا مافي معلوم

P1: لا قبل مافي معلوم

Mohammad: مرة ما في معلوم

P1: لا مافي

Mohammad: يجي هنا يتعلم

P1: ايه

Mohammad: طيب كيف انتا يتعلم عربي هنا

P1: هنا بس زبون شوف قرقر قرقر بس يجي..

Mohammad: وين شغل

P1: انا خرج

209

Mohammad: خرج ايه شغل.. هنا؟

P1: ايه اجلس هنا

Mohammad: ايه شغل هنا.. طيب انتا في باكستان (سيم سيم) هذا شغل نجار

P1: ايه نجار

Mohammad: نجار شغل (سيم سيم)..

P1: ايه

Mohammad: كم سنة .. متى يتعلم انتا نجار

P1: نجار انا تقريبا.. خمسطعشر سنة ستطعشر سنة انا يمشي شغل نجار.. ايه

Mohammad: يعني انتا عمر انتا خمسطعش سنة

P1: ايه

Mohammad: يعني صغير خمسطعش سنة؟ يبدا يتعلم

P1: اي نعم

Mohammad: طيب.. يدرس.. كم سنة يدرس يعني انتا

P1: باكستان؟

Mohammad: ايه باكستان

P1: الحين تقريبا عمر ثمنية واربعين سنة.. بعد اربعين سنة يجلس باكستان

Mohammad: لا لا انتا يروح مدرسة يروح مدرسة

P1: لا مافي روح مدرسة

Mohammad: مافي روح مدرسة

P1: لا مافي روح مدرسة

Mohammad: معلوم يكتب

P1: اكتب شوية معلوم مافي زيادة

Mohammad: بس مافي روح مدرسة

P1: ايه

Mohammad: طيب مافي يتعلم عربي مافي يتعلم لغة ثاني؟

P1: لا مافي

Mohammad: يعني ايش.. انتا كلام.. اردو؟

P1: باتاري بنجابي

Mohammad: بنجابي

P1: ايه

Mohammad: كلام اردو؟

P1: ايه اردو

Mohammad: عربي؟

P1: شوية عربي

Mohammad: شوي.. انجليزي في معلوم؟

210

Appendix A

P1: لا مافي

Mohammad: لغة ثاني مافي معلوم

P1: لا مافي

Mohammad: يعني ثلاثة لغة

P1: ايوه بس

Mohammad: عربي

P1: بتاري.. اردو

Mohammad: ايش باتاري؟ بتاري (سيم سيم) بنجابي ولا مختلف

P1: لا (سيم سيم) كلام بس شوية اختلاف

Mohammad: ايه.. طيب انتا زواج؟

P1: ايه فيه زواج

Mohammad: ااا.. كم موجود (بيبي)

P1: ثلاثة ولد ثلاث بنت

Mohammad: كم سنة قبل.. قبل كم سنة زواج

P1: زواج "interviewee chats with co-workers in Punjabi" تسعطعش

Mohammad: تسعطعش سنة..

P1: ايه

Mohammad: يعني قبل تسعطعش سنة

P1: ايه

Mohammad: انت الحين كم عمر؟

P1: تقريبا ثمنية واربعين.. سبعة واربعين

Mohammad: طيب مافي متأخر يعني انتا يجي كثير نفر يجي يجي سعودية يجي عشرين سنة عمر يعني لسه صغير انتا كبير بعدين يجي

P1: ايه بعدين بس رزق الله من الله

Mohammad: الحمد لله.. طيب نفر يبغى يجي باكستان يبغى يجي سعودية يجي في.. سهل يجي ولا في مشاكل كثير شغل بعدين يجي

P1: وين؟

Mohammad: يعني انتا.. انتا في باكستان تبغى تجي للسعودية اول مرة

P1: ايه

Mohammad: في شغل كثير روح سفارة روح (فيزا) كذا ولا على طول سرعة=

P1: على طول يجي بس.. رزق الله من الله.. انا يجي هنا على طول هنا.. تقريبا عشرة خمسطعشر يوم كذا انا يجي هنا

Mohammad: مم.. يعني ترتيب ايش.. كيف ترتيبات

P1: انا ولد عمي.. هو قول يبغى كفيل كلم (بي) انا يبغى نفر مية مية

Mohammad: ايه

211

Appendix A

P1: يبغى نجار مية مية

Mohammad: ايوه

P1: هو قول (بي) انتا نفسو بعد مية مية انتا روح سعودية

Mohammad: ايه

P1: انا يجي هنا سعودية

Mohammad: يعني انتا موجود ابن عمك في السعودية

P1: ايه

Mohammad: في نفر ثاني قريب موجود

P1: في اخت ولد موجود.. نسيب موجود

Mohammad: كلو في الخرج ولا في

P1: لا في.. جدة.. مكة.. اخت ولد رياض

Mohammad: يعني انتا مافي.. مافي شوف

P1: انا شوف

Mohammad: شوف

P1: ايه

Mohammad: روح مكة؟

P1: ايه روح مكة انا يجلس ثمنية عشرين رمضان انا روح مكة

Mohammad: ما شاء الله

P1: ايه

Mohammad: روح مكة كثير ولا

P1: لا مافي روح كثير

Mohammad: يعني واحد سنة روح مرة ولا كيف

P1: ايه بس ثلاثة مرة ثنين مرة

Mohammad: رمضان بس

P1: ايه.. رمضان=

Mohammad: حج مافي روح؟

P1: حج روح انا

Mohammad: كم مرة

P1: اثنين مرة روح

Mohammad: ما شاء الله. قبل كم سنة

P1: قبل ثلاثة سنة انا سوي حق انا حج

Mohammad:ايوه

P1: بعد ماما موت؟ انا سوي ماما خاطر حج.. بابا انا موت الحين انا فكر سوي خاطر بابا حج

Mohammad: السنة هذي ولا بعدين

P1: سنة هذي

212

Appendix A

Mohammad: طيب مافي خوف انفونزا خنازير هذي

P1: نعم

Mohammad: مافي خوف من انفلونزا خنازير هذا.. مرض.. مرض هذا يجي في..

P1: ايه ايه مرض

Mohammad: ايه.. مافي خوف انته؟

P1: ايش خوف.. في خوف.. لازم لبس هنا قماش ايه شي

Mohammad: ايه.. يعني لازم لبس قماش بعدين روح مافي خوف

P1: ايه نعم

Mohammad: طيب كويس.. انتا في.. في شغل دولة ثاني غير السعودية

P1: لا مافي

Mohammad: بس باكستان سعودية

P1: بس

Mohammad: على طول يجي

P1: ايه

Mohammad: طيب انتا كيف كلم زوجة.. اولاد هنا في السعودية ولا

P1: لا مافي باكستان

Mohammad: ما.. باكستان

P1: ايوه

Mohammad: ما.. ما في كلم..

P1: لا مافي..

Mohammad: كلم=

P1: فلوس كثير..واجد فلوس.. مافي..

Mohammad: مافي كلم يعني

P1: ايه لا مافي

Mohammad: طيب كيف يعني انتا معلوم اخبار

P1: نعم؟

Mohammad: كيف انتا معلوم اخبار يعني انت يرسل رسالة ولا كيف

P1: انا

Mohammad: ايه

P1: اعرف انا معلوم كلام

Mohammad: ايه.. بس مافي كلم جوال

P1: لا جوال.. على طول كلم.. ممكن. عشرة يوم كذا كلم السلام عليكم. كيف حال. كيف الاهل.. بس

Mohammad: ايه

P1: مافي على طول

Mohammad: طيب انتا كلام انتا مافي بنجابي في لغة ثاني بتري او..

213

Appendix A

P1: في بنجابي في سراكي

Mohammad: لا انتا انتا لغة انته

P1: انا

Mohammad: ايش لغة انته

P1: باكستان؟

Mohammad: ايه.. انتا انتا لغة.. بنجابي صح ولا

P1: بنجاب. باتواري

Mohammad: باتواري ايش فرق عن بنجابي

P1: بنجابي كلام ($) انا قول ($)

Mohammad: بس كلو باقي (سيم سيم)

P1: ايه

Mohammad: ايه.. انا اول مرة اسمع لغة هذا اول مرة يعني.. انا قبل معلوم سندي معلوم بنجابي معلوم

P1: بلوشي

Mohammad: بلوشي كلو معلوم بس هذي بتواري اول مرة

P1: كشميري

Mohammad: مافي كثير نفر يتكلم هذي باتوري

P1: لا شوف اي اسلام اباد

Mohammad: ايه

P1: نفس.. بعد روح جيلم

Mohammad: ايه

P1: هذا كلام كلو سوى سوى

Mohammad: ايه

P1: جيلم.. قدام كلو بنجاب

Mohammad: ايه

P1: بعد مترون سراكي

Mohammad: ايه.. طيب بنجابي في.. يعني في الهند في نفر كلم بنجابي ولا مافي

P1:هند؟

Mohammad: ايه

P1: لا اردو زيادة.. هند زيادة اردو

Mohammad: مافي نفر يتكلم=

P1: لا مافيه.. هند زيادة اردو

Mohammad: كشمير في

P1: كشمير.. هو كلام لحال شوية.. شوية فرق مافي زيادة

Mohammad: مافي فرق كثير

P1: ايه

214

Appendix A

Mohammad: طيب انتا كم.. كم يجلس شهر بعدين يتعلم عربي يبدا يتعلم عربي

P1: هنا؟

Mohammad: ايه

P1: سعودية؟

Mohammad: ايه

P1: تقريبا سبعة.. ثمنية شهر انا يجلس.. بعد شوية كلام يفتح محل يجي زبون قرقر معلوم.. الحين مافي زيادة معلوم عربي.. بس..

Mohammad: يعني سبعة شهر انتا يتعلم سرعة سرعة=

P1: ايه=

Mohammad: بعدين.. وقف خلاص مافي معلوم

P1: ايه

Mohammad: طيب انتا كيف عشان يتذكر كلام كيف.. سجل يعني ولا كيف

P1: انا؟

Mohammad: ايه

P1: لا مافي سجل.. بدون سجل

Mohammad: بدون سجل يعني انتا يتذكر على طول

P1: انا (بي) هذا شغل نجار.. هذا انا معلوم

Mohammad: ايه

P1: الحين شغل سباكة كهرباء انا مافي معلوم

Mohammad: ايوه.. بس معلوم نجار=

P1: شوية معلوم ايوه

Mohammad: طيب.. في بنجابي في كلام (سيم سيم) في العربي عربي

P1: لا مافي.. باكستان مافي عربي

Mohammad: لا لا مافي عربي.. بس في كلمة يمكن واحد كلمة (سيم سيم) موجود في عربي

P1: ايوه؟ كلمة كلو مسلم

Mohammad: ايه

P1: مسلم كلو دنيا كلمة واحد

Mohammad: ايه

P1: قران واحد.. بس اكتب ممكن هندي اكتب لحال.. باكستان.. لحال.. سعودية لحال..

