



Nebiye Petek Kurtböke

A Corpus-driven study of Turkish-English
Language Contact
in Australia

A thesis submitted to the Faculty of Arts

Of
Monash University
For
The Degree of Doctor of Philosophy

Department of Linguistics
Monash University
Melbourne

December 1998

Synopsis

A Corpus-driven Study of Turkish-English Language Contact in Australia

This study investigates how the language of a particular minority group (L1) behaves outside its natural environment, in a setting where another language (L2) is used. In particular, it employs a machine-readable corpus of written language samples from a variety of newspapers published by a migrant community over a period of 15 years. While the content and role of the migrant press in the maintenance of L1 have been largely exploited, linguistic analysis of ethnic newspapers has often been neglected in language contact research. Unlike previous studies, which have examined spoken data, this study therefore considers the written language to investigate the patterns emerging out of language contact situations and the motivations for such patterning. The data have been obtained from the Turkish community in Australia. The language of Turkish migrants in Australia, unlike that of their counterparts in Europe, has not been the subject of much research. Although this study concentrates on a specific ethnic community, the results contribute to the cross-linguistic studies of language contact and therefore to the description and analysis of immigrant languages. The results also have implications for mother-tongue instruction in Australian schools.

CONTENTS

page

Synopsis
Dedication
Acknowledgements

PART ONE: INTRODUCTION

1. <i>Background to this study</i>	1
1.1 "What have you found?"	1
1.1.1 Mixed code or not?	2
1.1.2 The Place of the Lexicon in Language Contact	3
1.1.3 Singly occurring switches	4
1.1.4 Which elements facilitate mixing	4
1.1.5 Borrowing or codeswitching?	5
1.1.6 Morphological integration of words	5
1.1.7 Cultural borrowings	6
1.1.8 The question of psycholinguistic stress	6
1.1.9 An integrated approach or not?	7
1.1.10. The nature of language contact data	7
1.2 Where does <i>delivery yapýlý</i> come from?	9
1.3 Why do language contact, corpus research and translation come together?	11
1.4 The organization of the thesis	17
1.5 Summary	20
2. <i>Current State of Turkish</i>	23
2.1 In and around Europe	23
2.1.1 Germany	24
2.1.2 The Netherlands	25
2.1.3 Scandinavia	27
2.1.4 UK	29
2.2 In Turkey	30
2.3 The relevance of the European and Turkish situation to this study	34
2.4 Summary	36

PART TWO: A MEETING PLACE FOR LANGUAGE CONTACT, TRANSLATION AND CORPUS RESEARCH

3. <i>Language Contact research - An Overview</i>	41
--	----

3.1 Where does Language Contact belong?	44
3.2 Five decades of language contact research	50
3.3 First decade: 1950-60	54
3.4 Second decade: 1960-70	77
3.5 Third decade: 1970-80	98
3.6 Fourth decade: 1980-1990	115
3.7 Fifth decade: 1990 to present	130
3.8 To Myers-Scotton or not to Myers-Scotton?	139
3.9 Summary	143
4. Corpus research - an overview	147
4.1 Background	146
4.2 Major issues in corpus research	153
4.2.1 Corpus design	154
4.2.2 Corpus annotation	156
4.2.2.1 Encoding	157
4.2.2.2 Tagging	158
4.2.2.3 Parsing	159
4.2.2.4 Database	159
4.2.3 Linguistic exploration of corpora	160
4.2.3.1 Sinclair	161
4.2.3.2 Francis	164
4.3 The future of corpus research	168
4.4 Corpus studies in Australia and Turkey	170
4.5 Corpus-building for language contact research and the uniqueness of textual resources in migrant settings	171
4.6 Summary	173
5. The contribution of Translation Studies	176
5.1 Migrants and translation	176
5.2 Similarities between studies of Language Contact and Translation	178
5.3 The "equivalence syndrome"	180
5.4 Translation strategies in the absence of "equivalence"	181
5.5 Summary	188
PART THREE: OZTURK CORPUS	
6. Turks in Australia	191
6.1 Turks in Australia	191
6.1.1 Before 1967	196
6.1.2 After 1967	196
6.2 Language Needs	202
6.2.1 Government Services, Community Network And Resources for Migrants	208
6.2.2 Language Use	210
6.3 Summary	213
7. Methodology	
7.1 Introduction	215

7.2 Data collection and analysis in Language Contact	220
7.3 Sampling	222
7.3.1 Ethics Clearance and Copyright	228
7.3.2 Content Categories	229
7.3.3 Time span and size	232
7.4 Hardware and Software	233
7.4.1 Data capture	235
7.4.2 Data processing	235
7.5 About written language	239
7.5.1 Nominalization	242
7.5.2 Delexicalization	244
7.5.3 Collocation	245
7.6 Summary	249

PART FOUR: "INTERNAL OPERATIONS"

8. <i>Corpus Analysis</i>	253
8.1 Emerging Patterns	253
8.1.1 The question of "single switches"	254
8.1.1.1 <i>delivery</i>	255
8.1.1.2 <i>register</i>	263
8.1.2 The question of "cultural borrowings"	269
8.1.2.1 <i>ombudsman</i>	272
8.1.2.2 <i>grants</i>	277
8.1.3 The question of "islands"	284
8.1.3.1 <i>Mature Age Allowance</i>	285
8.1.3.2 <i>Aile birle"imi</i>	288
8.1.4 The question of delexicalised verbs	290
8.1.4.1 <i>yap-</i>	295
8.1.5 The question of functors	308
8.1.5.1 <i>Eör</i>	313
8.1.6 The question of "nonce"	315
8.1.6.1 <i>licenceli</i>	320
8.2 Significance of Patterns	323
8.3 Motivations	325
8.3.1 A word on speaker-innovation	328
8.4 Summary	331
9. <i>On the characterization of C_{mx}</i>	334
9.1 <i>Codeness vs languageness</i>	339
9.1.1 Mixed code and the absence of metaphor	347
9.2 The search for a unit of analysis	349
9.3 Empty C _{mx} lexicon first	354
9.4 Mixed code and change: a comparison of 80s and 90s	361
9.5 Summary	363

PART FIVE: CONCLUSION

10. <i>Conclusions</i>	366
10.1 Lexicogrammatical tendencies in language contact	367

10.2 A Lexicogrammatical approach to language contact	370
10.3 "What have I found?"	373
10.3.1 Mixed code or not?	374
10.3.2 The Place of the Lexicon in Language Contact	375
10.3.3 Singly occurring switches	376
10.3.4 Which elements facilitate mixing	376
10.3.5 Borrowing vs codeswitching	376
10.3.6 Morphological integration of words	377
10.3.7 Cultural borrowings	378
10.3.8 The question of psycholinguistic stress	378
10.3.9 Waiting for an integrated framework: Will it ever come?	378
10.3.10. The nature of language contact data	379
10.4 Limitations of this thesis	379
10.5 Implications of this study for other areas of research	381
10.5.1 Lexicography	381
10.5.2 Grammar	382
10.5.3 Mental lexicon	383
10.5.4 Translation	384
10.5.5. Natural Language Processing	385
10.5.6. Teaching	385
10.6 Language Contact 2000	

11. Appendices

12. References

.....
List of Tables

Glossary

CHAPTER 1: Background to this study

1.1 “What have you found?”

I am frequently asked, especially by fellow-students and colleagues who have written on the same topic, the following question about my thesis: “What have you found?”. I answer this question by trying to explain where my thesis stands in relation to a set of questions faced by all language contact scholars at some stage of their research. This section is about that set of questions and the position adopted in this thesis. The section on Language Contact (Chapter 3) traces the developments around the broader issues for the last five decades in a more detailed way. They are different from the research questions listed in 1.3.

From the way the questions below are formulated, it will be obvious that the focus of this study is not on whether Turkish is holding its own despite the influence of English in Australia. Nor is it on whether the members of the Turkish community have assimilated to the dominant culture and where along the bilingual continuum they stand at this particular moment.

It rather brings a number of research areas together; namely, Language Contact, Corpus Research and Translation Studies, to answer various questions of general concern. The reasons for bringing these areas together are explained in 1.3. The theoretical framework of the thesis draws on the neo-Firthian tradition and follows the developments along the Birmingham-Sydney axis. In particular, the *corpus methodology* elaborated by John Sinclair and colleagues (e.g. Sinclair 1987a) and the linguistic insights provided as a result of the *Cobuild* project (Sinclair 1987b) have shaped the present study. Similarly, Michael Halliday's work with Jim Martin (e.g. 1993) on *nominalization* and *metaphor* as linguistic processes has been significant in the treatment of delexicalised verbs here.

The broader issues of concern to language contact researchers and have seemed problematic to me in the course of this study can be summarised as follows:

1.1.1 Mixed code or not?

Two contrasting views can be identified with respect to the outcome of language contact. In one view the outcome can be identified in terms of the participating languages. In the other view, there is a mixed code developing. Researchers subscribing to the second view, however, rarely agree as to how the emerging 'mixed code' can be characterised. Further questions are inevitable here: why has the notion of 'mixed code' never matured enough? Has this been a methodological problem? If so, can corpus studies contribute a solution? Consequently, this thesis jumps on the "mixed code" bandwagon, too, but from a corpus perspective. Such an approach of course requires a major change in the way the lexicon is perceived, as proposed by recent corpus research. This brings us to the second question, namely the lexicon in language contact.

1.1.2 The place of the lexicon in language contact

Among other things, this thesis traces the way the lexicon has been perceived in Language Contact over the past five decades. While recent corpus research has elevated the status of the lexicon to more than an inventory of items, the new perspective on the lexicon has not been widely recognised in language contact so far. This is worrying especially when we consider the fact that what constitutes language contact data is mainly lexical. The question to be answered here is: if the bulk of language contact material is lexical, why do we search for an explanation elsewhere? Particularly, two major areas of investigation within the field, borrowing and codeswitching, have been dominated by what has been perceived as individual lexemes. Hence, the lexicon has seemingly attracted more attention, which, in itself, has suffered from a narrow view such as 'the repository of items'.

It is common practice to conclude a book or an article in language contact by acknowledging the frequency of nouns and the rarity of other grammatical classes. Yet, lexicogrammatical research shows that nominalization is not only a fact of language contact but it is a device commonly used in monolingual texts. My view is that the lexicon in Language Contact, after five decades of inferior status, deserves a better place, and this can be done only by paying closer attention to the

lexical relations in data. An examination of the current status of the so-called 'singly-occurring switches' is sufficient to validate this claim.

1.1.3 Singly-occurring switches

The corpus methodology proves particularly useful when it comes to analysing what has so far been variously called as 'lexical insertions', 'singly-occurring switches', 'gap-fillers' and so on. Corpus research recognises that words do not occur in isolation but co-select each other. The question here is: Can concepts such as co-selection used in corpus research help if it is acknowledged that collocations occur independently of grammatical structures, or word and sentence boundaries? The position of this thesis is that when we observe an L2 item in L1 context it is not a matter of insertion but co-selection. Obviously, co-selection tendencies will differ in the new context and an L2 item will not necessarily co-occur with its expected collocates in L1. A further point this thesis makes is that the mixed code starts when new co-occurrence patterns acquire fixed status.

1.1.4 Which elements facilitate mixing?

A recent issue that has arisen from corpus research has been delexicalisation. This notion is particularly useful in explaining the frequent use of certain verbs in Turkish such as *yapmak* awaiting a satisfactory explanation in Turkish grammars and dictionaries. The question here is: If mixing is facilitated by certain elements in language contact, what is the function of delexicalised verbs in that facilitation? There is evidence from a number of language pairs that they play an important role in the construction of mixed patterns in contact situations. Similar delexicalised verbs can be found in other cases of contact and this issue is dealt with in Chapter 8.

1.1.5 Borrowing or codeswitching?

The question of integration has never been fully resolved in the study of language contact and such criteria help little when dealing with the syntagmatic level. Frequency is a notion related to the distinction between borrowing and codeswitching, at least in the work of Myers-Scotton. However, as the following pages display, frequency does not always reliably show what is integrated into the language and what is not. There is another aspect of the data that indicates which items are in the process of integration.

Rather than trying to establish what is a borrowing and what is a switch, the argument here centers around how an L1 item co-selects an L2 item and whether this co-selection pattern displays regularity. This thesis therefore makes the claim that the borrowing-codeswitching-mixing distinction is not fruitful in explaining lexical relations in the data.

1.1.6 Morphological integration of words

A challenge for those of us who deal with an agglutinative language in contact situations is suffixation. We often come across statements such as "Lx nouns are morphologically integrated into Turkish" without further elaboration. This is undoubtedly so, however, such English noun+Turkish suffix, English adjective+Turkish suffix etc. patterns are significant both in terms of combinability and in terms of the mental lexicon of a Turkish-English bilingual. Rather than integration, the point here is, certain classes (e.g. functors) in agglutinative languages require different treatment from those of Indo-European languages. Consequently, these should be given more attention in diaspora Turkish.

In relation to this picture, the present study questions why the 'integration' of loans has been such a big issue in contact linguistics. This is done by contrasting the language contact researcher's perspective with that of the lexicographer.

1.1.7 Cultural borrowings

Cultural borrowings are often left out of data analysis simply because they carry the cultural load. However, the so-called cultural borrowings have been demonstrated to gradually lose their cultural content and become mainstream lexical items over time. What should be taken into account, however, is not the culture-boundedness of L2 items but the way they form relationships with the L1 environment. There is an interesting example of this in Chapter 8.

1.1.8 The question of 'psycholinguistic stress'

It has been claimed that bilinguals use strategies to remove the 'psycholinguistic stress' (Myers-Scotton 1993a) caused by the simultaneous processing of two languages in contact situations. This study concentrates, to some extent, on translation as a factor that leads to the use of such strategies. While it pays little attention to the dynamics between the community and the host society, it recognizes that it is the dynamics of the immigration context that creates the conditions of stress. It then searches for parallel stressful conditions under which professional translators operate. These conditions, as well as the strategies used and the resulting outcome have often led the scholars of translation to consider similar issues such as an emerging third code.

1.1.9 An integrated approach or not?

This thesis does not make an overt attempt to combine linguistic, social and psychological aspects of the language use within the Turkish community in Australia. The reason for this lies specifically in the methodology chosen. What becomes central in a corpus-driven study is the text rather than the individual or the society. While there is reference to the other aspects, the premise of the study is that the mixed code is self-organizing in the development of the patterns. That is, it cannot be explained with reference to solely social and psychological issues.

1.1.10 The nature of language contact data

Traditionally, spoken data have been central in the characterisation of language contact phenomena. Here the written mode is central. The sources that provide the empirical basis for this thesis are introduced in section 1.2 and explained in detail in Chapter 7. Since the corpus does not contain spoken data, the study is not in the conversational analysis tradition focusing on how and to what end participants use their repertoire of codes. Instead, it uses 1000 texts written in the immigrant setting and makes the claim that for a study of this kind they provide invaluable material. It thus criticises the superior status granted to speech and the neglect of written sources in language contact. This study, therefore, does not deal with prosodic features of the languages involved, nor with issues of bilingual speech production. However, it has implications for models of lexical access in bilinguals. The differences between the mental representations of the speakers of two typologically unrelated languages are considered. Previous research (e.g. Hankamer 1989) has shown that a Turkish speaker recognises suffixes as words, unlike an English speaker. How, then, does a Turkish speaker access English words? More importantly, how does a Turkish-English bilingual access mixed patterns?

Clearly, the data used here are mainly in two languages that have come into contact in the context of immigration in Australia: Turkish and English. The current state of Turkish in similar contexts in Europe and the history of the development of the current linguistic situation in Turkey are also examined briefly as part of the background to this study. While the current situation of Turkish elsewhere (e.g. in the former Soviet Union and the Balkans) is also of great interest to the research community, the scope of this study does not embrace such cases of Turkish contact.

I use capital initials (e.g. Language Contact) throughout the thesis to refer to the area of study and lower case (e.g. language contact) to the actual event. Also, I use terms such as *community*, *ethnic*, *migrant languages* as well as *Languages Other Than English* (LOTEs) and *Non-English Speaking Background* interchangeably.

1.2 Where does *DELIVERY YAPILIR* come from?

Throughout this thesis there will be frequent reference to phrases containing English and Turkish elements. Perhaps the most illustrative one from my point of view is *DELIVERY YAPILIR* (delivery can be arranged/lit. delivery is done). In this phrase, an English noun, *delivery* is used with the Turkish delexical verb *yapmak*, in the passive form. This phrase, like many others, comes from the Ozturk Corpus, a collection of 1000 machine-readable texts produced by and for the Turkish migrants in Australia between 1980 and 1995. The choices and decisions made at the time of compilation are explained in detail in Part 3. The texts that make up the corpus are of two types: newspaper texts and information leaflets.

From both a synchronic and a diachronic perspective, community newspapers in immigrant settings provide invaluable material for language contact research. In fact, these texts are a rare record¹ of naturally occurring language in immigrant settings across time. In terms of linguistic analysis, however, they remain underexploited. Such a study, of course, is and has been possible in Australia thanks to the abundance of Turkish Community newspapers published on a relatively regular basis over the past two decades. The newspaper situation is inevitably different in European countries with considerably larger Turkish communities, due to the factors such as geographical proximity to Turkey, and easily received satellite transmissions from Turkish radio and TV.

Information leaflets are similarly produced for the migrants in the migrant's original language but the provider in this case is the Australian government and various public offices, not the migrants themselves. These texts also remain underexploited although their nature suits language contact research given the issues of non-equivalence (rather than equivalence) between and across languages frequently debated in the area of Translation.

¹ Ethnic radio broadcasts could also be used for the purposes of such a project. However, due to technical problems SBS Turkish Programs Unit cannot build an archive of past programs (Mr Tanju Yenisey, personal communication).