Mohammad: مم.. ايش في كلمة موجود (سيم سيم) بنجابي موجود في عربي

P1: لا كلمه كلو.. لا اله الا الله محمد رسول الله

Mohammad: موجود في بنجابي

P1: ايه

Mohammad: ايش ثاني.. كلو.. سوى سوى كلمة.. لا مثلا يعني بابا في بنجابي في بابا كلام بابا؟

P1: ايه

215

Appendix A

Mohammad: (سيم سيم) كلمة بابا ولا مختلف

P1: لا لا.. كلو (سيم سيم) كلمة.. كلمة.. كلو دنيا مسلم؟

Mohammad: ايه

P1: واحد كلمة

Mohammad: ايه.. لا غير غير كلام مسلم

P1: لا غير مافيه

Mohammad: غير مافيه

P1: لا غير مافيه

Mohammad: طيب انتا في باكستان مافي يتعلم عربي

P1: لا لا

Mohammad: مافي روح مدرسة يتعلم عربي

P1: لا مافي

Mohammad: يتعلم قران يتعلم

P1: ايه قران معلوم

Mohammad: طيب كيف يعلمونك قران هناك

P1: قران روح مدرسة حق.. ماما انا ماما موجود؟

Mohammad: ايوه

P1: هو يتكلم

Mohammad: طيب انتا معلوم يكتب عربي ولا

P1: لا مافي معلوم.. مافي اردو ما في معلوم.. عربي مافي معلوم.. انجليزي مافي معلوم..

Mohammad: ايه.. طيب انتا شوف تلفزيون؟

P1: لا مافيه.. مافي وقت زيادة

Mohammad: طيب مافي شوف مافي يسمع راديو؟ راديو راديو

P1: راديو؟

Mohammad: ايه

P1: لا ما

Mohammad: مافيه؟ بس (تيب).. بس هذا يعني مافي عربي

P1: لا عربي مافيه

Mohammad: ايه مافيه

P1: ايه

Mohammad: انتا يمكن لو يسمع (تيب) عربي ممكن شوي شوي معلوم عربي بعدين مية مية

P1: ايه

Mohammad: طيب.. كيف انتا يعني يجي للسعودية اول مرة رياض ولا يجي بس خرج على طول

P1: لا يجي خرج.. بعد سنة ونص تقريبا يجلس رياض

Mohammad: ايه

216

Appendix A

P1: انا كفيل يشتري ورشة

Mohammad: ايه

P1: هو يقول هذا انتا شغل هنا.. امشي شغل هنا.. انا يجي مافي معلوم عربي.. مافي معلوم وين يشتري اغراض.. كيف.. بعد غالي نفر جيت بعد انا كلام روح باكستان لازم.. هنا مافي فايدة

Mohammad: ايه

P1: هو قول لا انتا روح برى انتا روح برى اي محل انتا روح شغل؟ انتا مزاج

Mohammad: ايه

P1: بعد انا يجي هنا خرج موية فرزان

Mohammad: ايه

P1: ثلاثة شهر شغل

Mohammad: ايه

P1: بعد.. شوية معلوم كيف طريقة انا شغل.. عربي شوية معلوم

Mohammad: ايه

P1: بعد انا كلام.. كفيل ابغى هنا افتح محل

Mohammad: ايه

P1: هو قول مافي مشكلة.. انا افتح محل.. بعد هذا يشتري=

Mohammad: هذا محل.. ولا ثاني

P1: لا ثاني هنا بعد مفرق..

Mohammad: يعني انتا يجلس.. سنة ونص خرج؟

P1: لا؟ انا تقريبا كلو سبعة سنة واحد ونص سنة

Mohammad: واحد ونص رياض

P1: ايه رياض مافي واحد سنة

Mohammad: اي مكان في الرياض

P1: شفا صناعية

Mohammad: في الشفاء

P1: ايه

Mohammad: مافي كثير اجلس كثير

P1: لا ما في اجلس كثير هنا زيادة

Mohammad: ايش احسن الخرج ولا رياض

P1: الخرج احسن

Mohammad: ليش احسن الخرج

P1: احسن انا باكستان مدينة؟

Mohammad: مم

P1: (سيم سيم) مافي بعيد

Mohammad: مم مافي كبير يعني سيارة زحمة كثير

217

P1: ايه نعم

Mohammad: انتا ما.. انتا في مدينة صغير في باكستان

P1: ايه

Mohammad: (سيم سيم) خرج يعني

P1: (سيم سيم) خرج

Mohammad: طيب انتا ايش.. ايش سيارة هنا

P1: هنا مافي سيارة.. سيكل

Mohammad: هنا سيكل.. هناك سيكل ولا سيارة

P1: وين؟

Mohammad: باكستان

P1: باكستان دباب هذا دباب.. بعد سيكل

Mohammad: ايه.. مافي دباب بعدين ورى يعني..

P1: دباب كبير؟

Mohammad: لا مافي دباب كبير هذا انا شوف تلفزيون هذا.. دباب.. ورى في ركب عربية ورى صغير ورى كذا

P1: لا لا مافي

Mohammad: مافيه

P1: مافيه

X: ($)

X: ($)

Mohammad: ($) مافي انتا هناك

P1: لا مافي..

Mohammad:كثير هناك في=

P1: ايه باكستان في كثير

Mohammad: طيب هنا ليش ما يشتري دباب

P1: هنا مشكلة سيكل يشتري بيت روح زبون؟ درج

Mohammad: ايه

P1: مشكلة شيل.. ودي فوق.. برى ركب (علي بابا).. شيل

Mohammad: صح.. يعني هنا بس سيكل

P1: ايه سيكل

Mohammad: بعدين مافي يشتري دباب

P1: لا ما في يشتري

Mohammad: سيارة مافي يشتري

P1: لا

Mohammad: طيب انتا سيكل يمكن ما يقدر يروح بعيد

P1: لا روح بعيد.. قريب.. مافي مشكلة

218

Mohammad: يعني روح بعيد=

P1: سوق في سيارة انا سيكل سوى سوى

Mohammad: روح سيكل؟ مافي تعب

P1: تعب في بعيد.. مشكلة.. كيف هنا ممكن نص ساعة يجلس.. حصل سيارة.. بعد روح سوق

Mohammad: صح

P1: هنا شيل ثلاثة ريال بعد نزل سوق بعد روح

Mohammad: بس سيكل مافي تعب يعني روح..

P1: لا مافي تعب..

Mohammad: طيب.. ايش اكل في باكستان

P1: خبز.. خضرة.. ذرة.. لحم..

Mohammad: مم

P1: دجاج.. هنا.. خبز

Mohammad: هنا انتا ياكل اكل (سيم سيم) باكستان

P1: (سيم سيم)

Mohammad: (سيم سيم).. صلح انتا ولا

P1: لا انا صلح

Mohammad: انتا صلح.. معلوم صلح؟

P1: ايه معلوم

Mohammad: ايش طيب غير خبز.. ذرة

P1: بس خبز.. بيض. لحم. دجاج. خضرة. ذرة

Mohammad: ايش لحم.. لحم ايش غنم ولا

P1: ايه غنم

Mohammad: ما ياكل انتا لحم بقر ولا لحم=

P1: لا مافي.. بقر حصل انا صلح.. باقي.. مافي جمع مافي اكل

Mohammad: مافي اكل.. طيب في باكستان ياكلون جمل ولا ما ياكلون=

P1: لا مافي بقر.. غنم بقر زيادة

Mohammad: بس في جمل في باكستان ولا مافي

P1: في جمل.

Mohammad: كثير ولا

P1: مافي هنا.. مافي بيع

Mohammad: ما يذبح

P1: هنا بيع جمل.. باكستان ما في بيع جمل في شوية مافي زيادة

Mohammad: بس انتا هنا يعني هنا صلح يعني (سيم سيم) اكل هناك

P1: ايه (سيم سيم).. باكستان (سيم سيم) هذا صلح هناك

Mohammad: طيب كيف خبز.. خبز في باكستان (سيم سيم) خبز

P1: لا خبز احسن باكستان

Mohammad: ايه.. لا انا روح بقالة يشتري هذا خبز مدور هذا اربعة ريال

P1: لا مافي نفسه

Mohammad: (سيم سيم) ولا

P1: لا لا انا يشتري.. يشتري بر؟

Mohammad: مم

P1: بعد روح طحانه

Mohammad: ايوه

P1: بر جيب بيت..

Mohammad: ايوه

P1: حرمة سوي بر.. بعد سوي.. على طول خبز..

Mohammad: مم يعني خبز كذا ولا دقيق رهيف ولا كيف

P1: لا لا شوي خفيف

Mohammad: مم خفيف مرة.. مرة خفيف

P1: ايه مرة خفيف

Mohammad: يعني (سيم سيم) (تشاباتي)

P1: ايه باكستان في (تشاباتي)

Mohammad: ايوه.. طيب.. انتا سكن هنا في بيت لحال ولا في مع نفر ثاني

P1: في نفر ثاني

Mohammad: في نفر ثاني

P1: ايه

Mohammad: كلو (سيم سيم) لغة بنجابي ولا..

P1: ايه كلو سوى سوى

Mohammad: كلو سوى سوى لغة

P1: ايه

Mohammad: يعني انتا مافي.. مافي مشكلة كلم انتا هو

P1: لا مافي مشكلة

Mohammad: فهم

P1: فهم ان شاء الله

Mohammad: فهم انت.. طيب انتا كم ساعة يتكلم عربي في اليوم

P1: اي ناس زبون.. كلم.. بعد.. سواق انا قرقر.. لحال انا نفر باكستاني.. انا قرقر باكستاني.. يجي سعودي زبون.. مصري.. يمني

Mohammad: ايه

P1: سوى سوى قرقر

Mohammad: طيب انتا مافي مشكلة يعني كلم يفهم كلو كلام هو ولا شوي..

P1: ايه ان شاء الله

Mohammad: طيب لو في نفر ثنين نفر سعودي ثنين نفر

P1: لا مافي معلوم

Mohammad: مافي معلوم

P1: مافي معلوم

Mohammad: لو هو يتكلم شوية شوية انتا معلوم ولا

P1: ايه بس شوية معلوم مافي زيادة سعودي سوى سوى قرقر

Mohammad: ايه

P1: مافي معلوم

Mohammad: ليش طيب غريبه.. انا انتا يتكلم انتا معلوم كلو

P1: الحين انا انتا كلام شوية شوية

Mohammad: مم

P1: الحين يجي سعودي.. سوى سوى.. كلام مافي معلوم

Mohammad: انتا مافي معلوم

P1: مافي معلوم

Mohammad: طيب انتا كم.. كم يوم واحد اسبوع كم يوم شغل خمسة يوم ولا ستة يوم ولا كلو يوم شغل

P1: لا كلو يوم شغل

Mohammad: مافي اجازة

P1: مافيه

Mohammad: ايه.. كم ساعة شغل في اليوم

P1: شغل انتا في بس سبعة ساعة ثمنية يفتح.. بعد ظهر.. صلي اذان

Mohammad: مم

P1: روح بيت

Mohammad: ايه

P1: بعد يجلس.. عصر صلي؟ يجي صلي مسجد

Mohammad: مم

P1: صلي خلاص افتح محل ساعة عشرة.. ساعة حدعش.. في شغل؟ ممكن ساعة حدعش ساعة ثنعش

Mohammad: ايه

P1: مافي شغل ساعة عشرة سكر محل روح بيت

Mohammad: ايوه طيب يعني انتا مافي تعب كلو يوم روح شغل شغل كل يوم مافي.. يمكن في كلام كفيل يبغى اجازة شوف

P1: (بي) احسن يجي رزق.. فلوس

Mohammad: ايه

P1: احسن فلوس بعد يوم الحمد لله سكر محل.. مافي حصل لازم يجي ممكن مية ريال يمكن يجلس خمسين ريال

Mohammad: مم طيب..