The arguments deriving from the analysis of the *Ozturk Corpus* are also supported with a rich variety of examples from a monolingual corpus, *The Bank of English*. There are reasons for this. Firstly, behind the compilation of the Ozturk Corpus, lies the sound basis provided by two decades of corpus research carried out at the University of Birmingham (see Chapter 4 for an overview). Secondly, the lexical operations under investigation show similarities whether the corpus in hand has been compiled in monolingual or bilingual settings.

This study uses computational tools to investigate the regular patterns in written texts, difficult to establish manually on the basis of spoken data from individuals. It therefore questions the tradition that codeswitching involves two languages used in conversation as well as the reliance on grammaticalness and acceptability judgements of native speakers. This is again related to corpus research, which has shown that the native speaker's intuition frequently fails when it comes to naturalness criteria.

1.3 Why do language contact, corpus research and translation come together?

In the early days of Language Contact research, data from different parts of the world and in many different languages were lacking and the existing data collections were not readily accessible to everyone. These are at the research community's disposal now and more data become available everyday. Data analysis continues. Surprisingly, a satisfactory explanation as to the underlying mechanisms governing language contact phenomena has not been arrived at yet. Unfortunately, ongoing data production and data analysis do not solve the problems of the field. The problem, however, seems to be the way researchers have been looking at their corpus, rather than the corpus itself. Here, the approach to the data is corpus-driven in contrast to previous and current corpus-based approaches. In the data-driven approach the linguist investigates the corpus with an open mind to discover how language really works as opposed to the data-based approach where the linguist first establishes the model and then investigates the

corpus to find natural examples to fit into that model (see Clear *et al* 1996, Tognini-Bonelli 1996a). In the data-driven approach, intuitive knowledge of the language is not favoured:

While it is true that intuition and insight are valuable tools to language researchers, all assumptions must be tested against the facts of language as it is used. Linguists who spend hours examining their own intuitions of a particular pattern can easily construct examples which may be possible, but it is unlikely that they will actually occur. Those who work with language learners often face explanations in grammars, text books and dictionaries, which may seem obvious to an experienced speaker of the language, but are hardly supported by the facts of language as it is actually used (Hays 1997:12-13).

However, language contact researchers in the past generally approached their data to test the models operating elsewhere in linguistics. While this approach might have been successful with individual languages, it has proved problematic when the data in hand contained mixed elements from two different languages.

Corpus-based approaches to language contact are in line with that view of language that discards anything less than grammatical change as unworthy of study. Lexical analysis has been reduced to long lists of loan words and has not been explored any further. Surprisingly, taxonomic approaches to language contact data have remained in fashion for a few decades and the value of this approach has been questioned only relatively recently: "the study of loanwords out of context is a relic of the past" (Eastman 1992a:1). This results from a major shift of focus in the study of language in the last twenty years that has assigned the Lexicon a central role in many related areas of research (for example Sinclair 1987a, Levelt 1989). It was only appropriate, then, for Language Contact research to benefit from the recent developments in the related areas. In spite of some scattered attempts though, the place of the lexicon in language contact research, is still waiting to be defined. The ongoing preoccupation with singly-occurring switches shows that regardless of the theoretical claims in favour of lexical approaches, slot-and-filler view (for a criticism of this view see Sinclair

1991a:109-110) continues to prevail in the study of language contact (e.g. Silva-Corvalán 1995a).

The search for a unit of analysis has been a major concern in the study of language contact. While various units have been proposed for different levels of analysis, the sentence as the unit of analysis has remained central to most language contact research (e.g. intra- and intersentential codeswitching). The approach advocated in this study, on the other hand, is a lexical one and at the centre of the analysis stands the co-occurrence of lexemes from L1 and L2. Of special interest here, is how these elements co-select each other. The co-occurrence of word forms, the frequency of this co-occurrence and its statistical measurement is the basis of data-driven approach to language (Clear 1993). It is interesting to note that collocation as a unit of analysis was proposed by Firth in the 1950s, who was a contemporary of the first generation language contact researchers (e.g. Weinreich, Haugen, Ferguson, Mackey, see 3.3 for a review). However, Language Contact was established as an individual area of study on the other side of the Atlantic at a time when structuralism was in full swing. And the Firthian framework, not quite compatible with the Chomskian competence/ performance dichotomy, had not met acceptance even in Firth's own circle in the UK (see Palmer's introduction 1968, also Stubbs 1996:23-24), let alone in the US.

The development that has gradually pushed the lexicon to the centre of language research had also begun around the same time. It was the creation of the first machine-readable corpora, namely the Brown Corpus in the US in the early 1960s (see chap 4 for details). Nor the machine-readable corpus, however, readily made its way into language contact research, not for the first 3 decades at least. Generally speaking, however, language contact research has mainly relied on tape-recorded spoken data of limited size (see Dabène and Moore 1995 for a review) and searched for structural instances of change. This thesis is based on 1000 written texts and searches for lexical patterns of mixed nature, in line with the view that the main use of "a corpus is to identify what is central and typical in the language" (Sinclair 1991a:17).

This is done with reference to recent work on lexicogrammar that reconsiders the role of 'lexis' in linguistic description (e.g. Francis 1991, 1993). The main hypothesis is that grammar is largely determined by lexis and all lexical items have their own grammar. This view, inevitably, invites us to go below sentence level and revise the role of nouns in language contact research depicted as the most flexible class of loan material without much independence of their own. It is also advocated that in order to study a mixed code we should empty our lexicon first (Sinclair 1996a) rather than taking L1 and L2 grammars as a starting point.

The focus here is not what has so far been the centre of language contact research i.e. structural change. This is precisely because of the way the lexicon is perceived. There is no question of determining which language supplies the grammatical frame (as in Myers-Scotton's model 1993a). The frame as such is visibly Turkish. What is not immediately visible to the researcher is that the code establishes its own lexical patterning and that is the concern of this study. In particular, the lexicogrammatical approach adopted in this thesis throws light on the way single codeswitching forms work and where in a sentence a switch is possible, by proposing lexical constraints on switching. The constraints proposed have so far been syntactic (e.g. Pfaff 1979a, Poplack 1980; see Chap 3 for a review). The way singly occurring switches work and their location have been the major concerns in Language Contact since the earliest days (e.g. Hasselmo 1961). However, "corpus work has already called into question the way lexical and semantic studies have been dominated by single words" (Sinclair 1996a:101).

Likewise, the traditional framework has emphasised the centrality of verbal processes, and there has been no room for the grammar of nouns. More importantly, nominalization has always been treated as a language-contact-specific phenomenon. However, it is a common strategy used by monolinguals to increase the meaning potential in texts (Halliday and Martin 1993). In the process of nominalization delexicalized verbs prove to be particularly useful. Initial experiments on a limited number of language pairs (e.g. Spanish-English, Italian-English) lead to the observation that delexical verbs that do not perform the action but share the semantic load with the following noun phrase, are common across languages (e.g. Sinclair et al 1996). Consequently, a detailed study of their co-

occurrence tendencies in large corpora is essential for the new generation of language contact research as well as machine translation and lexicography.

Another insight provided by the written texts is the link between translation and language contact. While the analysis here concentrates on the authentic Turkish texts written by the migrants, it does not exclude another category of texts largely neglected in language contact research until present; information leaflets translated into the migrant's language. In fact, any selection of texts produced in an immigrant setting will prove that translation plays a crucial role in the daily life of migrants. Translation is, in fact, the link between the migrant and the host society. Socially and psychologically, the migrant is under constant pressure to translate, an aspect of immigrant settings that has not been given the attention it deserves. This study, therefore, is a link between Language Contact and Translation Studies. The questions central to Translation Studies (e.g. Baker 1992) have also been on the agenda of Language Contact researchers for a long time. Both areas are concerned mainly with lexical choice, and the range of lexical items actually used in relation to the range that could have been used in a particular situation. What is interesting here from the point of view of Language Contact research is the fact that in spite of their training, professional translators use the same strategies as the migrants in bilingual settings. This suggests that collaboration between these two areas of study would be beneficial to both. Scholars of Translation Studies, on the other hand, can gain from the expertise of Language Contact researchers in their search for bilingual description and the underlying mechanisms of bilinguals' lexical access.

This study makes a further link between these two areas and Corpus Research since both areas need to rely more on computational corpus research to compensate for traditionally intuitive data analysis. Furthermore, the similarities between the migrant's mixed code and the second language learner's interlanguage have implications for Second Language Acquisition research. Although publications connecting some of these areas do appear in the literature from time to time and joint conferences are organised (e.g. Gerver and Sinaiko 1978; House and Blum-Kulka 1986; Sinclair et al 1996; Jake 1998), these connections are rarely pursued in a systematic manner.

Within the framework outlined above, this study attempts to answer the following research questions:

1. Are there emerging patterns in the corpus?
2. What is the significance of these patterns in terms of frequency and distribution?
3. What are the motivations for such patterns in text?

1.4 The organization of the thesis

The issues outlined above are further dealt with in the subsequent chapters. This thesis is organized into five parts with each part containing a number of chapters. The chapters can be summarised as follows:

Chapter 2 is about the current state of the Turkish language in various Western European countries such as Germany, The Netherlands and so on, where a large number of Turkish migrant workers live. Their presence has stimulated considerable linguistic research and section 2.1 is a survey of the studies conducted on Turkish in Europe. The second part summarises the developments Turkish has undergone in Turkey since the Language Reform and considers the current situation. The chapter concludes with an attempt to link the state of Turkish in Europe and Turkey to the Australian context.

Chapter 3 offers a novel type of literature review of Language Contact research. Rather than analysing the language contact phenomena along a continuum as is usually done and treating these individually as transfer, borrowing, codeswitching, shift, loss and so on, I attempt an overview of the field of Language Contact in terms of five decades of research. The emphasis is on the development of the major areas of Language Contact since 1950, and how these developments have shaped the current picture.

Chapter 4 is a survey of computational corpus research and some areas of study where it has found immediate applications. This chapter is of particular significance in this thesis as computational corpus research is proposed as an alternative to the traditional data collection and analysis methods used by the language contact researcher. It is emphasized that the developments in the area of computational linguistics cannot be ignored if the study of Language Contact is to be furthered. This chapter also outlines in what way corpus research has contributed to the study of language through the work of John Sinclair and his colleagues.

Chapter 5 discusses the similarities between two areas of study that have not been traditionally related but have much in common: Translation studies and Language Contact. It also emphasizes the place of translation in the migrant's daily life, as well as the importance of translated texts for language contact research. The notion of equivalence that has been important for both areas of research is discussed and the idea of non-equivalence is explored. Importantly, a number of strategies, commonly developed and used by professional translators as well as bilingual migrants in the absence of equivalence, have been discussed.

Chapter 6 is a brief history of the Turkish migration to Australia and three decades of community life with special emphasis on how language needs are met and which community services are offered in Turkish. The section on language use summarises the answers provided by a number of community members involved in the Australian-Turkish media, to a questionnaire distributed as part of this study.

Chapter 7 describes the methodology adopted in this study in terms of text selection, corpus building, and the hardware and software involved. It emphasizes the experimental nature of this study since the community newspapers and information leaflets have never been used as corpus material for language contact research.

Chapter 8 deals with case studies based on the corpus data. It comes to the conclusion that a new description of the language contact phenomena is possible.

Concordances are displayed to show the co-occurrence patterns, and explain their significance rather than confirm preset hypotheses. The outcome is then interpreted with reference to an emerging mixed code and the relationship between translation and language contact.

Chapter 9 is a summary of findings in relation to a new unit of analysis. It proposes a number of lexical constraints on the language contact data with reference to recent corpus research outlined in chapter 4.

Chapter 10 summarises the implications of this study for Language Contact research in general and for Turkish in particular. It makes suggestions for future research and mother tongue teaching in immigrant settings. It also discusses the limitations of this thesis in terms of data collection and size, and finally draws conclusions on the basis of the theoretical and methodological framework in which the study has been carried out.

1.5 Summary

In this chapter the background and the scope of the present study have been outlined as well as the corpus, the central pillar of this thesis. The issues of broader focus have been summarised. This has been done in relation to three areas that have contributed to this study, namely Language Contact, Translation Studies and Corpus Linguistics. It has been emphasized that a lexical approach is suitable for language contact research and that the co-selection of L1 and L2 items should be in the centre of the analysis. The research questions have also been listed.

CHAPTER 2: Current State of Turkish

2.1 In and around Europe

Traditionally, research on migrants in Europe has focused on the acquisition of L2 rather than the migrant languages. Only relatively recently have projects in Britain (Stubbs 1985; Alladina and Edwards 1991), in Sweden (Boyd 1985), in France (Vermees 1988), and in the Netherlands (De Ruiter 1991) have dealt with a survey of the changing migrant languages. Similarly, research on migrant Turkish in Europe has been given some attention only over the last decade:

Through migration from Turkey and some Balkan countries, considerable Turkish-speaking diaspora groups have emerged in North Western Europe. These migrant speakers of Turkish have various backgrounds and do not constitute any uniform 'ethnic minorities'. Linguistically, they are, as a rule, living in unbalanced, asymmetrical contact situations, their first language fulfilling only community functions. Whereas several aspects of their acquisition and use of the dominant languages in the host societies have been studied, little attention has been given to Turkish as the dominated language. Only some preparatory work has been carried out in this field [...] (Johanson 1993).

The section below summarises that "preparatory work" conducted on Turkish in and around Europe in a number of countries where Turkish migrants live. In section 2.2, there is a brief overview of the Turkish Language Reform and its impact on today's Turkish. This chapter concludes with a section (2.3) that links the Australian context, in which the present study has been conducted, to the European and Turkish situation.

2.1.1 Germany

Due to the large number of Turkish speaking migrants in Germany (2.5 million) large scale research projects have been conducted in order to gather data on the bilingual development of Turkish children in German schools. The areas under

investigation in these studies (*Ausländerregelklassen* Study, SES, EKMAUS, KITA) have been the linguistic factors affecting L1 and L2 acquisition processes, the impact of L1 on L2 and vice versa, and the impact of the social conditions on the Turkish children's linguistic performance (Pfaff 1993). The speech of Turkish-dominant children was compared with that of German-dominant Turkish children and in spite of the differences between the structural inventories they employed, the German-dominant group also used the standard language and performed relatively error-free tasks. Researchers relate this to the regularity of Turkish morphosyntax (e.g. Pfaff 1988). The subjects who had frequent contact with German native speakers from birth used slightly less standard Turkish, interpreted as a sign of 'incipient language loss'. The low percentage of lexical mixing with regular Turkish case marking on German nouns, leads to the conclusion that the Turkish case system is not in danger of becoming reduced to fewer markers under pressure by L2. Turkish inflectional morphology appeared rather resistant to permeation by German but there was some evidence that 'erosion was starting to show in derivational morphology and syntactically complex structures' (Pfaff 1988, 1991). It is reported that instruction in Turkish is not widely available at the initial primary level in Germany and Turkish classes are attended by only a limited number of Turkish-background children (Pfaff 1991). Also in some German schools Turkish is not a compulsory subject for the Turkish background students whereas religion is (Özdemir and Poyrazođlu 1995). The maintenance of Turkish in Germany thus takes place mainly through informal discourse with Turkish adults and Turkish-speaking peers rather than through formal instruction. The social and educational framework in which Turkish is being learnt in Germany has also been the subject of a number of studies (Pfaff 1991, Röhr-Sendlmeier 1990, Kardam and Pfaff 1993). Since migrant workers from Turkey were initially accepted to Germany as 'transitory labor force', German educational policy and practices were not specifically designed to accommodate their needs, neither was the issue of mother-tongue instruction successfully addressed until the end of 1980s. In fact, researchers attribute the underdevelopment of Turkish lexicon and syntax, and the development of nonstandard Turkish and German in bilingual Turkish children's speech to the lack of a specific educational policy (Kardam and Pfaff 1993). Other studies include the problems of intercultural communication resulting from the limited German of Turkish migrant workers (Rehbein 1987),

and the contact between German and Turkish adolescents in West Berlin from a socio-psychological perspective (Apitzsch and Dittmar 1987).

2.1.2 The Netherlands

In the Netherlands, Turkish speaking migrants form the second biggest group of foreigners after the Surinamese, with a population of 205,898 (Kroon and Vallen 1995). Initially, research on Turkish migrants in the Netherlands focused on the acquisition of Dutch as L2 rather than the use of Turkish as L1 in an L2 environment. However, a growing interest in the acquisition, maintenance and change of Turkish in an immigrant setting has resulted in a number of projects in the Netherlands over the last decade (Extra and Verhoeven 1993a, b and c). Three main trends can be identified in research related to non-indigenous language varieties, especially Turkish and Moroccan-Arabic, in the Netherlands: 1. Study of these languages from the perspective of their potential contribution to second language learning in elementary schools (in late 1980s); 2. Study on the acquisition of these languages in a Dutch language submersion environment (early 1990s); 3. Study on the codeswitching behaviour of Turks and Moroccans (Extra and Verhoeven 1993b).

The Dutch government's programs for the teaching of minority languages and cultures had originally aimed at helping minority children "achieve or maintain a level of competence in their native language and culture which would enable them to return to school in their home country without too many problems" (Driessen 1992). However, in the early 1980s it became clear that the migrants were there to stay and the emphasis shifted from remigration to the retention of the native culture. Currently, government programs provide access to the native language and culture through a limited number of hours of instruction incorporated in the Dutch curriculum as well as a number of hours outside school. The effects of such programs on the acquisition of both Turkish and Dutch of Turkish and Moroccan children were investigated and positive effects were reported (Driessen 1992). In contrast with Germany, the degree of participation of Turkish children in such programs in the Netherlands is high (Extra and Verhoeven 1993b). Turkish bilingual children in the Netherlands were also compared with monolingual

children in Turkey and it was found that the second generation Turkish children in the Netherlands acquired fewer lexical items and morphosyntactic devices than their monolingual peers in Turkey (Verhoeven 1991, Schaufeli 1991). Studies on the codeswitching behaviour of Turkish children, adolescents and young adults in the Netherlands found that a "community vernacular" has developed characterized by a high amount of certain types of codeswitching (Backus 1992, 1993a and b, Backus and Van der Heijden 1994), and that this norm is gradually taking over. While the younger members use Turkish with the older members of the Turkish community, they switch to Dutch if Dutch speakers enter the scene, and use a kind of Dutch-Turkish with each other.