Appendix A

P1: انا بيت اجار مية ريال

Mohammad: مم

P1: واحد شهر

Mohammad: ايه

P1: هذا فاتورة مية ريال مية خمسة مية عشرة.. واحد شهر ميتين ريال.. مصروف

Mohammad: ايه

P1: هذا بيت.. بعد سبعمية اجار جوعان بطاطا شاي ببسي

Mohammad: اكل كلو

P1: ايه

Mohammad: يعني لازم انتا.. لازم شغل عشان فلوس

P1: ايه

Mohammad: طيب انتا كم مرة روح باكستان يعني في

P1: سبعة شهر ثنين مرة روح

Mohammad: كيف؟

P1: سبعة سنة انا هنا موجود؟

Mohammad: ايه

P1: ثنين مرة روح

Mohammad: ثنين مرة.. يعني كل.. اربعة سنة روح.. ثلاثة سنة روح

P1: لا.. اول مرة روح.. ثمنية عشرين شهر بعد روح

Mohammad: مم

P1: ثنين شهر=

Mohammad: سنتين واربعة شهور

P1: ثنين شهر يجلس=

Mohammad: مم

P1: بعد انا يرجع.. بعد.. تقريبا.. ثنين سنة روح بعد باكستان

Mohammad: ايه

P1: ثلاثة شهر عشرة يوم يجلس باكستان.. بعد يرجع

Mohammad: يعني اخر مرة قبل كم.. اخر مرة روح باكستان

P1: في.. قبل رمضان.. تسعة وعشرين رمضان.. انا روح سفر باكستان.. بعد حج؟

Mohammad: ايه

P1: محرم؟ شهر واحد.. شهر واحد.. سبعة طريق انا يجي

Mohammad: ايه.. يعني قبل واحد سنة

P1: ايه

Mohammad: ايه.. طيب انتا معلوم.. هذا مرض انفلونزا؟

P1: نعم

222

Mohammad: انفلونزا هذا معلوم.. مرض انفلونزا معلوم.. نفر هذا يجي ينتشر كثير يروح كثير ناس يعني

مرض.. معلوم انفلونزا خنازير

"Interviewee's co-worker explains the interviewer's question in Punjabi"

P1: اه اه.. باكستان مافيه

Mohammad: باكستان مافيه

P1: لا باكستان مافيه

Mohammad: مم

P1: هند كثير زيادة باكستان نيبال زيادة هند..

Mohammad: مم.. طيب انتا فكر فكر يعني يمكن.. كيف انا وقف يعني هذا مرض مافي ينتشر كثير.. كيف نفر

مافي.. يعني مافي خلاص يروح.. يمكن بعدين مرض هذا يروح كلو مكان باكستان يروح الهند يروح

P1: ما ادري من

Mohammad: ممكن ايه كيف انا وقف مرض هذا مافي

P1: لا مافي معلوم

Mohammad: مافي معلوم

P1: مافي معلوم

Mohammad: طيب انتا يعني عمر انتا كم؟ ثمنية واربعين سنة

P1: تقريبا سبعة.. ثمنية

Mohammad: ثمنية واربعين

P1: ايوه

Mohammad: طيب متى خلاص يعني خلص شغل.. يروح مرة

P1: باكستان؟

Mohammad: ايه

P1: رزق الله من الله.. انا ثنين مرة كلام كفيل انا روح سفر على طول.. هو قول لا مافي اجلس هنا

Mohammad: كفيل كويس هذا

P1: ايه لا برى مافي روح

Mohammad: ايه.. يعني انتا في فايدة يجلس في فايدة

P1: اه في فايدة.. مافي فايدة.. شوية نص نص

Mohammad: انتا قبل ما يجي يمكن انتا فكر في فايدة كثير بعدين يجي مافي حصل فايدة

P1: ايه

Mohammad: طيب.. مافي فكر انتا روح مكان ثاني

P1: لا مافي فكر

Mohammad: دبي مثلا

P1: لا

Mohammad: دبي فلوس كثير ممكن=

P1: لا.. هنا مرتاح زيادة

Appendix A

Mohammad: مم

P1: بعد=

Mohammad: ما فكر روح مكة مثلا

P1: نعم

Mohammad: ما في فكر روح مكة طيب

P1: لا بس سوي عمرة.. سوي عمرة

Mohammad: شغل مكة يمكن فلوس كويس

P1: لا مافي.. هنا راس كويس مية مية

Mohammad: @ ايه

P1: ايوه

Mohammad: يعني خرج كويس

P1:ايه خرج كويس

Mohammad: ايه.. طيب كويس

P1: فيه فايدة.. فيه خسارة.. حمد لله كويس

Mohammad: الحمد لله

P1: ايه

Mohammad: طيب.. خلاص.. انتا في.. في كلام يبغى يقول زيادة ولا خلاص

P1: مافي مشكلة ايش في..

Mohammad: لا لا انت لو يبغى في كلام انتا يعني يبغى كلام مثلا في كفيل

P1: ايه

Mohammad: كيف انتا مع كفيل كويس مرتاح

P1: لا كفيل الحمد لله كويس

Mohammad: كفيل في الخرج ولا في الرياض

P1: لا خرج.. داوود

Mohammad: ايه.. موجود في الخرج

P1: ايه

Mohammad: هو كويس مية مية

P1: لا الحمد لله كويس

Mohammad: الحمد لله.. طيب (بيبي) انتا كبير ولا صغير

P1: ثلاثة ولد.. كبير

Mohammad: كم سنة

P1: ولد كبير ثمنطعشر سنة

Mohammad: مم

P1: بعد تقريبا اثنعش.. بعد.. تسعة سنة تقريبا.. بعد (بيبي) ستة سنة

Mohammad: مم

224

Appendix A

P1: ستة سبعة سنة

Mohammad: يروح مدرسة كلو

P1: كلو حمد لله

Mohammad: كلو الحمد لله يروح مدرسة.. يعني هو مافي.. مافي تاثير انتا روح بعيد عنو مافي تاثير هو

P1: لا مافي

Mohammad: ماشي كويس

P1: امشي الله

Mohammad: الحمد لله

P1: انا قول انا مافي روح مدرسة.. انتا لازم روح مدرسة

Mohammad: ايوه

P1: فلوس يدفع انا.. لازم روح مدرسة..لازم

Mohammad: باكستان يدفع فلوس انته

P1: ايوه انا ادفع فلوس

Mohammad: ايه

P1: بابا في موجود هو شغل نجار ورشة كبير عامل موجود

Mohammad: ايه كويس,, يعني الحمد لله يعني هو ما=

P1: انتا ايش يبغى موية عصير

Mohammad: لا لا خلاص انا خلاص يمشي

P1: ايه خلاص وقف هذا

Appendix B. Interview Schedule

1. Interviewee demography:

Where are you from?

Which city in …?

Your job in Saudi Arabia…

Have you received training in your home country or in Saudi Arabia for your current

job?

Did you go to school? If yes, till which level?

Your previous job in …?

Marital status, if so how many kids?

Age?

Do you have siblings/ relatives living in Saudi Arabia? Back home? Do you meet them?

For how many years have you been living in KSA, the Gulf?

Have you been working/ living in any other Arabic speaking country before you come to

Saudi Arabia?

How do you contact your family? By phone? Post? Internet? Other? How often?

2. Linguistic background:

What is your first language?

Do you speak a variety of this language?

How often do you speak UPA? Do you speak it at home? Do you speak it in your home

country? Do you speak it in your workplace?

Was it difficult for you to learn UPA? Do you think you need to learn it more?

Did you find linguistic similarities between your mother language and Arabic?

Did you have Arabic courses before coming to Saudi Arabia?

If yes, were the courses in Standard or in non Standard Arabic? Were

they helpful?

If no, are there any? Why did not you consider taking one?

Do you watch Arabic TV channels, if so how many hours a day/ a week? What kinds of

programs do you watch? Do you watch here TV channels in your first language? Any

other language?

Do you listen to radio? If yes, in which language?

3. Other:

How was your experience of working abroad?

How did you find Saudi Arabia, things you like and thing you don't like?

Do you have traditional foods in your home country? How do you prepare them?

Do you eat here the same kinds of food you eat in your home country?

Do you live alone? With friends? Family?

Do you feel homesick?

What do you do in weekends? In your spare time?

What do you use for daily commutation? Why not use a (car, bus, bicycle, etc.)?

Have you ever been in a situation where you were about to die?

Are you worried about epidemics (e.g. swine flu, bird flue, HIV, etc.)? How do you

think we can stop them?

What are your plans for the future?

4. Direct elicitation (PowerPoint Presentation)

9

10

11

12

www.lucylearns.com

13

www.lucylearns.com

14

15

16

22
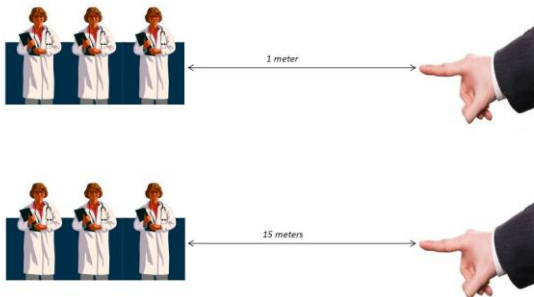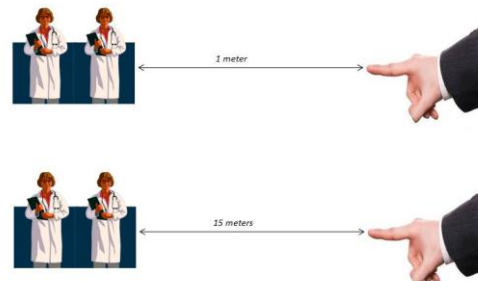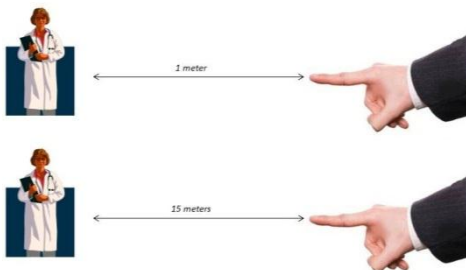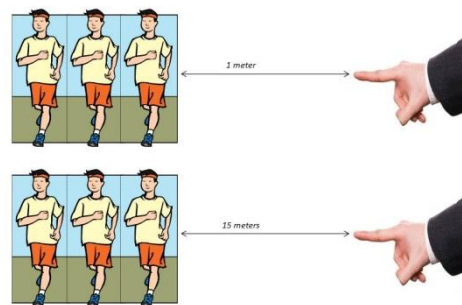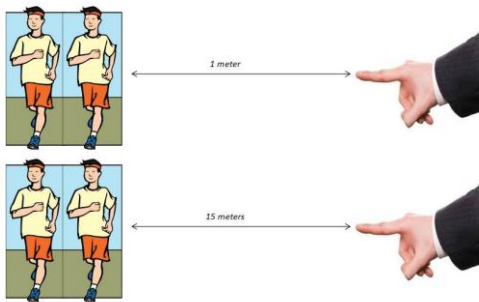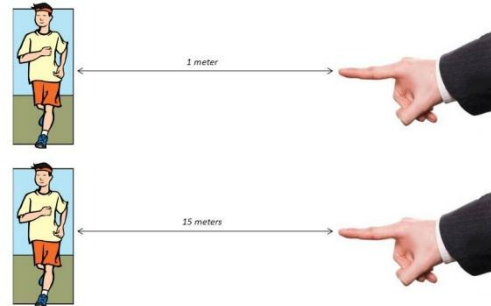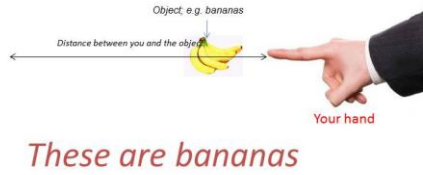
Appendix C. Consent Forms, Signed and Dated

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:
1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _M. AL Moosly_

Signature of Interviewee: _____

Date of Interview: _13 09.2009_

!