2.1.3 Scandinavia

Turkish is the language of everyday conversation amongst some 10,000 Turks currently living in Norway, particularly in major domains such as school, work, media, family and friends (Türker 1993). She concludes that lexical, phonological and semantic copying as well as switches from Norwegian can be observed in her informants' Turkish while the effect of Norwegian on Turkish syntax has been minimal.

In Denmark, out of a total of 11815 minority pupils, 4038 Turkish pupils form the biggest group of foreign students in Danish schools although Romaine points out that "in Denmark during the years 1975-78, not a single child of Turkish or Pakistani origin (the two largest minority groups in Denmark) finished secondary school" (1989:192). Research has mainly concentrated on the educational and social context in which Danish is acquired rather than linguistic aspects of Turkish migration. Turkish is used as the medium of instruction beside Danish in some experimental schools (Hetmar and Jørgensen 1993). However, the lack of qualified teachers continues to be a major problem. Gimbel's (1988) investigation of the impact of socio-economic background on the acquisition of Danish by minority children concludes that the Turkish pupils have reached the lowest educational achievement. He also reports on the acquisition of Danish word order in various different sentence types by minority children in Denmark where the Turkish speaking children have scored below the other groups (e.g. Serbo-Croatian

students). Jørgensen's (1993) study of the development of codeswitching behaviour of Turkish bilingual children in Denmark show that codeswitching is acquired as a power tool for purposes of social control at an early age.

Boyd (1985:32) states that in Sweden, immigrants from Turkey "represent the stereotype for immigrants in Sweden for many Swedes" along with Greeks and Finns. They "tend to live and work in close contact with one another" and , community institutions such as Saturday schools have also been organised independently from the state to support language maintenance (Boyd 1985:33).

2.1.4 U.K.

The number of Turkish speaking migrants in the U.K. is estimated between 80 - 100,000 with the majority being Cypriot Turks (Ali 1991). Small scale migration from Turkey to the U.K. took place when the rest of Europe closed the gates to Turkish workers in the 1970s. The Turkish speaking community in the U.K. has not been studied widely, however, Ali (1991) reports that codeswitching is more a behaviour of Turkish Cypriots who have been there longer than more recent arrivals from Turkey. The Linguistic Minorities Project (LMP) Adult Language Use Survey (ALUS) conducted in the early 1980s remains to be the only source of information on the use of Turkish in the U.K. However, according to the Linguistic Minorities in Britain report (Stubbs 1985:62), Greeks and Turks are treated together due to birthplace information in the census which, for most of them, is Cyprus. Therefore, it is difficult to obtain exact figures regarding the Turkish speakers in Britain. Turkish Cypriots have higher levels of English proficiency than mainland Turks. The results of the survey show that nearly all children have a receptive competence in Turkish and many use it in conversation with their parents and older members of the community. However, in the company of their siblings, they behave differently as 61 per cent were reported to use only or mostly English. Recently, there has been more emphasis on the instruction of Turkish at secondary level (Ali 1991) although the absence of Turkish from the curriculum remains to be a major problem.

The number of Turkish-speaking children in British schools is around 15,000, the second largest group after the Bengali (Sonyel 1988). However, the rate of underachievement has been high amongst Turkish children due to similar problems faced in other Western countries with large Turkish population. For example, 23 % of Turkish children were affected by the work situation of their parents, who were both unemployed. Turkish students and their parents interviewed by Sonyel (1988) attributed their failure in the British School system to a number of factors. These include the lack of recognition for the student's culture and religion in the schools, prejudice and racism, problems with English, and the failure of the schools to inform parents about the differences between British and Turkish education systems.

The studies mentioned above can be seen as "preparatory work" towards a description of Migrant Turkish in Europe (Johanson 1993). The researchers cited so far agree that Turkish in Europe is "starting to erode towards a less complex and less synthetic morphosyntax and a more restricted vocabulary" (Schaufeli 1991) and "the mother tongue development of Turkish children in Western European countries is weakened at an early stage" (Verhoeven 1991).

2.2 In Turkey

The Turkish language today cannot be described without reference to the Turkish Language Reform (henceforth TLR) and its consequences. The origins of the TLR can be seen in the efforts of the 19th century literary figures to simplify the language as a reaction to the diglossic situation that had developed in Ottoman times (e.g. Heyd 1954). The High variety of the language was a mixture of Turkish, Arabic and Persian used by the courts and the administration as well as the followers of the Divan literary genre and Turkish was the Low variety used by the people and the writers and poets of Halk Edebiyatı (people's literature).

A gap thus developed between the vernacular of the upper classes and that of the masses. The simpler, purer Turkish of the masses came to be looked down upon as coarser folk vernacular used by the Anatolians and Rumelians, whereas the

upper classes claimed using Osmanlıca or Lisan-ı Osmani (Ottoman Turkish). The lack of public education for the masses further separated the two vernaculars. It was considered a sign of culture to elaborate and decorate speech and to write with borrowed Arabic words and Persian literary styles.

The efforts to simplify the language in the pre-language reform period were relatively successful and by World War I most writers had begun to use simpler Turkish with less Arabic and Persian elements (Göđüđ 1995). As the Ottoman Empire collapsed (1918) and the new Republic of Turkey was founded (1923), language came to be seen as the central pillar of the nationalist ideology. A major source of dissatisfaction with the Arabic alphabet then in use was the lack of correspondence between the Arabic letters and the Turkish sounds resulting in the replacement of the Arabic alphabet with the Latin in 1928. The Turkish Linguistic Society (TDK) was formed in 1932 with the aim of eliminating foreign elements from Turkish. A number of strategies were established for the modernization and renovation of the Turkish lexicon: collecting words from pre-Islamic Turkish texts (tarama), collecting words from folk vernacular (derleme), deriving new words using Turkish roots and derivational morphemes (türetme), and compounding (birleđirme).

New words and expressions based on the processes above started to be published in the form of long lists in newspapers from 1934 onwards (Tietze 1962) and the TDK's *Türkçe Söđük* (the Turkish Dictionary) was also published in 1944 after long years of research. The transition from the Ottoman to the new Turkish lexicon was not a smooth process and created heated debate amongst the public as well as the politicians and literary figures. The Linguistic Society had been strongly supported by the Republican People's Party in power until the change of government in 1950. While it had met severe criticism, its activities had been considerably successful in the twenty years following its foundation. From 1950 to 1980 the language reform continued, though with varying degrees of success and the TDK had maintained its semi-official status. After the military coup in 1980 it was turned into a government organisation. For example, soon after the military coup in 1980, many government institutions including the Turkish Radio and Television (TRT) received a list of words (approx. 200) previously coined by

the TDK, and were advised not to use these (Özdemir 1995). Since then TDK's focus has shifted from purification to the production of scholarly works on the Turkish language.

Three main groups are identified with different perspectives on the activities of the TDK at the time:

1. Those radical purists in favour of eliminating all foreign rules and words from the language.
2. Those conservatives wishing to maintain the language without changes.
3. Those who argued for the elimination of foreign rules and lexical items that are "felt" to be alien, but maintaining words accepted and used by the people in everyday communication (Doğançay-Aktuna 1995).

Although the issue has been, to some extent, depoliticized, the main poles can still be identified today². So far the left wing has pushed the language reform and the purification forward and the right wing has favoured the Arabic and Persian loanwords and even made an effort to revive them. Accordingly, depending on one's political preference, the choice of words will be different in one's speech and writing (Cücelođlu and Slobin 1980). The language of daily newspapers can be taken as an example to this although little difference has been found between the languages of the left and right wing newspapers (Ýmer 1993). Interestingly, Ersöz (1995) reports that while 29% of the vocabulary used in modern Turkish today is of foreign origin, newspapers employ both the loanword and its proposed Turkish equivalent for 9.8% of the foreign vocabulary. While the choice of words remains mainly political, it can also be generational in that the older generation educated prior to the language reform, generally uses the old vocabulary (Hazai 1970, Tietze 1962).

Although a silent agreement seems to have been reached by the opposing parties on the fact that the language reform can no longer be reversed (Demircan and

² The insider's and the outsider's views do not always correspond on this issue (compare for example Boeschoten 1997 with Doğançay-Aktuna 1995).

Erözden 1991), it can be said that a certain amount of schizoglossia (Tietze 1962, Kurtböke 1996a and b) still exists in today's Turkish society. While a relatively small number of loanwords of Arabic and Persian origin remain as a sign of political affiliation, words of English origin have become another source of debate since 1980. While the members and the followers of the pre-1980 TDK continue to propose Turkish equivalents for English loanwords and expressions rapidly invading Turkish, in certain domains of use English terminology may be expected to gain more space in the near future.

2.3 The relevance of the European and Turkish situation to this study

Turkish in contact with European languages has been investigated, more extensively than in Australia, since migration from Turkey to European countries began a decade earlier. The common point of linguistic studies on Turkish migrants in different parts of Europe has been their emphasis on the acquisition of the local language, and language education rights of the Turkish children in Europe (see for example Skutnabb-Kangas 1984).

This study attempts to answer some of the questions surrounding the Turkish language in the diaspora, as its European predecessors. Thus, it shares the same ground with the studies in Europe. However, its difference lies in the fact that it investigates Turkish in contact with English in Australia where studies on this particular pair of languages are rare and with a different focus (e.g. Yađmur 1993, 1997, Peköz 1993). Perhaps more prominent in the present study are the linguistic characteristics of the contact rather than the social functions of Australian Turkish in the community. The language investigated here is used exclusively by the Turkish Community members to communicate with each other in Australia and not by Turkish speakers to communicate with English speakers. The majority of the texts that make up the corpus come from the writings of the first generation Turkish migrants in Australia. It is, therefore, an investigation of the first generation's language use rather than that of the second or the third. In this respect this study looks different from the ones in Europe as regards the nature of its data.

But there is at least one striking similarity emerging out of the European and Australian data: the frequent use of delexicalised verbs (e.g. *yapmak*) by the Turkish migrants when they switch codes. This issue is dealt with in chapter 8.

As for the relevance of the Turkish situation to this study, the following argument can be put forward. The influence of English on today's Turkish in Turkey has been mentioned above. While a systematic study needs to be carried out, it is obvious that the English words borrowed into mainland Turkish are different from the ones borrowed into Australian Turkish. At the time of writing, I had no access to a Turkish corpus in Turkey for a comparison, so this question remains open for future study. One minor question at the beginning of this study was whether the politically-determined use of language had been carried over to Australia and whether the texts selected had some differences in terms of vocabulary use. There is a brief discussion of this in section 9.4. From a different perspective, the future of the Turkish language in Australia depends on the links between Turkey and Australia and the direction of the language planning activities in both countries in the 21st century. However, given the current strong link favoured with Asian countries, it seems unlikely that Turkish will remain as a major community language in Australia in the years to come.

2.4 Summary

This chapter has considered the state of the Turkish language in and outside Turkey. Turkish has undergone major changes in Turkey due to the language reform; and it has been also subjected to outside influences as a result of four decades of migration from Turkey to western countries. These changes have been summarised and the relevance of the linguistic situation in Turkey and in Europe to the situation in Australia has been highlighted.

CHAPTER 3: Language Contact research - an overview

Geographical and/or social proximity between communities results in mutual influence between languages and dialects. Such situations are known as language contact. According to Braunmüller (1997) contact situations produce four different ways of communication that can be summarised as follows:

1. The dominant section of the speakers establish their own language as the official language and the inferior language may survive as a restricted variety.
2. In trading situations, a third language, a lingua franca or a pidgin is used.
3. When the languages involved are genetically related the population becomes passively bilingual.
4. The population may become bilingual.

The context of immigration investigated in this study is one of the ways in which languages come into contact. Being confronted with speakers of another language in the host country means new options of communication for the migrant community. These options, not necessarily very different from the ones outlined above (Braunmüller 1997), have been described by Fase, Jaspaert and Kroon (1992:4-5):

1. The migrant community can choose not to communicate with the monolingual speakers of the host country.
2. The migrant community can try to establish communication in its own language.
3. The two groups can choose to communicate in a third language.
4. The migrant community can try to communicate in the language of the host country.

On the basis of previous and present research one more option should be added to this list as part of (3) above. Semi-communication may develop as a result of interaction efforts between the host country and the migrant group. Depending on the group's choice from these options then, new linguistic situations emerge in immigrant settings. Different tendencies can be identified with respect to the

description of these situations dating back to the early debate on whether linguistic analysis can and should be separated from non-linguistic analysis (e.g. Pike 1967:33). One group of researchers hold that defining these situations in purely linguistic terms at the expense of social and psychological factors would not work, so they favour a unified approach (e.g. Lüdi 1992). Then, there are those who recognize the importance of psychological and socio-cultural considerations but what matters for them is the restructuring of grammar/s as a result of contact. For example, Muysken and associates (e.g. Van Hout and Muysken 1994) seem to work along these lines, although they do not explicitly say so. A third group, instead, emphasizes the social and/or psychological factors impacting on contact where linguistic aspects remain secondary. For example Gillian Sankoff finds it puzzling that "many linguists still find it difficult to see how [...] the social world is relevant to the internal structure of language" (1980:xix). The strongest support for this view came from Fishman (e.g. 1968a:7) during the developmental stages of the field:

[L]anguage "per se", in the form of a corpus of sounds and smaller or larger units of meaning, has been examined for its patterns, as if it were something that existed above and beyond its users and its uses. Psychologizing and sociologizing have not only been ignored (as leading in "exolingistic" directions) but have been attacked in former years [...] as dangerous and misleading pursuits.

Naturally, these tendencies, all perfectly legitimate, depend on the researcher's main area of expertise, usually one or a combination of the fields listed in section 3.1. A note of clarification is necessary here. The boundaries of these fields were not well-defined before mid-1940s. Hymes defines "the years from about 1942 to 1965" in the U.S. as a period of "academic expansion and of cold war" (1972:245):

Linguistics was institutionalized in separate departments for the first time and grew greatly in numbers and influence. The relation between linguistics and anthropology, once so obvious and intimate, came to seem a problem, as methodological developments absorbed the attention of linguists and made their work and talk esoteric to many anthropologists.

Consequently, the hope that "an anthropological linguistics in the years after World War II" would "enrich linguistics with anthropology" did not materialize and so came the division (Hymes and Fought 1975:74). During these years, also Fishman (1968a:7) notes that "[i]n contrast to the mainstream of American linguistics, linguists with strong socio-cultural interests [...] represented a smaller parallel tradition, usually under the rubric of anthropological linguistics". Initially, Sankoff (1980:xvii) confirms that "the goals of sociolinguistics were thought to be modest, i.e., the investigation of the social circumstances of the use of language". But gradually a transition took place from the "social context of language use to [...] how the social nature of language use may constrain, influence, and shape language structure" (Sankoff 1980 xvii-xviii). Also when Fishman (1968a:8) claimed new territory for what he named the 'sociology of language' involving the study of multilingualism, language maintenance, language shift, language standardization, language planning and so on, there was no welcome either from the sociologists or the linguists (1968a:8-9). This was a period when even the meaning of 'sociolinguistics' was not clear, let alone the sociology of language (Fishman 1972a). This debate brought about a further distinction, that of macro and micro levels of analysis and the question of how two combine them. Macro studies of language contact have addressed language processes of uniformization, creolization, standardization and so on; whereas, "most studies of social processes related to language contact (i.e., repertoire selection and codeswitching) have been limited to the micro level" (Grimshaw 1987a).

3.1 Where does language contact belong?

Today, there is an overwhelming amount of input coming into language contact research from different directions. Familiarity with these would help understand the current picture and how such crossfertilization has developed. Collaboration among fields was encouraged early (e.g. Haugen 1956:9):

It is important that linguists learn about the contributions made to the field by psychologists, sociologists and educators. But it is equally important that experts in the latter fields learn about the body of linguistic knowledge available today.

Undoubtedly, "coordinated efforts of all the disciplines interested in the problems" would be beneficial for all (Weinreich 1953:5) but the question of how these should be coordinated was not clear. Participation from different fields has indeed enlarged the scope of language contact research but the question of how to combine contributions from different fields into a single framework or theory, in spite of the attempts made by the subsequent generation of researchers, remains unanswered. As De Beaugrande (1997:279) points out "the relation between theory and practice can be a difficult issue in nearly all domains of human activity" including linguistics. This is because human practices usually precede theories and thus decide what sort of theories should be constructed. Since the practices brought about by language contact are multidimensional, it is only to be expected that so would be the theories. However, without a given recipe, it was no surprise that an interdisciplinary framework could not easily emerge. In fact, collaboration among different fields did not take off immediately. Each of these fields, Fishman wrote a decade later (1968b:21),

has normally conducted its work in isolation from the other. As a result, not only have the methods stemming from one discipline been unexamined by investigators associated with the others but there has been insufficient concern for integrating the particular aspects of reality that each of these disciplines recognizes into a single, inclusive theory of bilingual behaviour.

After some 12 years, Hymes (1980:x) noted that the initial enthusiasm to share a framework had weakened:

The congeries of interests that coalesced in the 1960s around the goal of a sustained social study of language have tended to separate out again. In arguing for the social study of language, each had its specific opponent, its specific disciplinary world to conquer. For some, it was conventional sociology, for some conventional linguistics, for others philosophy, for still others anthropology, or some combination of these. The impulse to band together depended on a sense of marginality in a home discipline. Achieved legitimacy has weakened the impulse.

Also, what Appel and Muysken wrote, two decades later, shows that the idea of a unified approach was still struggling (1987:8):

Distinguishing between sociological, psychological, sociolinguistic and linguistic contributions to the study of bilingualism and language contact is in many ways unsatisfactory and artificial, because they are so intricately interrelated. It is impossible to study a psychological topic such as the cognitive consequences of individual bilingualism without taking social factors into account such as the relative status of the languages involved.

Today there is no question as to the importance of different kinds of input from different fields but feeding multi-level practices into a single theory still proves to be problematic in language contact. The problem may be summarised in terms of methodology and terminology.

While methodologies differ, some generalization can still be made. In a recent article De Beaugrande (1997:280) has distinguished between homework linguists (Chomsky and followers) and fieldwork linguists. In De Beaugrande's sense, language contact researchers can be identified with the latter as they

go out to work in the field of cultural and social activities and carefully record and describe what [...] speakers are actually observed to say. Their theories are both data-driven insofar as observation and induction exert prominent control, and practice-driven insofar as fieldwork must join in the social practices of the community of speakers in order to gather and interpret the data.

Except for those working on historical aspects of language contact who deal with historical material, and those in psycholinguistics who rely on laboratory experiments, language contact researchers work in the way De Beaugrande's (1997:281) fieldworkers do:

1. The fieldworker [...] tries to learn from a gradually accumulating circle of contacts and acquaintances, who represent the community's language from multiple perspectives.

2. Fieldworkers [...] assemble their own systematic data corpus.
3. The fieldworker [has] an intense and round-the-clock commitment to the task.
4. The fieldworker has received highly specialized training skills for writing down utterances in a reliable phonetic alphabet or for distilling out regular patterns from the data that might correspond to categories like 'nouns' and 'verbs'.
5. The fieldworker is operating under conditions of total cultural immersion. Language is being observed in relation to human activities that provide both important clues as to what certain utterances mean and a pervasive sensitivity to the general cultural setting of the target language, whether or not such factors are addressed in the fieldworker's official linguistic theory.
6. The fieldworker is conversant with explicit theoretical concepts about language, directed partly toward the native language and partly toward experience with other languages during a period of pre-fieldwork training.
7. The fieldworker's results are intended for presentation to a scientific community of other experts, some of whom may re-examine the data and compare it to the theoretical statements made about the language.

These principles were established, long before the systematization of Language Contact, within the anthropological tradition in the U.S. pioneered by Boas and his followers (see Hymes and Fought 1975). While data-collection procedures are similar, differing foci of research in the major participating areas in relation to language contact can be summarised as follows:

Descriptive linguistics - investigates what happens to the structure of languages in contact and how new varieties emerging out of contact situations can be described. Depending on the researcher's theoretical inclination, the data can be used either as a test bed to confirm a pre-established theory or as a springboard to develop new models.

Sociolinguistics - investigates the role of social stimuli (e.g. cultural pressure) and the speaker's sociolinguistic background in the explanation of language contact phenomena.

Psycholinguistics - investigates bilingual behaviour as part of the psychological make-up of the individual who has access to two linguistic codes. Various techniques for the measurement and description of a bilingual's use of two languages have been drawn from experimental psychology such as picture naming, word association, word frequency rating and so on.

Anthropological linguistics - investigates language contact as a way in which language use reflects social change and cultural values. A significant contribution from here to language contact is the use of ethnographic method that involves a detailed analysis of speech in context.

Historical linguistics - investigates how languages change internally across time through processes such as borrowing and substratum interference with or without taking external factors into account.

The object of language contact is also of interest to other fields such as social psychology, neurolinguistics, political science, demography and so on. These fields have contributed to the study of language contact not only in terms of research techniques but also in terms of new perspectives.

The question to be addressed at this point is where this thesis belongs. It is descriptive in nature although it is driven by a corpus rather than based on it. This is an important difference since the corpus acts as the major informant as explained in detail in chapter 7. It also has a (relatively short) historical dimension and makes references to social and psychological issues when necessary although these aspects remain secondary throughout and there is no overt attempt to unify these dimensions.

3.2 Five decades of Language Contact Research

Obviously, many questions in relation to contact situations had been raised long before 1950s. In fact, Schuchardt's work in late 19th century, as well as Ronjat's (1913) and Leopold's (1939-49) work early this century were important studies of language contact. However, the direction of the field was properly set in 1950s as a number of influential works marked the turning point and the study of Language Contact has been more systematic and well-recognised since then. Until 1950s, "most previous studies in bilingualism ha[d] concentrated on static bilingual areas" such as border communities or enclaves where "bilingualism [had] already been established and maintained for some generations rather than of migrant communities" (Clyne 1967:10). This type of stable bilingualism differed from the dynamic bilingual areas found in the case of for example recent migrants in Australian cities where closed settlements as in the Americas hardly existed. Rather, with recent migrants there was a constant dynamic two-way process of change involving both languages. This thesis was also born out of a bilingual situation of the dynamic type.

The literature review that follows goes differently from the usual formula. I survey the central issues as voiced by the predecessors of Language Contact rather than refer to only the most recent work. The reason for this is my conviction that most of us get the early decades second-hand. Perhaps this is inevitable in a field so saturated with the ideas of the pioneers but to find out what exactly they said would help us understand how exactly the central issues of today's Language Contact have developed into what they are now, particularly those relevant to this thesis. While surveying Language Contact research, my aim has been to stay as close as possible to the original words of the researchers. As De Beaugrande (1991:1) points out this strategy is "particularly instrumental in tracing the development of terminology, and the continuity, evolution, or change in the major lines of argument not merely between theorists, but within the work of an individual theorist".

This section then is a survey of language contact research in terms of five decades starting from 1950. A number of issues were central in the first decade, and these have been followed up by the subsequent generations of language contact researchers. The literature review here is an attempt to trace the development of concepts and approaches around these issues that helped shape today's picture contrary to the conventional approach dealing with language contact in terms of 'phenomena' usually investigated individually as *codeswitching*, *borrowing*, *attrition*, *shift*, *death* and so on. In other words, the approach in this case is vertical rather than horizontal and the idea comes from computer research where the machines developed in each decade are referred to as first generation computers, second generation computers and so on. However, this analogy with computer research does not go beyond the surface level since in computer research each new generation of machines has replaced the previous generation. Such a dramatic development has not taken place in Language Contact.

Another reason for this approach is the difficulty of defining the so-called phenomena as there is considerable overlap among these and the contribution of different areas does not necessarily clarify the picture. In fact, I argue in this thesis that regardless of the previously established labels such as *borrowing*, *switching*, *mixing* and so on, the outcome of language contact can be studied as a single process of co-selection where language is seen as a self-organizing and pattern-making system.

Here the emphasis is on the prevailing trends within and outside Language Contact in each decade rather than the development of the work of individual researchers. It goes without saying that the work of most language contact researchers span across a few decades and a number of these had been writing on this subject before the 1950s (e.g. Haugen 1938). Occasionally, there are references to the other decades and to current work on language contact and this is done in order to show the link of early work to the present. Throughout the sections 3.3-3.7 the emphasis remains on the key concepts. The numerical sequence in each decade refers to the same concepts in order to facilitate cross-referencing. The sequence is broadly as follows:

1. bilingualism and language contact
2. interference, integration, different levels of analysis
3. borrowing, resistance to borrowing, predictability of what can be borrowed
4. switching, switching vs borrowing distinction, nonce-borrowing, location of switches
5. mixing, mixture, code-mixing, mixed-code etc
6. languages with similar and different typology, equivalence, contrastive analysis
7. language change, restriction, shift, loss, convergence vs. divergence
8. the place of the lexicon in language contact, classification of loans, frequency, domains
9. methodological issues
10. summary of the decade

Although the key concepts are highlighted, subheadings are deliberately avoided in this chapter. This strategy makes it obvious that the overlap between concepts and approaches is omnipresent and another attempt to force further boundaries would only make things more complicated. Similarly, I avoid the individual definitions of the so-called "phenomena" and the simple reason being from whichever angle we look at the phenomena, they refer to the use of L1 and L2 elements in the same environment. Consequently, in each subsection, the key concept is highlighted and acts as a magnet attracting a few other related points. The idea of surveying the literature in terms of five decades is simply a matter of convenience, and researchers with different perspectives would undoubtedly perceive different time periods in language contact research.

3.3 First Decade: 1950-1960

In the field of Language Contact there has been considerable reshuffling of approaches and frameworks and rewording of the issues and problems. This has stimulated a constant shift of focus from one aspect of the field to the other (e.g. definition and measurement of bilingualism has lost prominence while a topic such as codeswitching has gained popularity). However, the issues discussed in the first decade remain the same. Undeniably, the impact of this decade determined the

direction of language contact research for the subsequent generations. In fact, today's taken-for-granted notions such as *equivalence*, *restriction*, *code* and so on, which are critically examined in this thesis, date back to the first decade.

3.3.1 As the bilingual individual came to be seen at the centre of language contact it was inevitable that any reference to language contact would contain reference to some aspect of bilingualism. Haugen confirmed the emergence of bilingualism as a separate area of study (1956:3):

When I started collecting bilingual bibliography, it was in order to gain a wider perspective of the problem I had undertaken to solve in analyzing the linguistic experiences of the Norwegians in America. But **bilingualism grew into a problem of its own**, and the writings on the subject proved to be mountainous. [...] The subject was reduced to something like **scientific order** by Uriel Weinreich's *Languages in Contact* in 1953 [emphasis added].

Nevertheless, both Haugen and Weinreich treated bilingualism and language contact together in their respective publications and their boundaries were not very clearcut. For example, Haugen put the following questions on the baking tray all at once (1953:10):

- is there a correlation between the amount of bilingualism and the amount of confusion?
- are both languages equally affected, or is the influence likely to be one-sided?
- are the influences permanent or evanescent?
- under what circumstances can language mixture really disturb the essential core or basic patterns of a language?
- why are some forms adopted and not others?
- is there such a thing as a 'hybrid' or 'mixed' language?
- what is the dilemma bilinguals face, and how can he most conveniently make his peace with that dilemma?

In these questions we can identify the structural and psycholinguistic concerns at the same time and the origins of the *process/product* distinction in language

contact. However, Haugen (1953:383) did not seem totally convinced about the psycholinguistic line of research as "we [were] unable to disentangle the psychological factors" at work or "watch directly" "the mental processes" involved. We could neither take the "remarks people make about their own mental experiences" as these were "generally inaccurate" and could not be "accepted as completely valid evidence". So we just had to observe the bilingual behaviour and "adopt a purely linguistic point of view" (1953:388).

During this decade, the definition of the bilingual individual moved away from the Bloomfieldian requirement of native-like control of two languages and the prevailing negative attitude to bilingualism was criticised:

The question of whether bilingualism is a Good Thing or a Bad Thing has occupied the thoughts of many writers on the subject. Even when these terms are not used, there is implicit in many writings an emotional attitude of justifying or attacking the practice. Much of the research devoted to show that it is a handicap appears to be based on an unconscious bias against bilingualism as such, at least among immigrant groups (Haugen 1956:116).

While the claim now was a more positive approach to the status of the bilingual, some of the language used to describe bilingual behaviour makes one wonder if this really was the case. That is, there was frequent use of such expressions as *deviation* (Weinreich 1953:1), *interference* (which, in itself, had negative connotations although adjustments to this term were made in the subsequent decades) (Weinreich 1953:1; Haugen 1956:39), *distortion* (Weinreich 1953:21; Haugen 1956:55), *inadequacies of the bilingual's lexicon* (Weinreich 1953:31), *confusion* (Haugen 1953:53), *dilemma bilinguals face* (Haugen 1953:1), *confounding of patterns* (Haugen 1953:10), *pattern disturbances, complexities involved in the use of two languages* (Mackey 1956); *conflicting sets of linguistic habits* (Martinet 1953:viii) and so on. Although this is not the place to discuss language contact rhetoric, the expressions above indicate, to a certain extent, that their focus had not sharpened yet.

For Haugen (1956:13) the questions to be answered by language contact researchers were as follows [Emphasis added]:

1. What happens when **a second language** is being **learned** and when it is being used?
2. How does the individual learn **to control the responses of each language and keep them apart?**
- 3a. Do speakers actually have **two distinct codes** or are they to some extent coalesced,
b. and what happens when the speaker **switches** from one to the other?
4. What **effect** does **switching** have on the individual in terms of clarity of thought, integration of personality, and life goals?

These questions played an important role in the development of Language Contact in that each has turned into a major area of investigation itself. The first question stimulated research into the measurement of proficiency, and found immediate applications in the area of language teaching. An evaluation of this, however, is not within the scope of this thesis.

The second and third questions occupied the Language Contact researchers working in the newly-emerging field of the time: *Psycholinguistics*. Here, the dichotomy of *Compound* vs. *Coordinate* bilinguals was introduced. It was about whether the bilingual's two languages merged into a single system or co-existed (Weinreich 1953:9-10).

Weinreich was mainly concerned with the relationships between words and their referents and how bilinguals interpreted these relationships. This distinction was soon interpreted by Ervin and Osgood (1954) in terms of the cognitive organisation of individual bilinguals and the context in which the two languages were acquired, and "has stayed with us through the empirical studies of compound and coordinate bilingualism" (Hakuta 1986:101). According to Hakuta, it is the

emphasis on individual differences that led to confusion. Instead, researchers should have looked for "concepts that are organized in either coordinate or a compound fashion within bilingual minds" (Hakuta 1986:101). It was also the reinterpretation of this notion in terms of acquisition context that reinforced the confusion and consequently led to its disuse (e.g. Lambert et al 1958). Since this decade dichotomies have been particularly popular in bilingual description. The distinction between *individual* and *societal bilingualism* (Weinreich 1953:3,83) and *adult* and *childhood bilingualism* (Haugen 1953:371) was also made at this time.

The third question has led to the development of *Codeswitching* as a separate area of study not only for researchers with psycholinguistic concerns but also for others operating in the areas outlined above (see 3.1 Where does language contact belong?). A relevant issue that has stayed with us since this decade, has been the *location of switches*. According to Haugen "a switch is normally made at lexeme boundaries, and it need not embrace more than a single lexeme" (1956:50). In fact, most of Haugen's and Weinreich's examples were single switches analysed out of context. This may be due to the fact that phonemic aspects of language contact were given priority at that time (e.g Sapir 1921:197-202, 1933; Bloomfield 1933:138; Pike and Fries 1949). The location of switches has been much discussed since then. The fourth question has also been studied within the domain of codeswitching and there was no surprise as to the result of the enquiry, then and ever since: switching has positive effects on the bilingual individual (Weinreich 1953:121).

This decade closed with Ferguson's Diglossia (1959). He focused on the functional distribution of codes in speech communities, that is the roles language varieties (e.g. the superposed variety used in formal education and the native language used in ordinary conversation) played in the society and the attitudes of the speakers to these. Although Ferguson's model of diglossia (the high/low model) was a theoretical construct based on "the few available descriptions of the distribution of language functions in plurilingual communities" at the time (Mackey 1993), it stimulated research into group bilingualism and the consecutive development of other models of diglossia.

3.3.2 True, bilinguals had been freed from the tight boundaries of the Bloomfieldian definition but they were now facing new charges, those of *interference*. In fact, Haugen saw the linguist's task as "to identify and describe all cases of interference and then to cooperate with other social scientists in accounting for them" (1956:11). The broader aim of interference studies since then has been to investigate "the mechanisms of interference" that "would appear to be the same whether the contact is between" two similar or unrelated languages (Weinreich 1953:1).

The effort to describe what exactly constituted interference resulted in forcing boundaries on language contact data. This was the point where labels such as switching³, integration, established loan sprung into existence (Haugen 1956:40). This was also the time when the appropriateness of terms previously proposed by American anthropologists, such as language mixture, hybrid language and borrowing was discussed (Haugen 1950, 1956:40).

Equally important, then, for the early researchers as well as their followers in the subsequent decades was the *integration of loans*. At one end of the scale was the complete integration of loans with "no trace of foreign material left", and at the other end, the loanword with "a residue of the model" (Haugen 1956:56). The Prague School's idea of the *continuum* was beginning to gain ground and would be applied to different aspects of language contact from then on. In fact, today's popular continuum perspective on language contact starting with code-switching to language shift with borrowing in between, leading eventually to language death (e.g. Silva-Corvalán 1994) belongs to this decade (Weinreich 1953:69):

It remains to be determined empirically whether habitual switching of this type represents a transitional stage in the shift from the regular use of one language to the regular use of the other.

³ Attributed to Jakobson et al 1952 (see Weinreich 1953:7; Pike 1967:595)

Haugen observed that "[l]oans, which are accepted from bilinguals into a monolingual community, will pass through various stages of integration until they achieve their commonly accepted form (1956:55-56). What concerned them was the phonetic side of the coin, of course. This question has kept researchers busy for a long time and only recently it has been challenged: complete assimilation seldom occurs (Myers-Scotton 1993a:21). Long before, Sapir had also argued along the same lines "Such an English word as the recently introduced camouflage, as now ordinarily pronounced, corresponds to the typical usage of neither English nor French" (1921:197).

It was also at this time that the idea of distinguishing levels of analysis as phonology, morphology, syntax and lexicon and their further subdivisions gained popularity (Weinreich 1953:29; Haugen 1956:41, Mackey 1956). For example, for Weinreich (1953:1), *interference* meant the rearrangement of patterns in the phonemic as well as in *morphological, syntactic* and *lexical systems* of a language. However, the researchers themselves were well aware that the boundaries between these divisions were not clear:

To this day, there is little uniformity in the drawing of lines between morphology and syntax, grammar and lexicon. But this need not be an obstacle to a systematic analysis of grammatical interference (Weinreich 1953:29).