230

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:

1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _Mohammad N. Moaily_

Signature of Interviewee: _Yallais M._

Date of Interview: _13.9.2009_

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:
1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _____M. Mostly_____

Signature of Interviewee: _____

Date of Interview: _____13 . 9 . 2009_____

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:
1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _M. Moaily_

Signature of Interviewee: _____

Date of Interview: _2-9-09_

1

233

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin
Arabic. The interview is confidential, your real name and personal details are not to be
displayed in the study. Instead, pseudonyms will be used. It would be helpful if you
could start the interview by providing some background information about yourself to
the student who is interviewing you once you have both signed and dated this form
below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:
1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _K. Mohammed Ashraf_   M. Moaily

Signature of Interviewee: _Koshraf_

Date of Interview: _04.09.09._

1

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:
1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _M. Manily_

Signature of Interviewee: _Afu_

Date of Interview: _2-9-09_

1

CONSENT FORM

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:
1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _Muhammad Al Nuaily_

Signature of Interviewee: ظافر القرشي

Date of Interview: 30 - 9 - 2009

1

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:
1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: M. Maaily

Signature of Interviewee: _____

Date of Interview: 30. 9. 2009

1

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK.

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:

1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _Mohammad Al Moaily_

Signature of Interviewee: _____

Date of Interview: _5. 9 . 2009_

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:
1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _____ALavi M. Menily_____

Signature of Interviewee: _____

Date of Interview: _____5.9.2009_____

1

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:

1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _ABDUL WAHAB_

Signature of Interviewee: _____

Date of Interview: _04/9/209_

1

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:

1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _M. Moaily_

Sig nature of In terviewee: _____

Date of Interview: _21 - 8 - 2010_

241

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:

1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _M. Moaily_

Sig nature of In terviewee: _____

Date of Interview: _22-1-2010_

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:

1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: *M. Mosily*

Sig nature of In terviewee: *Wajed*

Date of Interview: 23 - 8 - 2010

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:

1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _M. Moaily_

Sig nature of In terviewee: _R_

Date of Interview: _21 - 8 - 2010_

CONSENT FORM

UNIVERSITY OF
NEWCASTLE UPON TYNE

School of English Literature,
Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RU, UK

The purpose of this interview is to provide data for linguistic description of Gulf Pidgin Arabic. The interview is confidential, your real name and personal details are not to be displayed in the study. Instead, pseudonyms will be used. It would be helpful if you could start the interview by providing some background information about yourself to the student who is interviewing you once you have both signed and dated this form below.

AGREEMENT

I agree that the recording of my interview and accompanying material may be:

1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Interviewer: _M Moaily_

Sig nature of In terviewee: _____

Date of Interview: _23 - 8 - 2010_

245

Appendix D. Maps

Map 1: Arabian Gulf States (Source: Google Maps[1])
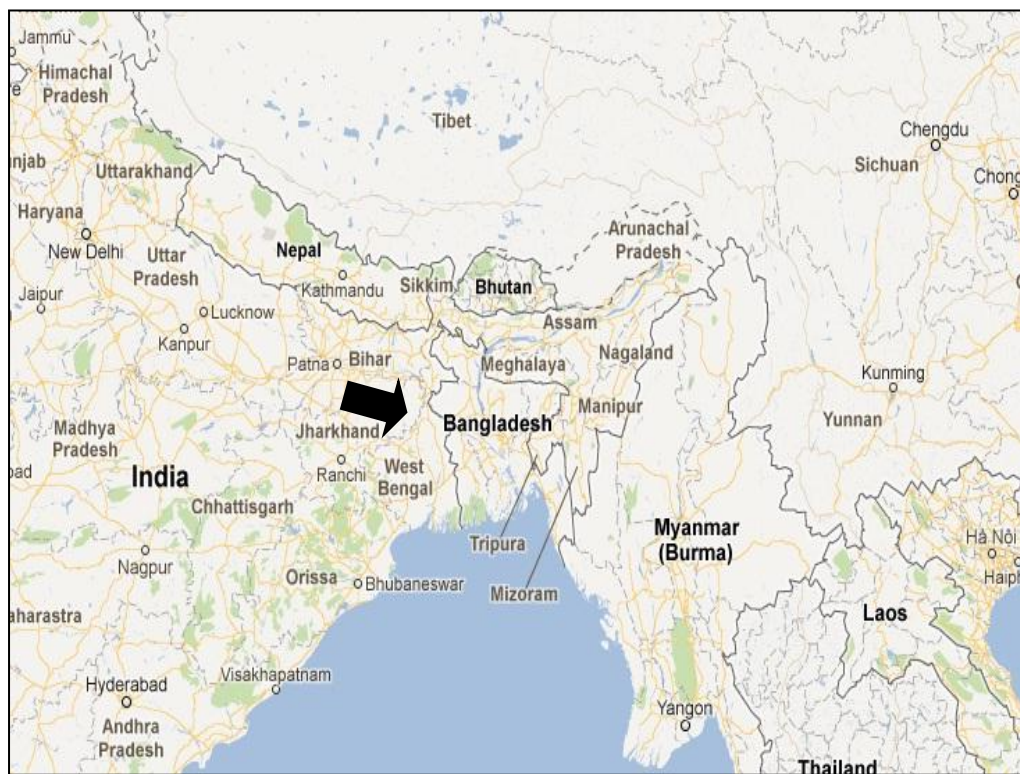
Map 2: Bangladesh and the West Bengal.

---

[1] Retrieved 5/7/2012 from: http://maps.google.com/

Map 3: The Punjab Region

Map 4: Kerala

Appendix E. Cartoon to Elicit GPA Data from GA Speakers

Figure 1: Cartoon to elicit GPA data from GA speakers, source: Alriyadh
Newspaper[1]

References

Aceto, M. and Williams, J.P. (2003) *Contact Englishes of the Eastern Caribbean*. Amsterdam: John Benjamins Pub.

Agnihotri, R. K., Khanna, A. L., and Mukherjee, A. (1988) *Tense in Indian English: a sociolinguistic perspective*. New Delhi: Bahri Publications.

Al-Agmi, F. (1997) 'al-Lahagat al-arabiyya al-haditha bayna t-tahgin wa-t-tawlid', *Magallat algami'a.* Umm al-Qura University, 10, p. 374–426. [Arabic document]

Alaidros, M. (1998) *Tariikh elkhaleej al'ararabi alhadiith walmu'asir*. Kuwait: Ain Publishing Co. [Arabic document]

Alleyne, M. C. (1971) Acculturation and the cultural matrix of creolization, in D . Hymes (Ed), *Pidginization and Creolization of Languages*, Cambridge: Cambridge University Press.

Alleyne, M.C. (1980) *Comparative Afro-American: A Historical-Comparative Study of English-Based Afro-American Dialects of the New World*. Ann Arbor, MI: Karoma

Almoaily, M. (2008) *A data-based description of Urdu Pidgin Arabic*. MA: Newcastle University

Al-Mozainy, H. (1981) Vowel Alternations in Bedouin Hijazi Arabic: Abstractness and Stress. PhD dissertation. The University of Texas, Austin

Alshammari, W. (2010) An Investigation into Morpho-syntactic Simplification in the Structure of Arabic Based Pidgin in Saudi Arabia. MA Dissertation: Mu'tah University, Jordan.

Ammon, U., Dittmar, N., and Mattheier, J. (2005) *Sociolinguistics, an international handbook of the science of language and society*. Berlin: Walter de Gruyter.

Andersen, R. W. (1983) *Pidginization and Creolization as Language Acquisition*. Rowley, Mass.: Newbury House.

Ansaldo, U., Matthews S., and Smith, G. (2011) *The Cantonese substrate in China Coast Pidgin*. In: Lefevbre, C. (ed.) *Creoles, Their Substrates and Language Typology*. Amsterdam/Philadelphia: John Benjamins.

Arends, J. and Bruyn, A. (1995) *Gradualist and developmental Hypotheses*. In Arends, J., Muysken, P. and Smith, N. (eds) Pidgins and Creoles, an introduction (1995) Amsterdam; Philadelphia: J. Benjamins

Arends, J., Muysken, P. and Smith, N. (1995) *Pidgins and Creoles: an introduction* Amsterdam; Philadelphia: J. Benjamins

Asher, R. E. and Kumari, T. C. (1997) *Malayalam*. London, Routledge.

Asher, R. E. and Simpson, J. M. (1994) *The Encyclopaedia of Language and Linguistics*, Oxford, Pergamon

Avram, A. (2010) 'An outline of Romanian Pidgin Arabic', *Journal of Language Contact*. 3, p. 20-38

Baayen, R.H. and Schreuder, R. (2003) *Morphological structure in language processing*. Berlin: Mouton de Gruyter.

Babbie, E. R. (2010) *The practice of social research* (12th ed. ed.). Australia; United Kingdom: Wadsworth Cangage Learning.

Bailey, B. H. (1966) *Jamaican Creole Syntax. Cambridge*. Cambridge University Press

Bailey, B.H. (2002) *Language, race, and negotiation of identity: a study of Dominican Americans*. New York: LFB Scholarly Pub.

Baker, P. (1987) 'The historical developments in Chinese Pidgin English and the Nature of the Relationships between the Various Pidgin Englishes of the Pacific Region', *Journal of Pidgin and Creole Languages*, 2 (2). pp. 163-207.

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*. In Wurm, S., Mühlhäusler, P., and Tryon, D. (eds) Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas. De Gruyter Mouton: Berlin

Bakir, M. (2010) 'Notes on the verbal system of GPA', *Journal of Pidgin and Creole Languages 25:2*

Baker, P. and Corne, C. (1986) *Universals, Substrata and the Indian Ocean Creoles*. In Muysken, P. and Smith, N. (Eds.) Substrata versus Universals in Creole Genesis, pp. 163-183. Amsterdam and Philadelphia: John Benjamins

Bakker, P. (1995) *Pidgins*. In Arends, J., Muysken, P., and Smith, N. (eds) Pidgins and creoles an introduction (1995) Amsterdam; Philadelphia: J. Benjamins

Bakker, P. (2003) 'Pidgin inflectional morphology and its implications for creole morphology', *Yearbook of Morphology*, 2003, Part 1, 3-33

Bakker, P. (2008) *Pidgins versus Creoles and Pidgincreoles*. In Kouwenberg, S. and Singler, J. (eds) The Handbook of Pidgin and Creole Studies (2008) New York, NY: B. Blackwell

Bakker, P. (2011) *Unpublished review of the book Une grammaire de la servitude*, by Bizri, F (2010) Librairie Orientaliste Paul Geuthner, p. 280.

Bakker, P., Daval-Markussen, A., Parkvall, O. M., and Plag, I. (2011) 'Creoles are Typologically Distinct from Noncreoles', *Journal of Pidgin and Creole Languages*, 26 (1) pp. 5-42.