3.3.3 The interest in borrowing, Haugen complained, had never been primary amongst linguists. Consequently, he set out to discover "the basic laws of borrowing" (1953:363) at structural and societal level. At the societal level, Haugen discussed the bilingual norms (1953:60-62), as part of his search for the laws pertinent to bilingual communities in terms of the adoption of foreign material or resistance to it. He underlined the power of group pressure in controlling the too rapid advancement of foreign material into the language (Haugen 1953:71). Haugen's (1950) early work on borrowing had already given it special status in the study of language contact and when he introduced *switch* "to designate a clean break between the use of one language and the other" (1953:65), he clearly distinguished between the two:

Switching is different from borrowing, in that the two languages are not superimposed, but follow one another.

It was clear that Haugen (1953:384) considered his definition of borrowing as a historical process, which required that

the linguist **prove** for **each borrowing** that it **first existed in language A** and only after the first contact of A with B did it appear in language B. In studying AmN borrowings we must show then (a) that any given item was used in the English heard by the immigrants; (b) that it was not previously used in the Norwegian they knew; and (c) that they could not have made it up independently. [emphasis added].

But how was the linguist to prove these? Here, Haugen made a statement, and it would be fair to say that it served as the basis for the ongoing war between borrowing and codeswitching: "Since we cannot actually observe the entry of borrowings, we must be content to make certain assumptions of high probability" (1953:384). In other words, despite his emphasis on the historical aspect, he acknowledged that the entry point of a borrowing remained open to speculation.

Simultaneously, came the myth of *nonce-borrowing*, that is when a loan was first uttered by the bilingual (Weinreich 1953:11). Although this term remained dormant for a while after its first use, it stormed back into the third decade and has been the subject of heated debate since. This also fitted with the distinction between *synchronic* and *diachronic* dimensions of linguistic research as well as the influential *langue* and *parole* distinction (e.g. Mackey 1967). These researchers insisted that the two phases of interference should be distinguished (Weinreich 1953:11):

In speech, it occurs anew in the utterances of the bilingual speaker as a result of his personal knowledge of the other tongue. In language we find interference phenomena which, having frequently occurred in the speech of bilinguals, have become habitualized and established.

Saussure's synchronic vs diachronic distinction is to blame if we are now so confused about borrowing vs codeswitching distinction. It is at this point in time that the difference becomes blurred and we cannot work out which is which. Theoretically, as has been traditionally done, it should be possible to attribute borrowing to diachrony and codeswitching to synchrony but this is not the only point of confusion. It should be possible, also structurally, to distinguish between the two if we are to follow the differing structural definitions Haugen proposed for the two. Neither has been.

Another area of inquiry subsequently developed from the work in this decade was 'resistance to borrowing'. This question was already investigated earlier (e.g. Sapir 1921:194) although the emphasis had been on cultural and psychological aspects. Now, it was the time to question the structural aspects:

Is there [...] a structural, or purely linguistic resistance to borrowing?
(Haugen 1956:66; also Weinreich 1953:61).

to what extent interference is determined by the structure of two
languages in contact (Weinreich 1953:4).

At the time of Haugen's writing, previous literature suggested that the answer to this question was dependent on the language involved. What interested him and his contemporaries however was "the degree of boundness or independence of the linguistic items" in relation to linguistic resistance to borrowing (Haugen 1956:66). Their view has been dominant ever since:

"Full" words, which can occur either as parts of utterances, or as complete utterances, are rather easily borrowed. Function words, which normally occur only as parts of utterances, are seldom borrowed. Bound morphemes, whose independent meaning does not become apparent except by comparison of utterances (e.g., derivatives, or plural suffixes), are seldom if ever borrowed (Haugen 1956:66-67).

However, all this was not enough to predict that a given word would or would not be borrowed (Haugen 1953:97) and which "potential forms of interference [would]

actually materialize" (Weinreich 1953:3). Predictability, Haugen felt, could not be accounted for in terms of necessity as "there was no strict need for Norwegian immigrants to adopt the English words for 'river', 'field', 'fence', 'barn', 'cousin', or 'pail'. They had adequate Norwegian words for all of these objects"(1953:373-4). The search for predictability criteria still continues (Myers-Scotton 1993a:3).

3.4 Second decade: 1960-1970

The preceding decade had introduced language contact "as a social phenomenon, in such a way that it might offer something of value to students of history and sociology" (Haugen 1953:xi), and it did. Consequently, the decade 1960-1970 saw a sociolinguistic boom from which there was no going back. Later, in an overview of this decade, it was confirmed that in these years the most significant contribution to the study of language contact had come from Sociolinguistics.

From this decade on, language would be seen as an index of social processes (Labov 1968:240) "in which utterances are selected in accordance with socially recognized norms and expectations" (Gumperz 1968:381). Linguistic analysis meant the study of usage, a reflection of "more general behaviour norms", "within a socially defined universe" (Gumperz 1968:381). And it was this "socially defined universe" that constituted the speech community where the "grammatical rules define[d] the bounds of the linguistically acceptable" (Gumperz 1968:381). The emphasis was now on how the norms varied among speech communities and the similarities and differences among the speech varieties. In other words, the dimension of variation had made its way into the study of linguistic phenomena (Labov 1968). The question was whether to handle variation on the basis of "informant reaction (the so-called 'intuition' of the native speaker) or on distribution-frequency measures" or on some other criteria (Ferguson 1966a). This question was obviously important for those who held that we had to distinguish between "variations that are an aspect of a certain code" and "those that are due to the presence of another code" (Hasselmo 1969:122).

Immigrant groups as such were among speech communities with their own norms "often quite different from those prevailing in [...] monolingual communities" (Gumperz 1964). In this respect, *Haugen's the Norwegian Language in America* (1953) had been a significant contribution as it had drawn attention to the fact that a speech community was not always "coterminous with a single language" (Gumperz 1962:30). Importantly, there was also recognition that "studies of individual languages var[ied] greatly in range" but they had hardly been "based upon comparable sets of data" (Gumperz 1962:28). This line of thinking would lead to "more generalized formulations" of language contact situations "through the application of concepts such as code matrix, role distinctness, language distance and language loyalty" to bilingual communities (Gumperz 1962:33). The long search for language contact universals was starting. Obviously, this search was not independent from the general trends in linguistics at the time. In addition to the issue of universals, two other recurring themes of this decade were the "behaviorism versus mentalism" and "competence versus performance" debates (Garvin 1970:10). With respect to the universals debate there were "two extreme positions" neither of which contributed to the search. According to the first position "universals were rejected as wholly premature" and according to the second every "linguistic effort", to be valid, had to deal with universals. The main problem, however, was "what kinds of universals" were to be postulated and how these universals would "relate to the remaining assumptions" held "about natural language" (Garvin 1970:10).

For a "balanced view of linguistics" neither a purely behavioristic nor a purely mentalistic attitude would work. Likewise, it was pointless to discuss "whether the proper province of linguistics [was] the study of competence or of performance" as the "unobservable features of competence" could only be arrived at through performance (Garvin 1970:10). The competence/performance question inevitably occupied a central position in research on bilingualism.

3.4.1 It was now becoming more widely recognised that bilingualism was an asset both to the individual and to the society. On the one hand, there was the traditional quantitative study of bilingualism, and on the other the sociolinguist's emphasis on the qualitative aspects. The measurement of bilingualism was based

on the assumption that bilingualism was "a variable phenomenon", therefore "it [could] be measured" (Mackey 1969:5). However, before it could be measured, it had to be defined. The definition came in terms of categories (e.g. proficiency), dichotomies (e.g. stable vs. unstable), and scales (e.g. bilingual semantic differentials) none of which proved totally successful (Mackey 1969:5-6). The measurement, likewise, came in terms of "function, stability, and distribution of the languages involved, in relation to their location, origin, and dominance" (Mackey 1969:6) but there were problems with the identification of the unit of measurement.

These problems added to the feeling that in Psychology, Sociology and Linguistics, bilingualism had not been adequately dealt with, both in terms of theoretical models and methodological procedures. In Psychology, the bilingual individual had been of interest in relation to the measurement of IQ and educational attainment (see Macnamara 1967a, Fishman 1968a for a review). Consequently, the psychological study of bilingualism had been confined to testing speed (e.g. Lambert 1955), fluency and flexibility (eg Lambert *et al.* 1959a and b), and dominance in bilinguals (eg Lambert 1955). In Linguistics, interference was assumed between "two entities" in contact, that otherwise would live "in a pure and unsullied state" (Fishman 1968a:27). Also in Sociology, while census data had been useful in exploring a wide range of linguistic issues with respect to bilingual populations (Lieberson 1966), sole reliance on it had been misleading. This was the state of affairs, when Fishman (1968a:34) protested to make room for Sociolinguistics or as he preferred to call it: *The Sociology of Language*:

The three disciplines [Psychology, Sociology and Linguistics] [...] are quite similar in their shortcomings with respect to the study of bilingualism. Bilingualism has normally been a fringe topic rather than a central topic with each of them. [...] Bilingualism research has more frequently suffered from the use of discipline-derived (rather than problem-derived) cut-and-dried approaches. Investigators have employed impoverished models of bilingualism, utilized conceptually impoverished data gathering approaches, obtained impoverished results and, as a result, reinforced an impoverished view of the field. It would be good to break out of this circle (Haugen 1973:511).

The structuralists were to blame for including in their work "precious little either about bilingualism or any other kind of linguistic variability" (Haugen 1970:3). Chomskyan view, being concerned with an ideal speaker in a homogeneous speech community, left no room for the bilingual who was neither an ideal speaker as such nor lived in a homogeneous speech community.

How could this 'impoverished view' of bilingualism change? It was not "so much languages" but "subsets of individuals whose" "verbal repertoire[s]" came into contact and revealed the patterns to be investigated (Fishman 1968a:29). As a result, Fishman's sociolinguistic recipe included notions such as 'community norms', 'the identification with cultural values', 'role relations', 'networks' and 'language choice'. The aim was to move the focus of language contact research over to the relationship between socio-cultural organization and language use in bilingual communities. The view now in vogue was that speakers shared "not only rules of grammar, but also rules of appropriate speech usage" with other members of their society (Sankoff 1974:47-79).

In connection to this argument, the notion of competence was extended from "the mastery of a set of grammatical rules to the mastery of a set of cultural rules that include the appropriate ways to apply grammatical rules in all speech situations possible for that society" (Sankoff 1974:47-79; Haugen 1970:5).

The debate on compound-coordinate distinction was not bearing fruit. The unanimous view was that the way Weinreich had intended it was misunderstood. Haugen (1970:41) tried to clarify the misunderstanding:

compound-coordinate distinction [...] builds on a partial misunderstanding of Uriel Weinreich's book (1953). I have reread very carefully what he wrote on this subject. He was very skeptical about the notion that people could be classified as compound or coordinate. He merely wrote that some signs in the language could be stored as compound signs and some as coordinate, while in addition, some could be subordinate.

The target of his criticism was the "concept of compound and coordinate as used by the psychologists" including Ervin and Osgood, Lambert and associates. However, the study of bilinguals suffering from aphasia carried out by Lambert and associates (1959b) at the end of the first decade promised insights regarding this distinction.

3.4.2 "[T]he linguist's concern with interference", according to Fishman, was legitimate only when it answered the question "which language is being used" (1966a:426). What was the point of measuring the degree of interference otherwise if not as an indication of maintenance and/or shift? Regardless of the sociolinguist's criticism, however, interference was a significant aspect of bilingual research although what exactly constituted interference and whether the term was 'noxious' or 'neutral' (Haugen 1970:2) was still much debated.

It was by now widely accepted that "[t]he best way to approach" the investigation of interference was "to study the performance of persons" in bilingual contexts (Macnamara 1967b:69). Performance errors, as such, were considered important as they arose from "overtaxing the programming capacity of the speaker" and "violate[d] co-occurrence restrictions socially or linguistically, producing interference" (Ervin-Tripp 1969:33). It was now generally accepted that bilinguals had difficulty keeping their "codes wholly apart", and "to save themselves effort", they had to "take shortcuts" (Haugen 1973). Neither in this decade, however, a performance model was proposed for the production of sentences containing instances of interference.

Elsewhere in the interference debate, the search for "a universal theory of interference" was still missing (Diebold 1962:43). A universal theory of interference had to account not only for "the differential resistance to interference in the structural hierarchy" (Diebold 1962:43) but also for the psychological mechanisms underlying the types of interference encountered in data (Ferguson 1966b). However, much of the interference debate was sociolinguistic as Gumperz had reinterpreted it in terms of an integrated communication matrix within a bilingual society: "Code switching in everyday interaction sets up cross currents of diffusion which materially change the structure of local speech

varieties" (1967:49) and such interaction would generate its own norms of correctness (1967:50). Here, he was building on the "bilingual norms" pointed out earlier by Haugen (e.g.1953:60-62).

Like interference, integration was treated at different levels, phonological, grammatical and lexical (e.g. Hasselmo 1961). The focus was on the measurement of integration (Hasselmo 1969; Mackey 1970) and the development of tests to this end. However, after two decades of work on the topic Mackey (1970) would acknowledge that the question of integration was still a problem: "One of the most difficult puzzles in the study of bilingualism has been the separation of cases of integration (borrowing) into the code from cases of interference in the message". In other words, there was the problem of how to identify the integrated items as well as how to measure their stability with respect to the overall rate of change in the language. Frequency of occurrence as a criterion had already been questioned, so had the range of texts in which a particular item would be used. In other words, measuring integration in terms of "the message" did not work and we had to measure from "the code". This was another way of packaging the "langue-parole" question and its relevance to language contact data. Measuring from the code included "tests of availability, acceptability, and translatability" (Hasselmo 1969; Mackey 1970). *Availability* represented the totality of the words that might be used in a bilingual situation, *acceptability* meant a selected item's relation to the norm, and *translatability* was "the bilingual's ability to furnish equivalents" between the two languages (Mackey 1970).

The debate on the levels of analysis continued also during this decade without much solution. In an evaluation of the available levels of analysis, Hasselmo (1969:124) drew attention to the previous disagreement between Haugen, whose favourites were phonemic and morphemic, and Weinreich, who separated grammatical and lexical levels. There was also Mackey's (1965) scheme that recognized all the levels in the hierarchical structure of sentences. While brushing the lexical level aside, Hasselmo (1961:51-52) himself declared that interference "on the phonological level" and morphological level (1969) were "more easily defined". The view that "phonological and grammatical pattern[s]" seen as "the structural core of a language", were "more resistant to change" still prevailed, and

the elements of the lexicon tended "to be adapted to the patterns of the recipient language" (Gumperz 1968:382).

3.4.3 In an overview of this decade at the beginning of 1970s, Haugen (1973:521) summarised the way borrowing was seen:

A case of interference which has been accepted and shows by its form that the speaker has made it part of his language can then be said to have been integrated which most nearly corresponds to the traditional concept of borrowing.

In other words, it was interference but integrated as opposed to switching which was not or was it? Borrowing was, in a sense, losing popularity as it referred "to the end result of a process of change" failing to "provide an insight into the dynamics of this process" (Gumperz 1962:30); and also persistent borrowing "obscure[d] pre-existing genetic distinctions" and thus led to stronger similarities in linguistic structure of two languages (Gumperz 1968:382).

In this decade, with the emphasis on social factors, the question of resistance to borrowing, would involve not only linguistic factors such as compatibility between the two languages but also contextual factors such as the type and duration of contact, the social function of the languages, and the speaker's attitude to foreign elements (e.g. Diebold 1961:111).

3.4.4 What exactly was *switching* in this decade? It was considered "part of the bilingual's skill" but it could also "take the form of linguistic interference" (Macnamara 1967b:70). It could be with or without interference, involved translation, was not related to the degree of bilingualism (Macnamara 1967b:70-72), was triggered by structural factors such as word order (e.g. Clyne 1967:65); and/or extralinguistic factors such as "persons, places and topics" (e.g. Macnamara 1967b:71) and so on.

At the beginning of this decade, Hasselmo (1961:40-42) came up with the notion of the base language. As it was coined, this notion looked fairly innocuous but it was to cause controversy in the decades to come:

each situation possesses certain characteristics which call for the use of one or the other of the languages [...] As an interview progresses, a situation may also change its characteristics and thus a change of base language also takes place (1961:41).

In Hasselmo's view the concept of the base language would "limit our task to that of explaining only the cases of switching from the base language into the other, but not vice versa" (1961:42). What would determine the bilingual's choice of the base language? Hasselmo's concept involved the location, participants and topic. More classification was on its way in terms of 'clean', 'ragged' and 'pathological' switching (i.e. the presence or absence of phonic and/or morphic overlapping in switches); 'limited' and 'unlimited' (i.e. the introduction of complete immediate constituents of sentences or complete discourse segments vs the introduction of stretches not describable as units); 'marked' and 'unmarked' (i.e. the presence or absence of signals of a change of language) 'anticipational' and 'consequential' (Hasselmo 1961:53-54; 1969:127, Clyne 1967) and so on.

The analysis of switches out of their linguistic context had proved unsatisfactory: Haugen had put the study of switching on the map but he paid little attention to the linguistic environment as a cause of the phenomenon. From this decade on attention would be paid to larger units (e.g. discourse) in language contact data (e.g. Hasselmo 1970) although they had been commonly used in the ethnographic method to explain the relation between language and culture (see Hymes 1970 for a review). This was only one way of approaching the analysis of switching, that is "starting with the broad patterning of use of two languages in a community and working down toward a finer and finer analysis of code-switching in a given discourse". The other way would be

starting with a minute analysis of a set of recorded discourses, and working up toward the broader patterning in the community as a whole (Hasselmo 1969:136).

Hasselmo had observed that switching did not "take place haphazardly at any point in the sentence" but in "certain crucial borders" and the location of switches were subjected to restrictions of structural nature. Particularly adjective+noun combinations in noun phrases were not interrupted (1961:55). He concluded that switching took place at clause boundaries (1961:56) and between free morphemes (1961:58). The drums were beginning to beat.

Later in this decade, there was some effort to further Hasselmo's earlier finding that a loanword may carry over into the following clause and lead to a switch "where the speaker is stuck for a word" (Clyne 1967:16-17). The lack of phonological borders in such situations would create a neutral zone particularly in the use of two similar languages such as German and English, facilitating the switches.

Borrowing vs code-switching distinction was acquiring official status although the boundary was still blurred. For Hasselmo (1961:19) both "switching" and "borrowing" were "voluntary" ways of "satisfying the new needs" of migrants for "communication". However, they led to "interference" which was "the involuntary result of these two ways of coping with the expressive needs". The inherited negative view of interference was firmly there. When would interference "cease to be a case of interference?" (Haugen 1973:521). The answer was when it became part of the norm and frequency of occurrence was one way of distinguishing between integrated and unintegrated items but that was not successful, either (Hasselmo 1961:63).

Where did nonce borrowings stand? They were at the transfer end of the integration continuum moving gradually towards the loan stage. The debate on integration, however, led to the "standard procedure in linguistic descriptions to delete from one's corpus all recent loan-material" as the "inclusion of such materials is rarely tolerated until the interference has been sufficiently long-standing so as to finally affect paradigmatic changes in the language under study" (Diebold 1962:45).

3.4.5 In terms of mixing languages, the study of pidgins and creoles had "been a stepchild" in linguistics. It would gain importance in this decade (Hymes 1971a:3). As pidgins and creoles were part of activities such as colonialism, trade and migration; and involved the processes of reduction and expansion, their description and history would benefit the study of language contact in general (Hymes 1971b). The similarity with migrant languages, in the industrial context of 1960s in Europe was obvious: pidgins and creoles had always been labelled as 'marginal languages' associated with "poorer and darker members of a society"; and they were seen as "deviations from other systems" born out of the processes of "hybridization, simplification, convergence, and acculturation" (Hymes 1971a:7). All of these arguments have since been made with respect to migrant languages (e.g. Turkish) in contact with the language/s of the host country (e.g. German and Dutch) (see Chap. 2).

Accordingly, a theory of pidgins and creoles had to include (Hymes 1971a:9):

1. The universal tendencies to adapt speech, and varieties of a language, by simplification in some circumstances, expansion in others.
2. The occurrence of these tendencies in situations of language contact, so as to give rise to partial confluence of linguistic traditions.
3. The conditions, linguistic and social, under which forms of speech so adapted and influenced become and remain independent of norms of any contributing tradition.
4. The subsequent histories of languages so formed.

The student of language contact has also been very familiar with these theoretical considerations although their integration into a theory of language contact as such has been more difficult. This may be due to the differences between the types of language contact leading to a mixture in the sense of pidgins and creoles; and mixing as understood in the context of migration investigated in this study.

Mixing in this narrower sense did not seem to be used widely in the second decade as Haugen's previous criticism of the term was taken seriously. In relation to "the question of whether mixing fulfill[ed] a social function" (1953:54) he had proposed "a detailed study of the social responses called forth by mixing" and although the term was not popular with the researchers of this decade, the study of the social responses was being carried out.

3.4.6 In the previous decade it was observed that contact between similar languages led to easier integration of items. Similarly, in the second decade, the use of languages with parallel structures such as German and English was shown to be facilitating transfer mechanisms. This was due to the overlapping area, created by the similarity between the languages, in the bilingual's mind (e.g. Hasselmo 1961:Chap.4). The "words and expressions" belonging to the overlapping area were found to "exercise a triggering effect" on the bilingual, which would be either *consequential* (switching to the other language after a trigger word) or *anticipational* (switching in advance of a trigger word).

3.4.7 Contrastive analysis was in full swing. According to Hasselmo (1961:22) before the issues of interference and integration could be described it was "necessary to present a description" and a comparison "of the participating structures". This was done to separate "the purely linguistic causes of interference" "from the extra-linguistic ones". Consequently, he presented the classes and categories of Swedish and English grammars and identified the overlapping areas (1961: Chap 4). More than the differences, *equivalence* between systems was the major premise of this approach (Hasselmo 1969:123, 129). In this view, where the systems differed, interference occurred. The nature of this interference could be predicted if the researcher carried out in advance a contrastive study of two languages, commonly seen as "inventories of units" (Hasselmo 1969:125). Naturally, the most significant units to be considered were "the phones and the morphs", their distribution in the respective languages and redistribution as a result of contact.

In the 1960s however "no theoretical rationale" was available for the contrastive study of languages except for a number of attempts "to formulate principles" that

would "yield fairly precise predictions" (Ferguson 1996c). The problem here was that a "whole range of external variables [would] interfere with the validation of the principles" and would "turn out to be inadequately formulated" (Ferguson 1996c). During the first half of the decade, in particular, "the publication and use of contrastive studies, especially those between English and another language" showed a considerable increase (Ferguson 1996c). Still, there was no agreement as to the "procedures of contrastive analysis" and what was even more serious was the lack of a "coherent theory of interference which would bridge the gap between even the most sophisticated contrastive analyses and observed language behaviour" (Ferguson 1966a).

3.4.7 The view that "the development of an immigrant language is [...] normally predetermined to be one towards eventual extinction" (Hasselmo 1961:26) was now popular. Although language loss in this decade was still waiting to be picked up as an area of inquiry, extensive studies of language shift would gradually further language contact research into the extreme case of loss.

Since Jespersen the view that "there was a general trend in language toward simplification" had been dominant. Later, the evolutionary view of language began with simple monosyllabic words and gradually developed complex forms (Swadesh 1972:3). In terms of change then it was not only the internal vs. external factors debate but also simplification vs evolution. For Fishman (1966b:440-441) change in language could not be considered without reference to "change in human behaviour" in general. The challenge waiting to be met was the "determination of the circumstances under which language and non-language behaviours change concurrently, consecutively or independently" (Fishman 1966b:441). What this decade had sealed in terms of the perception of language change was that the impact of social factors could not be taken for granted. "How" then would "the observed changes" be "embedded in the matrix of linguistic and extralinguistic" structures (Weinreich, Labov and Herzog 1968:101)? This question has kept researchers busy since it was first asked in this decade. A reliable account of change would have to "show the influence of the structural environment upon the feature in question" (1968:172).

Fishman (1966b:425) was dissatisfied with the study of interference per se as the primary concern of Language Contact because it did not relate language choice to socio-cultural settings (1966b:429). Therefore, he proposed the concept of *domains*, "a socio-cultural construct abstracted from topics of communication, relationships and interaction between communicators and locales of communication [...]" (1966b:430). The main premise of the domain argument was to show the gradual "overlap" and "restriction" of "disadvantaged" languages (1966b:433). Previous research supported this view: "the immigrant language remains - for a varying length of time - the language of inter-group activities" while "the language of prestige" dominates "outward-directed communication" (Hasselmo 1961:26).

From this decade on language shift came to have a counterpart, language maintenance referred to by Haugen as a rare event: "Within a single, cohesive social group the use of a single language is the rule and the maintenance of a second language the exception" (1953:5). Fishman (1966b:424) drew attention to the link between the two and contact situations: "change or stability in habitual use, on the one hand, and ongoing psychological, social and cultural processes, on the other" rather than interference as the object of study. The main factors that, according to him, contribute to language maintenance and inhibit language assimilation were as follows: religio-societal insulation; time of immigration; existence of language islands; parochial schools; pre-emigration language mixing experience; former language use. In addition to these there were also other factors such as the educational level or the size of the migrant group that had some impact on the maintenance efforts (Kloss 1966:206). Kloss (1966) thus differentiated between clearcut and ambivalent factors in language maintenance and shift.

3.4.8 Where did the lexicon stand in this decade? Simply, where it did in the previous decade. Phonology and grammar were seen as systems whereas the lexicon was not. So it was more vulnerable to extra-linguistic influences (Hasselmo 1961:23).

More contradiction was on the way. In this view, lexical analysis was quickly discarded on the grounds that it was: "combined with considerable difficulties, and

ha[d] little practical value" (Hasselmo 1961:51). The cases of "complete substitution" with appropriate inflections would be considered part of the lexicons of both languages (Hasselmo 1961:51) but what we would do with the rest was open to question. Hasselmo's approach was not really doing the lexicon much good. There was also the view deriving from the standardization argument that for a language to be modernized its lexicon had to expand either through its "own processes of word formation or from extensive borrowing from another language" (Ferguson 1968:228). This point was made in spite of the 'simplification' argument.

At the beginning of the decade, Hasselmo (1961:62-66) used frequency as a criterion to determine the stability of loanwords in American Swedish. His conclusion contributed to the borrowing vs code-switching dilemma (1961:63):

it has turned out that the frequencies of such [limited] switching do not differ significantly from the frequency of borrowing.

Later, frequency of occurrence was criticized on the grounds that it was not enough to work out the bilingual's dominant language. According to Mackey (1965) frequencies were unstable and therefore unreliable when dealing with interference, and Fishman claimed it was the types of use that counted not the frequency of use (1966b:436). Yet, "the loanword count" remained in vogue as "the most common type of measurement of interference" (Hasselmo 1969:129).

3.4.9 Previously used data gathering techniques such as the 'written questionnaire' and 'direct interrogation' were criticized and the description of pictures was added to methodological options in bilingual research (Clyne 1967:22). Earlier in the decade, Hasselmo (1961:24) had used switching, translation, and lexical tests. Some of these tests are still with us today. "Such techniques" however, proved to be "only marginally relevant" to "sociolinguistically oriented" surveys. Fishman (1968a:36) had combined participant observation, interview and self-report techniques with the aim of emphasizing the qualitative aspects of bilingual behaviour (1966b). The other effective techniques were "the culling of information from published sources" as

well as "consultation with experts and persons knowledgeable about specific areas of problems" (Ferguson 1966c).

3.4.10 As this decade closed, it was predicted that "the future of sociolinguistics [lay]" in the "explanation of the origin, maintenance, change and loss" of languages as well as "the structure of variation" and the "adaptation of lexical and syntactic" material. Equally important would be "the study of the emergence and adaptation of the phenomena variously identified as code-repertoires, registers, speech levels, speech styles, and the like" (Hymes 1971b:5-6). The next decade would also see a search for the nonrandomness of what was becoming a central issue in language contact research: switching. With the sociological emphasis on the variation, use and function of language, it was inevitable that functional aspects of switching would come to the fore and even the structural analysis of code-switching would involve questions of functional relevance (Hymes 1968).

The debate on general properties of language, namely *universals*, was also opened in this decade thanks to the works of Chomsky and Greenberg. The problem of the universal constraints on language contact processes would, if it were solved, specify which switches were possible and which were impossible, and predict which switches would take place in particular circumstances. These questions would be partially answered in the third decade as the path set during the first and second decade remained unchallenged.

3.3.5 It is also appropriate to refer to *mixing* here, an equally controversial term. Haugen criticized the older generation of linguists such as Hermann Paul, Whitney and Schuchardt for using this metaphor. According to him, it "suggest[ed] that two languages can be put into a kind of cocktail shaker and then emerge as an entirely different concoction". While objecting to the term **mixture** on the grounds that it "imply[ed] the **creation of an entirely new entity** and the

disappearance of both constituents, or a jumbling of a more or less haphazard nature" his argument for abandoning this term was not truly convincing: "they [bilinguals] may **switch rapidly** from one to the other, **but** at any given moment **they are speaking only one**, even when they resort to the other for assistance" (emphasis added) (Haugen 1953:362).

Haugen was not alone in his contribution to the source of today's dilemma regarding the outcome of language contact: "Can the descriptive linguist in an actual contact situation derive any indications from the speech of bilinguals as to whether a new language is in the making?" asked Weinreich (1953:69, 105). Also Martinet, in his preface to Weinreich's book, had questioned the possibility of a 'structural merger' between the bilingual's languages.

So far there have been two opposite camps around this question. One view is that the outcome can be accounted for in terms of the participating languages. Whereas, the other view, to which this thesis also subscribes, attempts to develop a method for the search of the 'new language in the making'.

It was in relation to this question that the possibility of a transfer grammar was first investigated. This was postulated by Harris (1954) though not immediately followed up (see Haugen 1972). However, the notion has stayed with us since then. There, Harris was interested in establishing a grammar that would take into account the differences and similarities in the grammatical structure of the two languages in question. Interestingly, Harris (1954) saw this method

relevant to a proceduralized system of translation, and indeed can be put in the form of routine instructions for machine translations; and this not only because of the inherent connection between transfer and translation, but also because sentence-pairs under translation are used in certain transfer foundations.

Through detailed listing of similar and different structures in the languages involved, Harris wanted to obtain a revised grammar of the language A based as far as possible on the categories of B. Later attempts (e.g. Oksaar 1972, Joshi

1984) to construct a third code grammar out of the participating languages in contact can be seen as a follow-up to this early study of a transfer grammar.

3.3.6 Languages with parallel structures, Haugen claimed integrated loans easily and in this process, nouns would be accepted as nouns, adjectives as adjectives and so on with appropriate inflection (1956:57). Consequently, in this view, borrowings rarely established new categories in the receiving language (Haugen 1956:58).

The notion of *equivalence* between languages that is critically examined in this thesis was already popular in this decade. The importance of this notion for the bilingual individual and its suitability as a method of analysis of the bilingual situation were discussed. Weinreich's (1953:7) *interlingual identification* was about the tendency of bilingual speakers to equate items between the languages involved. Haugen discussed (1956:48) the identification of lexemes in relation to specific contexts. However, he was aware that not all meanings of a lexeme in one language would find an equivalent in the other (Haugen 1956:48). Contrastive analysis gained popularity at this time as a method of establishing similarities and differences between languages (Haugen 1956:41). It has since regained popularity after being discredited and turned into a sub-field within Linguistics (Kachru 1980). This was also the time when the distance between languages was beginning to be taken into account in the investigation of equivalence (Haugen 1956:49).

3.3.7 The concern with the ever-prominent themes of Language Contact such as *change*, *restriction* and *shift* was also apparent during this time. Haugen (1956:27) pointed out that:

[...] **changes** occur in languages which live under conditions of strong bilingual pressure, where the number of monolingual speakers of the immigrant language is constantly being reduced. The immigrant is being gradually integrated into the main cultural stream of his country; as the anthropologists put it, he is being acculturated, whether he wishes or no. This process finds its expression in his first language, first by the **changes it undergoes**, and secondly by **the gradual restriction** of its use. It does not take long before he finds its

resources inadequate to express the culture he is entering, and he is forced to resort to the usual devices for **extending its vocabulary**. It is inevitable that he chooses those expressions which are already at hand for describing the new culture, viz. those of the national language in his new country. A **gradual shifting of content** occurs, whereby the immigrant language becomes more and more restricted in its functions and requires more and more borrowing from the other language to eke out its vocabulary (emphasis added).

Particularly, the notion of change had always been closely related to language contact since 19th and early 20th centuries. Language contact, in fact was considered of crucial importance to the hottest issue of linguistics at the time, language change. However, not all changes in language, Sapir had observed, came as a result of contact (1933):

there is a tendency among anthropologists and sociologists to operate far too hastily with wholesale linguistic changes due to external ethnic and cultural influences. The enormous amount of study [...] shows very clearly that the most powerful differentiating factors are not outside influences, as ordinarily understood, but rather the very slow but powerful unconscious changes in certain directions which seem to be implicit in the phonemic systems and morphologies of the languages themselves.

Consequently, the investigation of external linguistic and social factors came to be known as *contact-induced change*. It came through "a gradual infiltration of loans" that "were not limited to actual cultural novelties or so-called *necessary* words". This process, Haugen concluded, led to a discrepancy between migrants' "experience in the new language" and "their experience in the old" and this "pressure led to linguistic change" (1953:372). Here, we have a persisting confusion regarding the source of change going back to, unsurprisingly, Saussure again, who himself was confused as to whether change was the speaker's or the language's doing. Saussure "both claimed and denied that speakers can discern word-parts, change language, and observe language change" (De Beaugrande 1991:33ff). We also inherited from Saussure the negative view of change including 'deteriorations', 'vicissitudes', 'damage', 'disturbance', 'breaking' and

'effacement' in spite of which 'language continues to function' (De Beaugrande 1991:20).

Restriction has been a politically useful argument due to a number of reasons. Fishman was the main promoter of the *Sociology of Language*. It embraced all the areas of language contact where restriction as such was seen as the key issue and was elaborated in relation to the notion of *domains*. It was of prime importance if Fishman was to mobilize the community for the support of certain immigrant groups and languages. The notion of restriction would later be supported by the sociolinguistically-imposed "dichotomy of dominant vs. dominated, with English as dominant and all other languages of whatever source as dominated" (Haugen 1973:513).

Also *loss* was used in connection to these arguments (Haugen 1953:74), and became popular later on. While the Norwegian immigrants Haugen studied were learning new words, "at the other end of their vocabulary a **loss** was simultaneously going on, an atrophy of terms no longer needed or heard" (Haugen 1953:74) [emphasis added]. As this section summarises, Haugen used most of these terms interchangeably, although he had proposed the distinctions himself.

As a result of these researchers' interest in the individual efforts to maintain the languages then under investigation (Haugen 1953:292; Weinreich 1953:86) came also what Fishman (1966) would later turn into a study of its own right: *language loyalty*.