Banglapedia (2006) The Bangla Language. Retrieved 13 June 2011 from http://www.banglapedia.org

Bardel, C. and Lindqvist, C. (2010) 'Approaches to Third Language Acquisition', *IRAL* 48/2-3 [Special issue]

Basch, M. (2009) *Pidgins and Creoles*. Seminar paper aus dem Jahr 2009 im Fachbereich Anglistik - Linguistik, Note: 2,7, Universität Paderborn (Institut für Anglistik und Amerikanistik), Veranstaltung: Sociolinguistics, Sprache: Englisch

Bassiouney, R. (2009) *Arabic Sociolinguistic.* Edinburgh: Edinburgh University Press.

Basu, D. and Wilbur, R. (2010) 'Complex Predicates in Bangla: an event-based analysis', *Rice Working Papers in Linguistics* 2, Spring 2010.

Bayley, R. and Preston, D. R. (1996) *Second Language Acquisition and Linguistic Variation*. Amsterdam: Benjamins.

Bekker, A. and Veenstra, T. (2003) *Creole Prototypes as Basic Varieties and Inflectional Morphology*. In Dimroth, C. and Starren, M. (eds) Information structure and the dynamics of language acquisition. Amsterdam: John Benjamins 235-64

Bell, A. and Holmes, J. (1990) *New Zealand ways of speaking English*. Clevedon: Multilingual Matters.

Besten, H. Muysken, P. and Smith, N. (1995) *Theories focusing on the European input*. In Arends, J., Muysken, P. and Smith, N. (eds) (1995) Amsterdam; Philadelphia: J. Benjamins

Bhatia, T. K. and Koul, A. (2000) *Colloquial Urdu : the complete course for beginners*. London: Routledge.

Bhatia, T. K. (1993) *Punjabi: a cognitive-descriptive grammar.* London: Routledge

Bhatt, R. (2007) *Ergativity in Indo-Aryan Languages.* Retrieved 10 October 2011 from http://people.umass.edu/bhatt/papers/mit-nov2007-handout.pdf

Bickerton, D. (1974) 'Creolisation, linguistic universals, natural semantics and the brain', University of Hawaii Woking Papers in Linguistics 6 (3) 124-141

Bickerton, D. (1981) *Roots of language*. Ann Arbor, MI: Karoma

Bickerton, D. (1984) 'The language bioprogram hypothesis', *Behavorial and Brain Sciences* 7:173-221

Birdsong, D. (1999) *Second Language Acquisition and the Critical Period Hypothesis*. London: Erlbuam.

Bizri, F. (2005) 'The Pidgin Madam, a New Arabic Pidgin', *La Linguistique* 41 (2), pp: 53-66

Bizri, F. (2010) *Pidgin Madame. Une grammaire de la servitude*. Paris: Librairie Orientaliste Paul Geuthner

Bley-Vroman, R. (1989) *What is the Logical Problem of Foreign Language Learning?* in Gass, J. and Schachter (eds.) Linguistic Perspectives on Second Language Acquisition. CUP New York: 41-68.

Bresnan, J. (2000) *The Emergence of the Unmarked Pronoun*. In Geraldine, L., Vikner, S., and Grimshaw, J. (eds) *Optimality-theoretic Syntax*. MIT Press.

Brown, E. K. and Ogilvie, S. (2009) *Concise Encyclopedia of Languages of the World*. Amsterdam ; Oxford: Elsevier Science.

Bruyn, A. (1993) *Question words in 18th century and 20th century Sranan*. In: Martle, J (ed) Historical linguistics (1991) 31-47 Amsterdam: Bemjamins

Buchstaller, I. (1999) *Causality, Conditionalit and Concessivity in Hawaiian Creole*. Unpublished MA Thesis. Konstanz University

Buchstaller, I. and Khattab, G. (To appear) *Sampling in Linguistic Research*. In Sharma, D. and R. Podeswa (eds.), Research Methods in Linguistics. Cambridge University Press.

Bukhari, N. (2009) *The Syntax of Serial Verbs in Gojri.* Ph.D. Thesis. Newcastle University, UK.

Carden, G. and Stewart, S. (1988) 'Binding theory, bioprogram, and creolization: evidence from Haitian Creole', *Journal of Pidgin and Creole Languages* 3. 1-67

Cavalli-Sforza V., Soudi, A. Mitamura, T. (2000) 'Arabic Morphology Generation Using a Concatenative Strategy', *Proceedings of the 6th Applied Natural Language ProcessingConference (ANLP 2000)*, pages 86–93, Seattle, Washington, USA.

Chambers, J. K., Trudgill, P., and Schilling-Estes, N. (2004) *The Handbook of Language Variation and Change*. Malden, Mass; Oxford: Blackwell.

Chambers, R., Drinkwater, C., and Boath, E. (2003) *Involving patients and the public: how to do it better*. Abingdon: Radcliffe Medical Press.

Chaudenson, R. (1974) '*Le lexique du parler creole de la Reunion'*, Champion. 2 vols. 37

Cheshire, J., Kerswill, P., and Williams, A. (2005) *Phonology, Grammar and Discourse in Dialect Convergence*. In: Auer, P., Hinskens, F., and Kerswill, P. (eds). Cambridge: Cambridge University Press.

Coelho, F. A. (1880-86) Os dialectos romancious ou neolatinos na Africa, Asia ne America. Bolletim da soceidade de Geographia de Lisoa

Collins, J. T. (1980) *Ambonese Malay and Creolization Theory*. Kuala Lumpur: Dewan Bahasa Dan Pustaka Kementerian Pelajaran Malaysia

Corder, G. W. and Foreman, D. I. (2009) *Nonparametric Statistics for Non-statisticians: a step-by-step approach*. Oxford: Wiley-Blackwell.

Corriente, F. (1976) 'From Old Arabic to Classical Arabic through the pre-Islamic koine: some notes on the native grammarians' sources, attitudes and goals' *Journal of Semitic Studies* 21, 62-98.

Crowley, T. (2008) Non-Indo European Pidgins and Creoles. In Kouwenberg, S. And Singler, J. (eds) The Handbook of Pidgin and Creole Studies (2008) New York, NY: B. Blackwell

Davies, A. (2003) *The Native Speaker: Myth and Reality*. Clevedon, England: Multilingual Matters LTD.

Dayal, V. and Mahajan, A. (2004) *Clause Structure in South Asian Languages*. Boston, Mass; London: Kluwer Academic Publishers.

De Vaus, D. A. (2001) *Research Design in Social Research*. London: SAGE.

DeCamp, D. (1971) *Toward a Generative Analysis of a Post-creole Speech Continuum*. In: Hymes, D. (ed.), Pidginization and creolization of languages: 394-370

DeCamp, D. (1977) *The Development of Pidgin and Creole Studies*. In Valdman, A. (ed) 3-20

DeGraff, M. (1999) *Creolisation, Language Change, and Language Acquisition, and Epilogue*. In DeGraff (ed.) Language Creation and Language Change, Creolisation Diachrony and Development. Cambridge/MA London: MIT Press

Deumert, A. and Durrleman, S. (2006) *Structure and Variation in Language Contact*. Amsterdam; Philadelphia: J. Benjamins.

Dillard, J. L. (1970) 'Principles in the history of American English: Paradox, virginity, and cafeteria' *Florida Foreign Language Reporter* 8, pp. 32-33

Drecshel, E. J. (1996) *Native American Contact Languages of the Contiguous United States*. In: Wurm, S. A., Mühlhäusler, P., and Tryon, D. T. (eds.), Atlas of Intercultural Communication in the Pacific, Asia and the Americas, 2 pp. 1213–1239

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 http://wals.info/

Dryer, M. S. and Haspelmath, M. (eds.), (2011) *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. Available online at http://wals.info/ Accessed on 2011-11-24

Eckman, F. (1977) 'Markedness and the contrastive analysis hypothesis', *Language Learning* 27.315 – 330.

Edmondson, W. (2004) *Code-switching and World-switching in foreign Language Classroom Discourse*. In: House, J. and Rehbein, J. (eds) Multilingual Communication. Amsterdam: Brngamins, 155-178

Eklund, R (1996) *Derek Bickerton's Bioprogram: A Proposal and Its Critics*. Course Paper, Dialects and Language Contact. Lectures held by Peter Trudgill, Uppsala University Fall Term 1996. Retrieved Monday 14 May 2012 from: http://www.ida.liu.se/~g-robek/pdf/Eklund_1996_Bickertons_Bioprogram.pdf

El-Daly, O. (2005) *Egyptology: The Missing Millennium, Ancient Egypt in Medieval Arabic Writings*. London: University College of London

Eubank, L. (1991) *Point counterpoint: Universal grammar in the second language*. Amsterdam: Benjamins.

Faine, J. (1937) *Philologie Creole*. 2d ed. Port-au-Prince: Imprimerie de l'Etat

Faraclas, N. (1988) 'Nigerian Pidgin and the languages of southern Nigeria' *Journal of Pidgin and Creole Languages*, 3 (2), 77-97

Fassi Fehri, A. (1993) *Issues in the Structure of Arabic Clauses and Words.* Dordrecht: Kluwer Academic Publishers.

Feghali, H. J. (2004) *Gulf Arabic: the dialects of Riyadh and eastern Saudi Arabia: grammar, dialogues, and lexicon*. Springfield, VA, Dunwoody Press.

Ferguson, C. (1959). 'Diglossia'. *Word* 15: 325–340

Ferguson, C. A. (1971) *Absence of copula and the notion of simplicity: a study of normal speech, baby talk, foreogner talk and pidgins*. In: Hymes, D. (ed.) 1971: 141-150.

Ferguson, C. A. (1989) 'Grammatical Agreement in Classical Arabic and the Modern Dialects', *Al-'Arabiyya* 22.5-17.

Ferguson, C. A. (1991) *Individual and Social in Language Change: diachronic changes in politeness agreement in forms of address*. In Fishman, J. A., Cooper, R. L., and Spolsky, B. (eds). The Influence of language on culture and thought : essays in honor of Joshua A. Fishman's sixty-fifth birthday. Berlin ; New York: Mouton de Gruyter.

Fernández, M. (1993) *Diglossia : a comprehensive bibliography, 1960-1990 : and supplements*. Amsterdam: J. Benjamins.

Field, F. (2004) *Second Language Acquisition in Creole Genesis*. In Escure, G. and Schwegler, A. (eds). Creoles, contact and language change: Linguistics and social implications. Amsterdam: John Benjamins. 127-60

Finch, S. (2001) Copular Elements in Bengali and the Stage/individual Level Distinctions. University of Texas: Austin. Retrieved Monday 19 April 2010.

Fischer, J. L. (1958) 'Social Influences on the Choice of a Linguistic Variant', *Word, 14, 47-56*

Fischer, W. (1995) *Zum Verhältnis der neuarabischen Dialekte zum Klassisch-Arabischen. Dialectologia Arabica:* A collection of articles in honour of the sixtieth birthday of Professor Heikki Palva, 75–86. Helsinki: Finnish Oriental Society.

Fries, C. C. and Pike, K. L. (1949) 'Coexistent phonemic systems', *Language*, 25: 29–50.

Frohnmeyer, L. J. (1989) *A progressive grammar of the Malayalam language* New Delhi ; Madras: Asian Educational Services.

Gass, S. M. and Selinker, L. (2008) *Second Language Acquisition: an introductory course*. New York ; London: Routledge.

Gleitman, L. R. and Liberman, M. (1995) *An Invitation to Cognitive Science*. Cambridge, Mass; London: MIT Press.

Gomaa, Y. (2007) 'Arabic Pidginization: The Case of Pidgin in Saudi Arabic' *Journal of the Faculty of Arts*. Assiut University, Egypt, vol. 19, 85-120

Graddol, D. and Swann, J. (1989) *Gender voices*. Oxford: Basil Blackwell.

Granda, G. (1968) La Tipologia"criolla" de dos hablos del area linguistic hispanica, Thesaurus 23: 3-15, Bogota, Instituto Caro y Cuervo

Grant, R. W. and Sugarman, J. (2004) 'Ethics in Human Subjects Research: Do Incentives Matter?' *Journal of Medicine and Philosophy*, 29(6): 717–738, 2004

Greenfield, W. P. (1830) *A defence of the Surinam Negro-English Version of the New Testament. In reply to the animadversions of an anonymous writer in the Edinburgh Christian Instructor*. By W. G, London.

Guy, G. R. (1981) *Linguistic Variation in Brazilian Portuguese: Aspects of the phonology, syntax, and language history*. Doctoral Dissertation, University of Pennsylvania

Hall, R. A. (1966) *Pidgin and Creole Languages*. Ithaca: Cornell University Press.

Hall, R. A. (1968) Creole Linguistics. In Sebeok (ed) 1986 361-71

Han, Z. (2004) *Fossilization in Adult Second Language Acquisition*. Clevedon: Multilingual Matters.

Hancock, I. (1977) *Recovering Pidgin Genesis: Approaches and problems*. In Valdman, A. (ed.) Pidgin and Creole Linguistics. Blooming ton: Indiana University Press, 277-294.

Hancock, I. (1986) *The Domestic Hypothesis, Diffusion and Componentiality: An account of Atlantic Anglophone creole origins*. In Muysken, P. and Smith, N. (eds). Substrata versus Universals in creole genesis. Amsterdam: John Benjamins

Hancock, I. (1980) *Lexical Expansion in Creoles*. In Valdman, A. and Highfild, A. (eds) Theoretical Orientations in Creole Studies, 63-88. Bloomington: Indiana University Press

Haspelmath, M. and Tadmor, U. (2009) *Loanwords in the World's Languages*. Berlin: Walter de Gruyter GmbH and Co.

Haspelmath, M. Michaelis (to appear) *The Atlas of Pidgin and Creole Language Structures*. Accessed 14 May 2012:

http://lingweb.eva.mpg.de/apics/index.php/The_Atlas_of_Pidgin_and_Creole_Language _Structures_(APiCS)

Hassan, A.J. (1987) *A text-based model for the disambiguation of the temporal inerpretation of the verb in modern standard Arabic*. [electronic resource].

Heine, B. (1979) *Some Linguistic Characteristics of African-based Pidgins*. In Hancock, I., Polomé, E. C., Goodman, M.F., and Heine, B. (eds), Readings in creole linguistics. Gent, E Story-Scientia: 89-98

Heine, B. (1982) 'The Nubi Language of Kibera, an Arabic creole: grammatical sketch and vocabulary' *Language and dialect atlas of Kenya*. 3 Berlin: Reimer

Hobrom, A. (1996) *An Analysis of the Arabic Pidgin spoken by Indian Workers in Saudi Arabia*, MA thesis. Kansas University.

Holes, C. (1986) 'Review of Versteegh (1984)', *Bibliotheca Orientalis* 43.218–222.

Holes, C. (1990) *Gulf Arabic*. London, Routledge.

Holm, J. A. (1988) *Pidgins and Creoles*. Cambridge: Cambridge University Press.

Holm, J. A. (2000) *An Introduction to Pidgin and Creoles*. Cambridge: Cambridge University Press.

Hong, Y. (1991) 'A Sociolinguistic Study of Seoul Korean: with a special section on language divergence between North and South Korea' Seoul Research Center for Peace and Unification of Korea**:** xxiii,229p.

Hopper, P. J. and. Traugott, E. C. (2003) *Grammaticalization*. Cambridge: Cambridge University Press.

Huber, M. (1999) *Ghanaian Pidgin English in its West African Context: a sociohistorical and structural analysis*. Amsterdam; Philadelphia: J. Benjamins.

Hudson, R. A. (1996) *Sociolinguistics* (2nd ed). Cambridge: Cambridge University Press.

Humayoun, M. H. and Ranta, A. (2007) *Urdu Morphology, Orthography andLexicon Extraction*. CAASL-2: The Second Workshop on Computational Approaches to Arabic

Hymes, D. E. (1971) *Pidginization and Creolization of Languages.* Cambridge: Cambridge University Press.

Jarrah, M. A. S (1993) *The Phonology of Medina Hijazi Arabic: A Non-Linear Analysis.* PhD thesis, University of Essex

Jayaseelan, K.A. (2000) Lexical Anaphors and Pronouns in Malayalam. In: Lust et al. (eds.), *Lexical Anaphors and Pronouns in Selected South Asian Languages.* Berlin: Mouton de Gruyter.

Jayaseelan, K.A. (2004) *The Serial Verb Construction in Malayalam.* In Dayal, V. and Mahajan, A. (eds.), *Clause Structure in South Asian Languages*, pp 67-91. Dordrecht: Kluwer Academic Publisher

Jenkins, J. (2003) *World Englishes: A resource book for students.* London: Routledge.

Jourdan, C. (1991) 'Pidgins and Creoles, the Blurring Categories', *Annual Review of Anthropology* 20, 187-209

Kaye, A. and Tosco, M. (2001): *Pidgin and Creole Languages, a basic introduction.* München: Lincom Europa.

Kenstowicz, M. (1989) *The Null Subject Parameter in Modern Arabic Dialects.* In: Jaeggli, O. and Safir, K. (eds) *The Null Subject Parameter.* Dordrecht: Kluwer.

Klein, W. and Perdue, C. (1997) 'The basic variety (or couldn't natural languages be much simpler)', *Second Language Research* 13 (4). 301-347

Knapik, A. (2009) 'On the Origin of Pidgin and Creole Languages, an outline' *Styles of Communication*, 1 (1) 1-5

Kouwenberg, S. And Singler, J. (2008) The Handbook of Pidgin and Creole Studies New York, NY: B. Blackwell

Kouwenberg and La Charité (2001) *The Mysterious Case of Diminutive yala-yala.* In Le Page, R. and Christie, P. (eds) Due Respect: Essays on English and English-Related Creoles in the Caribbean. Mona, Jamaica: University of the West Indies Press

Kremers, J. (2005) *Arabic Pidgins and Arabic Colloquials: a syntactic comparison.* University of Nijmegen, the Netherlands. Unpublished paper. Retrieved Monday 14 May 2012 from http://wwwuser.gwdg.de/~jkremer/papers/Kremers-Pidgins.pdf

Kuster, W. (2003) *Linguistic Complexity, the influence of social change on verbal inflection.* Utrecht: LOT Publications

Kvale, S. and Brinkmann, S. (2008) *Interviews: Learning the craft of qualitative research interviewing* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Labov, W. (1966) *The social stratification of English in New York City.* Washington: Center for Applied Linguistics.

Labov, W. (1972) *Language in the Inner City; studies in the Black English vernacular*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1984) *Field Methods of the Project on Linguistic Change and Variation*. In Baugh, J. and Herzer, J. (eds), Language in use. Englewood Cliffs, New Jersey: Prentice-Hall, 28-53.

Le Page, R. B. and Tabouret-Keller, A. (1985) *Acts of Identity: Creole-based approaches to language and ethnicity*. Cambridge; New York: Cambridge University Press.

Lefebvre, C. (1998) *Creole Genesis and the Acquisition of Grammar: The case of Haitian Creole*. Cambridge: Cambridge University Press.

Lefebvre, C. (2004) *Issues in the Study of Pidgins and Creole Languages.* Amsterdam; Philadelphia, PA: J. Benjamins Pub

Leung, Y. I. (2005) 'L2 vs. L3 initial state: A comparative study of the acquisition of French DPs by Vietnamese monolinguals and Cantonese-English bilinguals', *Bilingualism: Language and Cognition*, 8, 39-61.

Lilliefors, H. W. (1967) 'On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown', *Journal of the American Statistical Association*, **62**, 399–402.

Llamas, C. (1999) 'A New Methodology: data elicitation for social and regional language variation studies', *Leeds Working Papers in Linguistics and Phonetics* 7: 95–118.

Long, M. H. and Doughty, C. (2009) *The handbook of language teaching*. Chichester: Wiley-Blackwell.

Lucas, C., Bayley, R. and Valli, C. (2001) *Sociolinguistic variation in American sign language*. Washington, D.C.: Gallaudet University Press.

Luffin, X. and Woidich, M. (to appear) *Bongor Pidgin Arabic*. In: Versteegh, K., Eid, M., Eljibali, A., Woidich, M., and Zaborski, A. (eds) Leiden: Brill

Luria, H., Seymour, D. M., and Smoke, T. (2006) *Language and Linguistics in Context: readings and applications for teachers*. Mahwah, N.J; London: Lawrence Erlbaum Associates.

Lutzeier, P. R. (2005) *International Handbook on the Nature and Structure of Words and Vocabularies.* Berlin: Walter de Gruyter

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.

Mahmud, U. A. (1979) *Variation and Change in the Aspectual System of Juba Arabic*. Ph.D. Thesis, Georgetown University.

Markey, T. (1982) 'Afrikaans: Creole or non-creole?', *Zeitschrift fur Dialektologie und Lin- guistik* 49: 169-207

Mather, P. (2000) *Cross-Linguistic Influence in Second Language Acquisition and in Creole Genesis*. PhD dissertation, University of Pittsburgh

McArthur, T. (1998) *The English Languages*. Cambridge: Cambridge University Press.

McMahon, A. M. S. (1994) *Understanding Language Change. Cambridge*; New York, NY: Cambridge University Press

McWhorter, J. H. (1995) 'The Scarcity of Spanish-Based Creoles Explained', *Language in Society*, 24: 213–44.

McWhorter, J. H. (2000) *Defining "creole" as a Synchronic Term*. In Neumann-Holzschuch, I. And Schneider, E. (eds) Degrees of restructuring in Creole languages. Amsterdam, J. Benjamins.

McWhorter, J. H. (2005) *Defining Creole*. Oxford: Oxford University Press.

Mehrutra, R. R. (1997) 'Reduplication in Indian Pidgin English', *English Today* 50, vol. 13 45-49

Mesthrie, R. (1989) 'The Origins of Fanakalo' *Journal of Pidgin and Creole Languages* 4, 211-40

Mesthrie, R. (2002) *South Africa: A Sociolinguistic Overview*. In Mesthrie, R. (ed) Language in South Africa, 11-26. Cambridge: Cambridge University Press.

Meyerhoff, M. (2008) *Forging Pacific Pidgin and Creole Syntax: substrate, discourse and inherent variability.* In Kouwenberg, S. and Singler, J. (eds) The Handbook of Pidgin and Creole Studies (2008) New York, NY: B. Blackwell

Miller, C. (2002) *The Relevance of Arabic-based Pidgins and Creoles for Arabic Linguistics*. In Mansur, G. and Doss, M. (eds) Al-Lugha, Cairo: Arab Development Centre, p. 7-46

Miller, C. and Woidich, M. (to appear) *Juba Arabic*. In: Versteegh, K., Eid, M., Eljibali, A., Woidich, M., and Zaborski, A. (eds) Leiden: Brill

Milne, W. S. (1993) *A Practical Bengali Grammar*. Calcutta: The Author.

Milroy, L. and Gordon, M. (2003) *Sociolinguistics: method and interpretation*. Malden, Mass; Oxford: Blackwell.