The migrants Haugen studied "could not escape the necessity of a lexical adaptation to their environment", which came particularly in the areas of "official, economic and social life". Haugen's concern was to list in detail "the new words that were introduced and [...] the old ones that were lost" (1953:75). However, in the section 'The Great Vocabulary Shift' dedicated to numerous borrowings from English into Norwegian, he concluded: "But in general the speakers of AmN were not borrowing beyond the number of words that one might expect; the core of their language was wholly Norwegian, and in no way the kind of non-descript pidgin that some have suggested" (1953:97). It looks like his frequency-based data

analysis supported this point but what was the threshold he had in mind when he said the borrowings did not exceed what 'one might expect'? And if they did not, why was he referring to a great shift?

Another aspect of language contact considered then (e.g. Martinet 1953) would also turn into a major area of investigation later on: whether contact breeds *convergence* or *divergence*. The complaint was that "linguistic research has so far favoured the study of divergence at the expense of convergence. It is time the right balance should be restored". The common assumption was that if geographical distance led to divergence, then contact would lead to convergence. Martinet (1953:viii) particularly encouraged the study of linguistic convergence resulting from the contact of two distant languages. Here, we observe the repositioning of historical linguistics as a result of the considerations of the language contact researchers of this decade:

It is a scientific exploration of contemporary bilingual patterns that will enable us to define exactly what shall be meant by such terms as substratum, superstratum, and adstratum, and to what extent we have a right to apply them to a given historical situation (Martinet 1953:viii-ix).

3.3.8 As for the vocabulary, it was a "loosely patterned domain of language" and that was it. Gleason, at the first major lexicography conference in 1960 protested that Bloomfield was to blame for seeing the lexicon as "an appendix of the grammar" (1933:274). Before him there was Saussure of course, whose lexicon was 'a listing of irregularities'. Gleason (1962) was not sure if linguists had

any clear idea what they mean when they say the word [lexicon]. Commonly it is treated as essentially equivalent to vocabulary, perhaps only slightly more elegant synonym. Certainly we descriptive linguists tend to be contemptuous of vocabulary. It is almost a dogma among us that vocabulary is the least significant part of language (save for a group among us who even doubt that vocabulary is really a part of language at all).

Consequently, a very firmly established belief, since the earliest studies of contact, has been the *mobility of nouns in borrowing* (Weinreich 1953:37). This observation was made much earlier by their predecessors (e.g. see Sapir 1921:chap. IX for a detailed discussion) and resulted in not more than long inventories of loans (This status of nouns in borrowing is critically examined in section 7.5.1). After all, vocabulary was a "sensitive index of the culture of a people" and as Sapir saw it "changes of the meaning, loss of old words, the creation and borrowing of new ones" were "all dependent on the history of culture itself" (Sapir 1933:27).

Naturally, most language contact studies afterwards have done what Haugen did then, list the loans in detail. This has been due largely to the unchanging image of the "grammar as an integrated system, and lexicon as a miscellaneous remainder" (Gleason 1962:96). "If lexicon can be given no better basis than this" Gleason wondered, could it "be considered as a basic segment of language structure in the way that the grammar and the phonology are to be considered?" This question still holds: Has it been given better basis over these years? The following sections on the subsequent decades examine the development of this basis up to present.

Frequency of occurrence (e.g. Haugen 1953: 92, 97; Weinreich 1953:82; Pike 1967:193) was, of course, an important criterion, not only for language contact but linguistics in general. It would soon make a big hit with the completion of Francis and Kuèera's Brown Corpus (see chapter 4, see also 3.3.10 for specific reference to Turkish). Haugen (1953:97) saw it "highly probable that words of high frequency in the native language are less likely to be displaced than those of low frequency". He then went on to establish another dictum of Language Contact, that of "any word can be borrowed":

Of the 1000 most frequent words in the writer's N[orwegian] word count, only 47 or less than one in twenty, are ever displaced by E[nglish] words. Most of these are terms referring to very specific aspects of N geographical or social life [...]. But individually and sporadically any word can be borrowed; the writer has even heard the conjunctions AND and BUT used in a N[orwegian] context by some speakers (1953:97).

The line of thinking outlined so far led to a long-lasting search for the *classification criteria* (Weinreich 1953:64-65; Haugen 1950, 1956:59-60). The attempt to classify loans, however, forced further boundaries on language contact data. It also led to the conclusion that the areas of influence of one language on another were evident. As Haugen put it: "The chief foci of influence were the store, the government, and the American neighbour". But in home and family life, in church and religion, the English expressions penetrated more slowly" (1953:93). This previously-made observation (e.g. Schmidt-Rohr 1932 quoted in Weinreich 1953:87) later turned into what Fishman would call *domains*.

3.3.9 And what were the methodological requirements? Given the picture of bilingual investigation at the time, the obvious methodology would be "observing closely the behaviour of bilinguals" (Haugen 1953:11). The preferred techniques of data gathering involved:

1. an adequate questionnaire
2. detailed phonetic transcription
3. reliable sampling of the speech community (Haugen 1953:321)

coupled with thorough information on "the local background and the atmosphere" as well as generational distinction (Haugen 1953:320-1). "Questionnaires as the chief tool of the dialect geographer" had been used since the late 1800s (Haugen 1953:321). Haugen adopted this method in his survey of American Norwegians and the questions he included in his questionnaires were of similar types: "1. general and cultural questions designed to bring out the informant's social and family background, 2. expressions referring to objects nearly always described in Norwegian, 3. expressions known to be very generally drawn from English" (1953:325). He also included detailed information on his informants in terms of their original dialect in Norway. It goes without saying that these fieldwork techniques are still largely in use today. Other data collection methods were also available, such as introspection (Weinreich 1953:12-13), and Leopold's work in the previous decade (1939-1949) pioneering a longitudinal case study. In linguistics, the prevailing research methods of the time, at least in the US, favoured

the spoken language, not the written, and were influenced by the anthropological tradition initiated by Boas and furthered by Bloomfield and Sapir (see Hymes and Fought 1981:102).

3.3.10 It is appropriate to pay tribute to Firth, a linguist of the first decade, whose notion of *collocation* is important for this thesis. Firth was not widely known in the US (see Kachru 1981), although his views "must have been at least partly known, through his presence at the 1949 Linguistic Institute" (Hymes and Fought 1981:153). Interestingly, neither Weinreich (1953) nor Haugen (1956) referred to Firth's work in their writings although they were greatly concerned themselves with *the* context of the bilingual situation and the *functions of bilingual usage* (e.g. Haugen 1956:93-96), two important concepts in Firth's work. Firth (1957:195-196) distinguished between the *contextual* and *collocational meaning* of words:

Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meanings of words. One of the meanings of night is its collocability with dark, and of dark, of course, collocation with night. This kind of mutuality may be paralleled in most languages [...].

This *syntagmatic* approach to the lexicon, however, was not popular with language contact researchers at the time and the view that it was no more than a repository of items prevailed. Firth died in 1960 at the end of this decade but his work would be furthered by Halliday, Sinclair and other colleagues in the subsequent decades.

At the boundary of this and the next decade it is also appropriate to pay tribute to a lesser-known researcher of the American Structuralist circle, namely Joe Pierce, whose work is of particular interest to this thesis. Pierce's frequency counts of spoken and written Turkish (1961, 1962a, 1962b, 1963a, 1963b, 1965) computed in person, in Turkey, 40 years ago, strike me as an unusual piece of work far ahead of its time. Yet, when placed within the academic climate of the US in 1950s and 1960s, it is not so difficult to understand his reasons for conducting such a study. Hymes (1972:230-231) summarises the development of structural linguistics in the US as follows:

In the second quarter of the century, the prospect for young linguists, especially those inspired by Sapir, included these tasks:

1. To develop the methods of the nascent structural linguistics and to test their application in the analysis of both exotic and well known languages.
2. To sustain the profession of linguistics, where almost no recognition existed so far as departments, chairs, specific courses, and autonomy of the discipline were concerned.
3. To continue to rescue disappearing languages.
4. To pursue proof and establishment of genetic relationships among languages.
5. To relate the results and methods of linguistic inquiry to other things - to other disciplines in the humanities and social sciences; to particular problems within these disciplines - such as cultural symbolism and patterning or personality and verbal art; and to practical affairs, such as education.

Pierce's work certainly linked to (1), (2) and (4) above, in that Turkish could be seen as an exotic language for a young American researcher in the 1950s. Secondly, Pierce did investigate the possibility of a genetic relationship between Turkish and Arabic on the basis of his data (e.g. 1965); and he also related his linguistic work to anthropological research (e.g. *Life in a Turkish Village* 1964), not surprisingly his main area of expertise.

This section has surveyed the major topics of investigation in the field of Language Contact, all dealt with in the first decade. These issues have remained central to the field since then and pursued by subsequent generations of language contact researchers. The following sections trace the developments around these points over time.

3.4 Second decade: 1960-1970

The preceding decade had introduced language contact "as a social phenomenon, in such a way that it might offer something of value to students of history and

sociology" (Haugen 1953:xi), and it did. Consequently, the decade 1960-1970 saw a sociolinguistic boom from which there was no going back. Later, in an overview of this decade, it was confirmed that in these years the most significant contribution to the study of language contact had come from Sociolinguistics.

From this decade on, language would be seen as an index of social processes (Labov 1968:240) "in which utterances are selected in accordance with socially recognized norms and expectations" (Gumperz 1968:381). Linguistic analysis meant the study of usage, a reflection of "more general behaviour norms", "within a socially defined universe" (Gumperz 1968:381). And it was this "socially defined universe" that constituted the speech community where the "grammatical rules define[d] the bounds of the linguistically acceptable" (Gumperz 1968:381). The emphasis was now on how the norms varied among speech communities and the similarities and differences among the speech varieties. In other words, the dimension of variation had made its way into the study of linguistic phenomena (Labov 1968). The question was whether to handle variation on the basis of "informant reaction (the so-called 'intuition' of the native speaker) or on distribution-frequency measures" or on some other criteria (Ferguson 1966a). This question was obviously important for those who held that we had to distinguish between "variations that are an aspect of a certain code" and "those that are due to the presence of another code" (Hasselmo 1969:122).

Immigrant groups as such were among speech communities with their own norms "often quite different from those prevailing in [...] monolingual communities" (Gumperz 1964). In this respect, *Haugen's the Norwegian Language in America* (1953) had been a significant contribution as it had drawn attention to the fact that a speech community was not always "coterminous with a single language" (Gumperz 1962:30). Importantly, there was also recognition that "studies of individual languages var[ied] greatly in range" but they had hardly been "based upon comparable sets of data" (Gumperz 1962:28). This line of thinking would lead to "more generalized formulations" of language contact situations "through the application of concepts such as code matrix, role distinctness, language distance and language loyalty" to bilingual communities (Gumperz 1962:33). The long search for language contact universals was starting. Obviously, this search

was not independent from the general trends in linguistics at the time. In addition to the issue of universals, two other recurring themes of this decade were the "behaviorism versus mentalism" and "competence versus performance" debates (Garvin 1970:10). With respect to the universals debate there were "two extreme positions" neither of which contributed to the search. According to the first position "universals were rejected as wholly premature" and according to the second every "linguistic effort", to be valid, had to deal with universals. The main problem, however, was "what kinds of universals" were to be postulated and how these universals would "relate to the remaining assumptions" held "about natural language" (Garvin 1970:10).

For a "balanced view of linguistics" neither a purely behavioristic nor a purely mentalistic attitude would work. Likewise, it was pointless to discuss "whether the proper province of linguistics [was] the study of competence or of performance" as the "unobservable features of competence" could only be arrived at through performance (Garvin 1970:10). The competence/performance question inevitably occupied a central position in research on bilingualism.

3.4.1 It was now becoming more widely recognised that bilingualism was an asset both to the individual and to the society. On the one hand, there was the traditional quantitative study of bilingualism, and on the other the sociolinguist's emphasis on the qualitative aspects. The measurement of bilingualism was based on the assumption that bilingualism was "a variable phenomenon", therefore "it [could] be measured" (Mackey 1969:5). However, before it could be measured, it had to be defined. The definition came in terms of categories (e.g. proficiency), dichotomies (e.g. stable vs. unstable), and scales (e.g. bilingual semantic differentials) none of which proved totally successful (Mackey 1969:5-6). The measurement, likewise, came in terms of "function, stability, and distribution of the languages involved, in relation to their location, origin, and dominance" (Mackey 1969:6) but there were problems with the identification of the unit of measurement.

These problems added to the feeling that in Psychology, Sociology and Linguistics, bilingualism had not been adequately dealt with, both in terms of

theoretical models and methodological procedures. In Psychology, the bilingual individual had been of interest in relation to the measurement of IQ and educational attainment (see Macnamara 1967a, Fishman 1968a for a review). Consequently, the psychological study of bilingualism had been confined to testing speed (e.g. Lambert 1955), fluency and flexibility (eg Lambert *et al.* 1959a and b), and dominance in bilinguals (eg Lambert 1955). In Linguistics, interference was assumed between "two entities" in contact, that otherwise would live "in a pure and unsullied state" (Fishman 1968a:27). Also in Sociology, while census data had been useful in exploring a wide range of linguistic issues with respect to bilingual populations (Lieberson 1966), sole reliance on it had been misleading. This was the state of affairs, when Fishman (1968a:34) protested to make room for Sociolinguistics or as he preferred to call it: *The Sociology of Language*:

The three disciplines [Psychology, Sociology and Linguistics] [...] are quite similar in their shortcomings with respect to the study of bilingualism. Bilingualism has normally been a fringe topic rather than a central topic with each of them. [...] Bilingualism research has more frequently suffered from the use of discipline-derived (rather than problem-derived) cut-and-dried approaches. Investigators have employed impoverished models of bilingualism, utilized conceptually impoverished data gathering approaches, obtained impoverished results and, as a result, reinforced an impoverished view of the field. It would be good to break out of this circle (Haugen 1973:511).

The structuralists were to blame for including in their work "precious little either about bilingualism or any other kind of linguistic variability" (Haugen 1970:3). Chomskyan view, being concerned with an ideal speaker in a homogeneous speech community, left no room for the bilingual who was neither an ideal speaker as such nor lived in a homogeneous speech community.

How could this 'impoverished view' of bilingualism change? It was not "so much languages" but "subsets of individuals whose" "verbal repertoire[s]" came into contact and revealed the patterns to be investigated (Fishman 1968a:29). As a result, Fishman's sociolinguistic recipe included notions such as 'community norms', 'the identification with cultural values', 'role relations', 'networks' and

'language choice'. The aim was to move the focus of language contact research over to the relationship between socio-cultural organization and language use in bilingual communities. The view now in vogue was that speakers shared "not only rules of grammar, but also rules of appropriate speech usage" with other members of their society (Sankoff 1974:47-79).

In connection to this argument, the notion of competence was extended from "the mastery of a set of grammatical rules to the mastery of a set of cultural rules that include the appropriate ways to apply grammatical rules in all speech situations possible for that society" (Sankoff 1974:47-79; Haugen 1970:5).

The debate on compound-coordinate distinction was not bearing fruit. The unanimous view was that the way Weinreich had intended it was misunderstood. Haugen (1970:41) tried to clarify the misunderstanding:

compound-coordinate distinction [...] builds on a partial misunderstanding of Uriel Weinreich's book (1953). I have reread very carefully what he wrote on this subject. He was very skeptical about the notion that people could be classified as compound or coordinate. He merely wrote that some signs in the language could be stored as compound signs and some as coordinate, while in addition, some could be subordinate.

The target of his criticism was the "concept of compound and coordinate as used by the psychologists" including Ervin and Osgood, Lambert and associates. However, the study of bilinguals suffering from aphasia carried out by Lambert and associates (1959b) at the end of the first decade promised insights regarding this distinction.

3.4.2 "[T]he linguist's concern with interference", according to Fishman, was legitimate only when it answered the question "which language is being used" (1966a:426). What was the point of measuring the degree of interference otherwise if not as an indication of maintenance and/or shift? Regardless of the sociolinguist's criticism, however, interference was a significant aspect of bilingual

research although what exactly constituted interference and whether the term was 'noxious' or 'neutral' (Haugen 1970:2) was still much debated.

It was by now widely accepted that "[t]he best way to approach" the investigation of interference was "to study the performance of persons" in bilingual contexts (Macnamara 1967b:69). Performance errors, as such, were considered important as they arose from "overtaxing the programming capacity of the speaker" and "violate[d] co-occurrence restrictions socially or linguistically, producing interference" (Ervin-Tripp 1969:33). It was now generally accepted that bilinguals had difficulty keeping their "codes wholly apart", and "to save themselves effort", they had to "take shortcuts" (Haugen 1973). Neither in this decade, however, a performance model was proposed for the production of sentences containing instances of interference.

Elsewhere in the interference debate, the search for "a universal theory of interference" was still missing (Diebold 1962:43). A universal theory of interference had to account not only for "the differential resistance to interference in the structural hierarchy" (Diebold 1962:43) but also for the psychological mechanisms underlying the types of interference encountered in data (Ferguson 1966b). However, much of the interference debate was sociolinguistic as Gumperz had reinterpreted it in terms of an integrated communication matrix within a bilingual society: "Code switching in everyday interaction sets up cross currents of diffusion which materially change the structure of local speech varieties" (1967:49) and such interaction would generate its own norms of correctness (1967:50). Here, he was building on the "bilingual norms" pointed out earlier by Haugen (e.g.1953:60-62).

Like interference, integration was treated at different levels, phonological, grammatical and lexical (e.g. Hasselmo 1961). The focus was on the measurement of integration (Hasselmo 1969; Mackey 1970) and the development of tests to this end. However, after two decades of work on the topic Mackey (1970) would acknowledge that the question of integration was still a problem: "One of the most difficult puzzles in the study of bilingualism has been the separation of cases of integration (borrowing) into the code from cases of interference in the message".