Montrul, S. (2004) *The Acquisition of Spanish: morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition*. Amsterdam; [Great Britain]: John Benjamins Publishing.

Mufwene, S. F. (1986) 'Les langues créoles peuvent-elles être définies sans allusion à leur histoire?' *Etudes Créoles* 9.135-50.

Mufwene, S. F. (1990) 'Transfer and the Substrate Influence in Creolistics', *Studies in Second Language acquisition*, 12:1-23

Mufwene, S. F. (1993) *Africanisms in Afro-American Language Varieties*. Athens: University of Georgia Press.

Mufwene, S. F. (2001) *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.

Mufwene, S. F. (2002) *Pidgins and Creole Languages.* Accessed: 26 April 2011 from http://humanities.uchicago.edu/faculty/mufwene/pidginCreoleLanguage.html

Mufwene, S. F. (2006) *Pidgins and Creoles*. In Kachru, B., Kachru, Y., and Nelson C. L. (eds.) The Handbook of World Englishes. New York: Blackwell Publishing

Mufwene, S. F. (2008) *Language Evolution, Contact, Competition and Change*. London: Continuum International Publishing Group.

Mühlhäusler P. (1986) *Pidgin and Creole Linguistics*. London, University of Westminster Press

Mühlhäusler, P. (1997) *Pidgin and Creole Linguistics*. London: Battlebridge Publications

Mühlhäusler P., Dutton, T. E., and Romaine, S. (2003) *Tok Pisin Texts: From the Beginning to the Present*. Amsterdam/Philadelphia: Benjamins.

Müller-Gotama, F. (1994) *Grammatical Relations: a cross-linguistic perspective on their syntax and semantics*. Berlin; New York: M. de Gruyter.

Mundy, P., D. Butchart, J. Ledger, and S. Piper (1992) *The Vultures of Africa*. London: Academic Press

Muysken, P. (1994) 'The search for Universals in language genesis: Etat de la question and research program', Amsterdam Creole Studies, 2, l6—28.

Muysken, P. and Meijer, G. (1979) *Introduction to Hesseling*, pp. vii-xix

Muysken, P. and Smith, N. (1995) *The Study of Pidgin and Creole Languages*. In Arends, J., Muysken, P. and Smith, N. (eds) Pidgins and creoles an introduction (1995) Amsterdam; Philadelphia: J. Benjamins

Muysken, P. and Veenstra, T. (1995) *Universalist Approaches*. In Arends, J., Muysken, P. and Smith, N. (eds) Pidgins and creoles an introduction (1995) Amsterdam; Philadelphia: J. Benjamins

Myers-Scotton, C. (1989) 'Codeswitching with English: Types of switching, types of communities' *World Englishes*, 8. 333-346.

Naess (2008) *Gulf Pidgin Arabic: Individual strategies or structured variety*. MA: University of Oslo

Naro, A. J. (1978) 'A study on the Origins of Pidginisation', *Language,* 54 (2): 314-349.

Nasrin, M. B. and Wurff, W. v. d. (2009) *Colloquial Bengali: the complete course for beginners*. London: Routledge.

Nemser, W. (1971) 'Approximative systems of foreign language learners', *IRAL*, 9, (2), 115-124.

Newman, P. and Ratliff, M. S. (2001) *Linguistic Fieldwork*. Cambridge: Cambridge University Press.

Odlin, T. (1989) *Language Transfer: cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Owens, J. (1977) *Aspects of Nubi Syntax*. Ph.D. thesis. University of London.

Owens, J. (1980) 'Monogenesis, the Universal and the Particular in Creole Studies', *Anthropological Linguistics*, 22: 97-117

Owens, J. (1989) 'Zur Pidginisieierung und Kreolisierung im Arabischen', *Afrika und Ubersee*. 72, 91-107

Owens, J. (1996) *Arabic-based Pidgins and Creoles.* In: Thomason, S. (ed) Contact Languages: a wider perspective, ed. Sarah G. Thomason. Amsterdam: John Benjamins.

Owens, J. (2001) 'Arabic Creoles: the orphan of all orphans' *Anthropological Linguistics,* 43: 348- 78. 2000

Patrick, P. (1999) *Urban Jamaican Creole: variation in the mesolect.* Amsterdam; Philadelphia, PA: J. Benjamins Pub.

Patrick, P. L. (2009) Pidgins, Creoles, and Variation. In Kouwenberg, S. And Singler, J. (eds) The Handbook of Pidgin and Creole Studies (2008) New York, NY: B. Blackwell

Paul, S. (2003) *Composition of Compound Verbs in Bangla. Multi-Verb constructions*. Trondheim Summer School.

Penny, R. J. (2000) *Variation and Change in Spanish*. Cambridge: Cambridge University Press.

Poplack, S. (1989) *The Care and Handling o f a Megacorpus: the Ottawa-Hull French project*. In Fasold, R. and Schiffrin, D. (eds), Language Change and Variation. Amsterdam: Benjamins, 4 1 1 -444.

Powers, W. R. (2005). *Transcription techniques for the spoken word*. Lanham, MD: AltaMira Press.

Precht, K. (2008) 'Sex Similarities and Differences in Stance in Informal American Conversation', *Journal of Sociolinguistics*, 12 (1) 89-111

Qafisheh, H. A. (1977) *A Short Reference Grammar of Gulf Arabic*. Tucson, Ariz., University of Arizona Press.

Rajaraja V. A. R. and Roy, C. J. (1999) *Kerala Paaniniiyam: a treatise on Malayalam grammar*. Trivandrum, Kerala, India: International School of Dravidian Linguistics.

Ramat, A. (2009) 'Typological Universals and Second Language Acquisition', *Universals of Language Today*, 76, 253-272

Rampton, M. B. H. (1990) 'Displacing the Native Speaker: expertise, affiliation, and inheritance', *ELT Journal* 44 (2): 97-101.

Rasinger, S. M. (2007) *Bengali-English in East London : a study in urban multilingualism*. Oxford: Peter Lang.

Ray, J. J. (1985) 'Random Sampling Might Not Be Impossible After All', *Political Psychology* 6: 141-46.

Ray, S. P., Abdul Hai, M., and Ray, L. (1966) *Bengali Language Handbook*. Massachusetts: Centre for Applied Linguistics.

Reinecke, J. E. (1937) *Marginal Languages: a sociological survey of the creole languages and trade jargons*. PhD dissertation, Yale University. Ann Arbor: University Microfilms international

Reinecke, J. E. (1977) *Foreword to Valdman*. In Valdman, A. (ed) vii-xi

Reinecke, J. E. (1981) *A Selective Chronology of Creole Studies*. Carrier Pidgin supplement. September 1981.

Reinecke, J. E., Tsuzaki, S. M., et al. (1975) *A Bibliography of Pidgin and Creole Languages*. Honolulu: University of Hawaii Press

Rice, K. (2006) 'Ethical Issues in Linguistic Fieldwork: An overview', *Journal of Academic Ethics, 4,* 123-155.

Rickford, J. (1998) *The Creole Origins of African American Vernacular English: Evidence from copula absence*. In Mufwene, S., Bailey, G., and Baugh, J. (eds), *African American English*. London: Routledge.

Rickford, J. R. and McNair-Knox, F. (1994) *Addressee and Topic Influenced Style Shift: a quantitative sociolinguistic study*. In: Biber, D. and Finegan, E. (eds). Oxford: Oxford University Press.

Roberts, S, and Bresnan, J. (2008) 'Retained Inflectional Morphology in Pidgins: A typological study', *Linguistic Typology* 12: 269–302

Roberts, S. (1999) *The TMA System of Hawaiian Creole and Diffusion*. In Rickford, R. and Romaine, S. (eds.), *Creole Genesis, Attitudes and Discourse: Studies Celebrating Charlene J. Sato.* Amsterdam/Philadelphia: Benjamins, 45–70.

Romaine, S. (1988) *Pidgin and Creole Languages*. London; New York: Longman

Romaine, S. (1992) Language, education, and development : urban and rural Tok Pisin in Papua New Guinea. Oxford: Clarendon.

Ryding, K. C. (2005) *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.

Sabourin, L.L., Stowe, L.A., and Haan, G.J. (2006) 'Transfer effects in learning and L2 grammatical gender system', *Second Language Research*, 27(1), 1-29.

Sakoda, K. and Siegel, J. (2003) *Pidgin Grammar: an introduction to the Creole English of Hawai'i*. Honolulu, Hawaii: Bess Press.

Samarin, W. J. (1971) *Salient and Substantive Pidginisation*. In. Hymes, (ed), 117-40

Samarin, W. J. (1982) 'Goals, Roles, and Language Skills in Colonizing Central Equatorial Africa', *Anthropological Linguistics* 24, (4) pp. 410-422

Samarin, W. J. (1986) 'Protestant Missions and the History of Colonialisms', *Journal of Religion in Africa* 16, 138-63.

Samarin, W. J. (2000) *The status of Sango in Fact and Fiction*. In McWhorter, J. H (ed) Language change and language contact in pidgins and creoles. Amsterdam, Netherlands ; [Great Britain]: J. Benjamins.

Sankoff, G. and Brown, P. (1976) 'The Origins of Syntax in Discourse: A case study of Tok Pisin relatives'. *Language* 52(3): 631-666

Sankoff, G. and Laberge, S. (1974) *On the Acquisition of Native Speakers by a Language*. In D. DeCamp and IF Hancock, (eds). Pidgins and creoles: Current trends and prospects, 73-84. Washington, DC: Georgetown University Press

Sapir, E. (1921) *Language : an introduction to the study of speech*. New York: Harcourt.

Satorra, A. and Bentler, P. M. (2001) 'A scaled difference chi-square test statistic for moment structure analysis' *Psychometrika*, 66, 507-514.

Schachter, J. (1988) 'Second Language Acquisition and its relationship to Universal Grammar', *Applied Linguistics* 9, 3, 219-235

Schäffner, C. (2002) *Third Ways and New Centres—Ideological unity or difference?* In: Pérez, M. (Ed.), Apropos of Ideology. Translation Studies on Ideology–Ideologies in Translation Studies. St. Jerome, Manchester, pp. 23–41.

Schilling-Estes, N. (2007) Sociolinguistic Fieldwork. In: Bayley, R. and Lucas, C. (eds.) Sociolinguistic Variation Theories, Methods, and Applications:. Cambridge: Cambridge University Press

Schmidt, R. L. (1999) *Urdu: an essential grammar*. London: Routledge.

Schramm, G. (1962) 'An Outline of Classical Arabic Verb Structure', *Language* 38:4.

Schuchardt, H.E.M. (1885) *Ueber die Lautgesetze. Gegen die Junggrammatiker*. Berlin.

Schulz, E. (2004) *A student grammar of modern standard Arabic*. Cambridge: Cambridge University Press.

Schumann, J. (1986) 'Locative and Directional Expressions in Basilang speech', *Language Learning*, 36, 277-294.

Sebba, M. (1997) *Contact Languages: pidgins and creoles*. London, Macmillan.

Sedlatschek, A. (2009) *Contemporary Indian English: variation and change*. Amsterdam; Philadelphia: John Benjamins Pub. Co.

Selinker, L. (1972) 'Interlanguage' *IRAL, 10,* (3), 209-231.

Selinker, L. (1996) *Fossilization: What We Think We Know*. London: Longman Group Limited

Serrano, J. M., Garzon, F. C., Manzanares, J. V (2003) *The Creolization of Pidgin: A connectionist exploration.* In Schmalhofer, F. Young, R. and Katz, G. (eds) Proceedings of EuroCogSci '03, the European Cognitive Science Conference. New Jersey: Laurence Erlbaum Associates, Inc.

Seuren, P. (1984) 'The Bioprogram Hypothesis: Facts and fancy', *Behavioral and Brain Sciences* 7: 208–209

Seuren, P. and Wekker, H. (1986) *Semantic Transparency as a Factor in Creole Genesis*. In Muysken, P. and Smith, N. J. (eds.) Substrata versus Universals in Creole Genesis. Amsterdam: Benjamins. 57-70.

Shackle, C. (1970) 'Punjabi in Lahore', *Modern Asian Studies*, 4, 239-67

Shuy, R. W., Riley, W. K., and Wolfram, W. A. (1968) *Field Techniques in an Urban Language Study.* Washington: Center for Applied Linguistics.

Sidnell, J. (2002) 'Habitual and Imperfective in Guyanese Creole'. *Journal of Pidgin and Creole Languages*, 17(2): 151–182

Siegel, J. (1982) 'Plantation Pidgin Fijian', *Oceanic Linguistics* 21:1-72

Siegel, J. (1999) 'Transfer Constraints and Substrate Influence in Melanesian Pidgin', *Journal of Pidgin and Creole Languages,* 14: 1–44.

Siegel, J. (2003) 'Substrate Influence in Creoles and the Role of Transfer in Second Language Acquisition', *Studies in Second Language Acquisition,* 25: 185–209

Siegel, J. (2004) 'Morphological Simplicity in Pidgins and Creoles', *Journal of Pidgin and Creole Languages,* 19: 139–62.

Siegel, J. (2008a) *Pidgins, Creoles, and Second Language Acquisition.* In Kouwenberg, S. And Singler, J. (eds) The Handbook of Pidgin and Creole Studies (2008) New York, NY: B. Blackwell

Siegel, J. (2008b) *The Emergence of Pidgin and Creole Languages.* Oxford: Oxford University Press

Siegel, J. (2010) *Pidgins and Creoles.* In Hornberger (ed.) Sociolinguistics and language education (2010) Uk, Short Run Press Ltd.

Singh, I. (2000) *Pidgins and Creoles: an introduction*. Oxford: Oxford University Press

Skousen, R. (1989) *Analogical Modelling of Language*. Dordrecht: London: Kluwer Academic.

Smart, R. J. (1990) 'Pidginisation in Gulf Arabic: a first report', Anthropological Linguistics 2 (1- 2) pp.37.

Smith, N. (1987) *The Genesis of the Creole Languages of Surinam*. PhD Thesis. University of Amsterdam

Smith, N. (1995) *An Annotated List of Pidgins, Creoles, and, Mixed Languages*. In Arends, J., Muysken, P. and Smith, N. (eds) Pidgins and creoles an introduction (1995) Amsterdam; Philadelphia: J. Benjamins

Smith, N. (2008) 'The origin of Portuguese words in Saramaccan: implications for sociohistory', *Journal of Portuguese Linguistics* 3: 115-147

Souag, L. (2006) *The Earliest Recorded Pidgin - Maridi Arabic?* Accessed 10 May 2011 from http://lughat.blogspot.com/2006/12/earliest-recorded-pidgin-maridi-arabic.html

Sridhar, K. (1991) *Speech Acts in an Indigenized Variety: Sociocultural values and language variation.* In: Cheshire, J. (ed.) *English Around the World: Sociolinguistic Perspectives*, 308-18. Cambridge: Cambridge University Press.

Sridhar, K. (1996) *Societal Multilingualism*. In. Mckay, S. L. and Hornberger, N. H. (eds.) Sociolinguistics and Language Teaching (pp. 47-71). Cambridge: Cambridge University Press.

Stanford University (2005) *Pidgins and Creoles: a guide to green library collections*. Retrieved Monday 11th of May 2009 from: http://www-sul.stanford.edu/depts/ssrg/pidgins/pidgin.html

Steever, S. B. (1987) *The Serial Verb Formation in the Dravidian Languages*. Delhi: Motilal Banarsidass, 1988.

Stewart, J. M. (1965) *The Typology of the Twi Tone System*. Reprint from the Bulletin of the Institute of African Studies 1, University of Ghana, Legon: Ghana

Stewart, W. A (1962) *Creole Languages in the Caribbean, study of the role of second languages in Asia, Africa, and Latin America*. In Rice, F. (ed) 34-53, Washington, DC: Center for Applied Linguistics

Subramoniam, V. I. (1997) *Dravidian Encyclopaedia*. vol. 3, Language and Literature. Thiruvananthapuram: International School of Dravidian Linguistics. Cit-P-487. Dravidian Encyclopedia

Suleiman, Y. (1999) *Language and Society in the Middle East and North Africa*. New York: Routledge.

Tagliamonte, S. (2006) *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Tarone, E. (1980) 'Some Influences on the Syllable Structure in Interlanguage Phonology', *IRAL*, 18-40

Taylor, D. R. (1971) *Grammatical and Lexical Affinities of Creoles*. In Hymes, D. (ed) 1971:293-296

Taylor, D. R. (1977) *Languages of the West Indies*. Baltimore: Johns Hopkins University Press.

Tennent, M. (2005) *Training for the New Millennium: pedagogies for translation and interpreting*. Amsterdam: John Benjamins Publishing Company.

Thomason, S. (2008) *Pidgins, Creoles, and Historical Linguistics.* In Kouwenberg, S. And Singler, J. (eds) The Handbook of Pidgin and Creole Studies (2008) New York, NY: B. Blackwell

Thomason, S. G. (2001) *Language Contact: an introduction*. Washington: Georgetown University Press.

Thomason, S. G. and Kaufman, T. (1991) *Language Contact, Creolisation, and Genetic Linguistics*. Berkeley/Los Angeles/Oxford: University of California Press.

Thomason, S. G. and Eljibali, A. (1986) 'Before the lingua franca: pidginised Arabic in the eleventh century AD.', *Lingua* 68: 407-436

Todd, L. (1974) *Pidgins and Creoles*. London: Routledge

Todd, L. (1980) *Annual Review of Applied Linguistics*. Rowley, MA: Newbury House.

Todd, L. (1990) *Pidgins and Creoles*. London; New York: Routledge

Tosco, M. (to appear) Pidginisation. In: Versteegh, K., Eid, M., Eljibali, A., Woidich, M., and Zaborski, A. (eds) Leiden: Brill

Towell, R. and Hawkins, R. (1994) *Approaches to Second Language Acquisition. Clevedon*: Multilingual Matters.

Tremblay, M.-C., (2006) 'Cross-linguistic Influence in Third Language Acquisition: The role of L2 proficiency and L2 exposure. *Cahiers', Linguistiques d'Ottawa*, 34, 109-119.

Trudgill, P. (1986) *Dialects in Contact.* Oxford: Blackwell.

Tryon, D. and Charpentier, J. (2004) *Pacific Pidgins and Creoles: origins, growth, and development.* Berlin/New York: Mouton de Gruyter.

Turner, R. (2004) *Who is a Native Speaker and What is it they Speak.* Retrieved Tuesday 11 November 2011 from http://neptune.spaceports.com/~words/native.html

Valdman, A. (1978) *Le Creole: structure, statut et origine*. Paris: Klincksieck.

Veenstra, T. (2008) *Creole Genesis, the Impact of the Language Bioprogram Hypothesis.* In Kouwenberg, S. And Singler, J. (eds) The Handbook of Pidgin and Creole Studies (2008) New York, NY: B. Blackwell

Verhaar, J. M. (1995) *Toward a Reference Grammar of Tok Pisin: An Experiment in Corpus Linguistics.* Honolulu: University of Hawai'i Press.

Véronique, D. (2003) *Iconicity and Finiteness in Development of Early Grammar in French as L2 and in French-based Creoles*. In: Giacalone Ramat, A. (ed). Typology and second language acquisition. Berlin, Mouton de Gruyter.

Versteegh, K. (1984) *Pidginization and Creolization: the case of Arabic*. Amsterdam, John Benjamins Publishing Company.

Versteegh, K. (2001) *The Arabic language*, Edinburgh: Edinburgh University Press

Versteegh, K. (2004) *Pidginization and Creolization revisited: The Case of Arabic*. In: Haak D.M., de Jong, D., and Versteegh, C. (eds.) Approaches to Arabic Dialects. Leiden: E. J. Bril. 343-357

Versteegh, K. (2008) *Non-Indo European Pidgins and Creoles*. In: Kouwenberg, S. and Singler, J. (eds) The Handbook of Pidgin and Creole Studies (2008) New York, NY: B. Blackwell

Versteegh, K. to appear. Pidgin Verbs, Infinitives or Imperatives. In: Buchstaller, I., Holmberg, A., and Almoaily, M. (eds) *Pidgins and Creoles beyond Africa-Europe Encounters.* Amsterdam: John Benjamins

Vincent, D. and Sankoff, D. (1992) 'Punctors: A pragmatic variable', *Language Variation and Change*, 4, (2) p. 205-216

Wahba, K. (1996) *Linguistic Variation in Alexandria Arabi: the feature of emphasis*. In Elgibali, A. (ed), Understanding Arabic: Essays in contemporary Arabiclinguistic in honor of El-Said Badawi. (pp. 91-101) Cairo: The American University in Cairo Press.

Wansbrough, J. E. (1996) *Lingua Franca in the Mediterranean*. Richmond, Curzon Press.

Wardhaugh, R. (2009) *An Introduction to Sociolinguistics 6th ed.* Oxford: Wiley.

Weinreich, U., Labov, W., and Herzog M. (1968) *Empirical Foundations for a Theory of Language Change*. In Lehmann, W. and Malkiel, Y. (eds.), Directions for Historical Linguistics. 97-195. Austin: University of Texas Press.

Wellens, I.H.W. (2003) *An Arabic creole in Africa: the Nubi language of Uganda*. Thesis (Ph.D.)_Katholiek Universiteit Nijmegen, 2003. Katholieke Universiteit Nijmegen.

Whinnom, K. (1965) 'The origin of the European-based creoles and pidgins', Orbis 14: 509-527

Whinnom, K. (1971) *Linguistic Hybridization and the 'special case' of pidgins and creoles*. In Hymes, D. (ed.), *Pidginization and Creolization of Languages*, 91-115.

White, L. (2003) *Second Language Acquisition and Universal Grammar.* Cambridge: Cambridge University Press.

Williams, J. (2000) 'Yima-Alamblak Tanim Tok: an indigenous trade pidgin of New Guinea', *Journal of Pidgin and Creole Languages* 15: 37-62.

Winford, D. (2003) *An Introduction to Contact Linguistics.* Oxford: Blackwell

Winford, D. and Migge, B. (2007) 'Substrate influence on the emergence of the TMA systems of the Surinamese Creoles', Journal of Pidgin and Creole languages 22: (1) 73-99

Wiswall, A. (2002) *Gulf Pidgin: An expanded Analysis*. On-line Paper for the University of Ohio, Linguistic Department, http://oak.cats.ohiou.edu/~aw321500/

Zadeh, F. J. and Winder, R. B. (2003) *An introduction to Modern Arabic*: Princeton: Princeton University Press

Zyhlar, E. (1932) 'Ursrpung und Sprachcharacter des Altägpytischen', Zeitschrift für Eingeborenen-Sprachen 23, 25-45.