In other words, there was the problem of how to identify the integrated items as well as how to measure their stability with respect to the overall rate of change in the language. Frequency of occurrence as a criterion had already been questioned, so had the range of texts in which a particular item would be used. In other words, measuring integration in terms of "the message" did not work and we had to measure from "the code". This was another way of packaging the "langue-parole" question and its relevance to language contact data. Measuring from the code included "tests of availability, acceptability, and translatability" (Hasselmo 1969; Mackey 1970). *Availability* represented the totality of the words that might be used in a bilingual situation, *acceptability* meant a selected item's relation to the norm, and *translatability* was "the bilingual's ability to furnish equivalents" between the two languages (Mackey 1970).

The debate on the levels of analysis continued also during this decade without much solution. In an evaluation of the available levels of analysis, Hasselmo (1969:124) drew attention to the previous disagreement between Haugen, whose favourites were phonemic and morphemic, and Weinreich, who separated grammatical and lexical levels. There was also Mackey's (1965) scheme that recognized all the levels in the hierarchical structure of sentences. While brushing the lexical level aside, Hasselmo (1961:51-52) himself declared that interference "on the phonological level" and morphological level (1969) were "more easily defined". The view that "phonological and grammatical pattern[s]" seen as "the structural core of a language", were "more resistant to change" still prevailed, and the elements of the lexicon tended "to be adapted to the patterns of the recipient language" (Gumperz 1968:382).

3.4.3 In an overview of this decade at the beginning of 1970s, Haugen (1973:521) summarised the way borrowing was seen:

A case of interference which has been accepted and shows by its form that the speaker has made it part of his language can then be said to have been integrated which most nearly corresponds to the traditional concept of borrowing.

In other words, it was interference but integrated as opposed to switching which was not or was it? Borrowing was, in a sense, losing popularity as it referred "to the end result of a process of change" failing to "provide an insight into the dynamics of this process" (Gumperz 1962:30); and also persistent borrowing "obscure[d] pre-existing genetic distinctions" and thus led to stronger similarities in linguistic structure of two languages (Gumperz 1968:382).

In this decade, with the emphasis on social factors, the question of resistance to borrowing, would involve not only linguistic factors such as compatibility between the two languages but also contextual factors such as the type and duration of contact, the social function of the languages, and the speaker's attitude to foreign elements (e.g. Diebold 1961:111).

3.4.4 What exactly was *switching* in this decade? It was considered "part of the bilingual's skill" but it could also "take the form of linguistic interference" (Macnamara 1967b:70). It could be with or without interference, involved translation, was not related to the degree of bilingualism (Macnamara 1967b:70-72), was triggered by structural factors such as word order (e.g. Clyne 1967:65); and/or extralinguistic factors such as "persons, places and topics" (e.g. Macnamara 1967b:71) and so on.

At the beginning of this decade, Hasselmo (1961:40-42) came up with the notion of the base language. As it was coined, this notion looked fairly innocuous but it was to cause controversy in the decades to come:

each situation possesses certain characteristics which call for the use of one or the other of the languages [...] As an interview progresses, a situation may also change its characteristics and thus a change of base language also takes place (1961:41).

In Hasselmo's view the concept of the base language would "limit our task to that of explaining only the cases of switching from the base language into the other, but not vice versa" (1961:42). What would determine the bilingual's choice of the base language? Hasselmo's concept involved the location, participants and topic.

More classification was on its way in terms of 'clean', 'ragged' and 'pathological' switching (i.e. the presence or absence of phonic and/or morphic overlapping in switches); 'limited' and 'unlimited' (i.e. the introduction of complete immediate constituents of sentences or complete discourse segments vs the introduction of stretches not describable as units); 'marked' and 'unmarked' (i.e. the presence or absence of signals of a change of language) 'anticipational' and 'consequential' (Hasselmo 1961:53-54; 1969:127, Clyne 1967) and so on.

The analysis of switches out of their linguistic context had proved unsatisfactory: Haugen had put the study of switching on the map but he paid little attention to the linguistic environment as a cause of the phenomenon. From this decade on attention would be paid to larger units (e.g. discourse) in language contact data (e.g. Hasselmo 1970) although they had been commonly used in the ethnographic method to explain the relation between language and culture (see Hymes 1970 for a review). This was only one way of approaching the analysis of switching, that is "starting with the broad patterning of use of two languages in a community and working down toward a finer and finer analysis of code-switching in a given discourse". The other way would be

starting with a minute analysis of a set of recorded discourses, and working up toward the broader patterning in the community as a whole (Hasselmo 1969:136).

Hasselmo had observed that switching did not "take place haphazardly at any point in the sentence" but in "certain crucial borders" and the location of switches were subjected to restrictions of structural nature. Particularly adjective+noun combinations in noun phrases were not interrupted (1961:55). He concluded that switching took place at clause boundaries (1961:56) and between free morphemes (1961:58). The drums were beginning to beat.

Later in this decade, there was some effort to further Hasselmo's earlier finding that a loanword may carry over into the following clause and lead to a switch "where the speaker is stuck for a word" (Clyne 1967:16-17). The lack of phonological borders in such situations would create a neutral zone particularly in

the use of two similar languages such as German and English, facilitating the switches.

Borrowing vs code-switching distinction was acquiring official status although the boundary was still blurred. For Hasselmo (1961:19) both "switching" and "borrowing" were "voluntary" ways of "satisfying the new needs" of migrants for "communication". However, they led to "interference" which was "the involuntary result of these two ways of coping with the expressive needs". The inherited negative view of interference was firmly there. When would interference "cease to be a case of interference?" (Haugen 1973:521). The answer was when it became part of the norm and frequency of occurrence was one way of distinguishing between integrated and unintegrated items but that was not successful, either (Hasselmo 1961:63).

Where did nonce borrowings stand? They were at the transfer end of the integration continuum moving gradually towards the loan stage. The debate on integration, however, led to the "standard procedure in linguistic descriptions to delete from one's corpus all recent loan-material" as the "inclusion of such materials is rarely tolerated until the interference has been sufficiently long-standing so as to finally affect paradigmatic changes in the language under study" (Diebold 1962:45).

3.4.5 In terms of mixing languages, the study of pidgins and creoles had "been a stepchild" in linguistics. It would gain importance in this decade (Hymes 1971a:3). As pidgins and creoles were part of activities such as colonialism, trade and migration; and involved the processes of reduction and expansion, their description and history would benefit the study of language contact in general (Hymes 1971b). The similarity with migrant languages, in the industrial context of 1960s in Europe was obvious: pidgins and creoles had always been labelled as 'marginal languages' associated with "poorer and darker members of a society"; and they were seen as "deviations from other systems" born out of the processes of "hybridization, simplification, convergence, and acculturation" (Hymes 1971a:7). All of these arguments have since been made with respect to migrant languages (e.g. Turkish)

in contact with the language/s of the host country (e.g. German and Dutch) (see Chap. 2).

Accordingly, a theory of pidgins and creoles had to include (Hymes 1971a:9):

1. The universal tendencies to adapt speech, and varieties of a language, by simplification in some circumstances, expansion in others.
2. The occurrence of these tendencies in situations of language contact, so as to give rise to partial confluence of linguistic traditions.
3. The conditions, linguistic and social, under which forms of speech so adapted and influenced become and remain independent of norms of any contributing tradition.
4. The subsequent histories of languages so formed.

The student of language contact has also been very familiar with these theoretical considerations although their integration into a theory of language contact as such has been more difficult. This may be due to the differences between the types of language contact leading to a mixture in the sense of pidgins and creoles; and mixing as understood in the context of migration investigated in this study.

Mixing in this narrower sense did not seem to be used widely in the second decade as Haugen's previous criticism of the term was taken seriously. In relation to "the question of whether mixing fulfill[ed] a social function" (1953:54) he had proposed "a detailed study of the social responses called forth by mixing" and although the term was not popular with the researchers of this decade, the study of the social responses was being carried out.

3.4.8 In the previous decade it was observed that contact between similar languages led to easier integration of items. Similarly, in the second decade, the use of languages with parallel structures such as German and English was shown to be facilitating transfer mechanisms. This was due to the overlapping area,

created by the similarity between the languages, in the bilingual's mind (e.g. Hasselmo 1961:Chap.4). The "words and expressions" belonging to the overlapping area were found to "exercise a triggering effect" on the bilingual, which would be either *consequential* (switching to the other language after a trigger word) or *anticipational* (switching in advance of a trigger word).

3.4.9 Contrastive analysis was in full swing. According to Hasselmo (1961:22) before the issues of interference and integration could be described it was "necessary to present a description" and a comparison "of the participating structures". This was done to separate "the purely linguistic causes of interference" "from the extra-linguistic ones". Consequently, he presented the classes and categories of Swedish and English grammars and identified the overlapping areas (1961: Chap 4). More than the differences, *equivalence* between systems was the major premise of this approach (Hasselmo 1969:123, 129). In this view, where the systems differed, interference occurred. The nature of this interference could be predicted if the researcher carried out in advance a contrastive study of two languages, commonly seen as "inventories of units" (Hasselmo 1969:125). Naturally, the most significant units to be considered were "the phones and the morphs", their distribution in the respective languages and redistribution as a result of contact.

In the 1960s however "no theoretical rationale" was available for the contrastive study of languages except for a number of attempts "to formulate principles" that would "yield fairly precise predictions" (Ferguson 1996c). The problem here was that a "whole range of external variables [would] interfere with the validation of the principles" and would "turn out to be inadequately formulated" (Ferguson 1996c). During the first half of the decade, in particular, "the publication and use of contrastive studies, especially those between English and another language" showed a considerable increase (Ferguson 1996c). Still, there was no agreement as to the "procedures of contrastive analysis" and what was even more serious was the lack of a "coherent theory of interference which would bridge the gap between even the most sophisticated contrastive analyses and observed language behaviour" (Ferguson 1966a).

3.4.7 The view that "the development of an immigrant language is [...] normally predetermined to be one towards eventual extinction" (Hasselmo 1961:26) was now popular. Although language loss in this decade was still waiting to be picked up as an area of inquiry, extensive studies of language shift would gradually further language contact research into the extreme case of loss.

Since Jespersen the view that "there was a general trend in language toward simplification" had been dominant. Later, the evolutionary view of language began with simple monosyllabic words and gradually developed complex forms (Swadesh 1972:3). In terms of change then it was not only the internal vs. external factors debate but also simplification vs evolution. For Fishman (1966b:440-441) change in language could not be considered without reference to "change in human behaviour" in general. The challenge waiting to be met was the "determination of the circumstances under which language and non-language behaviours change concurrently, consecutively or independently" (Fishman 1966b:441). What this decade had sealed in terms of the perception of language change was that the impact of social factors could not be taken for granted. "How" then would "the observed changes" be "embedded in the matrix of linguistic and extralinguistic" structures (Weinreich, Labov and Herzog 1968:101)? This question has kept researchers busy since it was first asked in this decade. A reliable account of change would have to "show the influence of the structural environment upon the feature in question" (1968:172).

Fishman (1966b:425) was dissatisfied with the study of interference per se as the primary concern of Language Contact because it did not relate language choice to socio-cultural settings (1966b:429). Therefore, he proposed the concept of *domains*, "a socio-cultural construct abstracted from topics of communication, relationships and interaction between communicators and locales of communication [...]" (1966b:430). The main premise of the domain argument was to show the gradual "overlap" and "restriction" of "disadvantaged" languages (1966b:433). Previous research supported this view: "the immigrant language remains - for a varying length of time - the language of inter-group activities" while "the language of prestige" dominates "outward-directed communication" (Hasselmo 1961:26).

From this decade on language shift came to have a counterpart, language maintenance referred to by Haugen as a rare event: "Within a single, cohesive social group the use of a single language is the rule and the maintenance of a second language the exception" (1953:5). Fishman (1966b:424) drew attention to the link between the two and contact situations: "change or stability in habitual use, on the one hand, and ongoing psychological, social and cultural processes, on the other" rather than interference as the object of study. The main factors that, according to him, contribute to language maintenance and inhibit language assimilation were as follows: religio-societal insulation; time of immigration; existence of language islands; parochial schools; pre-emigration language mixing experience; former language use. In addition to these there were also other factors such as the educational level or the size of the migrant group that had some impact on the maintenance efforts (Kloss 1966:206). Kloss (1966) thus differentiated between clearcut and ambivalent factors in language maintenance and shift.

3.4.8 Where did the lexicon stand in this decade? Simply, where it did in the previous decade. Phonology and grammar were seen as systems whereas the lexicon was not. So it was more vulnerable to extra-linguistic influences (Hasselmo 1961:23).

More contradiction was on the way. In this view, lexical analysis was quickly discarded on the grounds that it was: "combined with considerable difficulties, and ha[d] little practical value" (Hasselmo 1961:51). The cases of "complete substitution" with appropriate inflections would be considered part of the lexicons of both languages (Hasselmo 1961:51) but what we would do with the rest was open to question. Hasselmo's approach was not really doing the lexicon much good. There was also the view deriving from the standardization argument that for a language to be modernized its lexicon had to expand either through its "own processes of word formation or from extensive borrowing from another language" (Ferguson 1968:228). This point was made in spite of the 'simplification' argument.

At the beginning of the decade, Hasselmo (1961:62-66) used frequency as a criterion to determine the stability of loanwords in American Swedish. His conclusion contributed to the borrowing vs code-switching dilemma (1961:63):

it has turned out that the frequencies of such [limited] switching do not differ significantly from the frequency of borrowing.

Later, frequency of occurrence was criticized on the grounds that it was not enough to work out the bilingual's dominant language. According to Mackey (1965) frequencies were unstable and therefore unreliable when dealing with interference, and Fishman claimed it was the types of use that counted not the frequency of use (1966b:436). Yet, "the loanword count" remained in vogue as "the most common type of measurement of interference" (Hasselmo 1969:129).

3.4.9 Previously used data gathering techniques such as the 'written questionnaire' and 'direct interrogation' were criticized and the description of pictures was added to methodological options in bilingual research (Clyne 1967:22). Earlier in the decade, Hasselmo (1961:24) had used switching, translation, and lexical tests. Some of these tests are still with us today. "Such techniques" however, proved to be "only marginally relevant" to "sociolinguistically oriented" surveys. Fishman (1968a:36) had combined participant observation, interview and self-report techniques with the aim of emphasizing the qualitative aspects of bilingual behaviour (1966b). The other effective techniques were "the culling of information from published sources" as well as "consultation with experts and persons knowledgeable about specific areas of problems" (Ferguson 1966c).

3.4.10 As this decade closed, it was predicted that "the future of sociolinguistics [lay]" in the "explanation of the origin, maintenance, change and loss" of languages as well as "the structure of variation" and the "adaptation of lexical and syntactic" material. Equally important would be "the study of the emergence and adaptation of the phenomena variously identified as code-repertoires, registers, speech levels, speech styles, and the like" (Hymes 1971b:5-6). The next decade would also see a search for the nonrandomness of what was becoming a central

issue in language contact research: switching. With the sociological emphasis on the variation, use and function of language, it was inevitable that functional aspects of switching would come to the fore and even the structural analysis of code-switching would involve questions of functional relevance (Hymes 1968).

The debate on general properties of language, namely *universals*, was also opened in this decade thanks to the works of Chomsky and Greenberg. The problem of the universal constraints on language contact processes would, if it were solved, specify which switches were possible and which were impossible, and predict which switches would take place in particular circumstances. These questions would be partially answered in the third decade as the path set during the first and second decade remained unchallenged.

3.5 Third decade: 1970-1980

In the early 1970s Gumperz and Hymes published the sequel of *The Ethnography of Communication* they had edited in the previous decade (1964). In the preface to this new collection, they wrote (1972a:v):

The present work integrates [...] some major directions of research on the social basis of verbal communication, a subject which has become of considerable interest to social scientists working at both theoretical and applied levels. In recent years most linguists, in their concern with formal methods of description, have concentrated on the internal relations of linguistic signs, ruling out consideration of extra linguistic factors.

Thus, the editors were announcing another decade of interest in social influences on language although they were well aware of the disagreement among researchers as to the sources of data and methods of analysis. The work started in the preceding decade would continue and the intention to bring together linguistic and social levels of analysis would remain:

Today, semantics is once more a legitimate subject, and syntax is seen to raise questions of the status of sentences as acts of speech and parts

of discourse. Basic theoretical problems of the nature of grammar and its relation to speakers' verbal competence are therefore once more becoming relevant. Similarly, questions of the function of language are again receiving serious attention. The importance of language in social problems, especially education and national development, also is drawing linguistics into wider concerns.

Within this framework, language *usage* was subjected not only to *grammatical* but also to *social constraints*. *Acceptability* covered not only *grammaticality* but also the notion of *appropriateness* in a given context (Gumperz and Hymes 1972a:vii). By now it was very firmly established that "structural uniformity of languages [was] largely a matter of the linguist's basic assumptions", and given sufficient attention "all speech communities" would prove "linguistically diverse" (Gumperz 1972:13). Moreover, "this diversity serve[d] important communicative functions" such as "signalling interspeaker attitudes" or "providing information about speakers' social identities" (1972:13). *Speech events*, seen as "units of content or activities", stood at the centre of "the analysis of verbal interaction" (1972:17) and would be analyzed along two dimensions in accordance with the current linguistic practice:

(1) the syntagmatic, involving the temporal ordering of subunits, including allocation of rights to speaking, and (2) the paradigmatic, referring to the selection among alternates within a contextual frame (Gumperz 1972:17).

Both of these originally Saussurean dimensions were to prove useful for the study of code-switching in the long run. Also related to the "allocation of rights" and alternation between varieties, was the notion of *indexicality* (Garfinkel 1972). It had arrived to stay. All this was significant but left one guessing whether it contained enough linguistics. There was no doubt that such a method would provide increased undersanding of social processes though at the cost of leading Language Contact down to the inescapable path. Researchers of language contact had little choice but to work in the sociolinguistic fashion. There were attempts to link sociolinguistic data to the transformational model of the day (Dittmar 1976) but these were not successful. Fishman (1972a) viewed "linguistics as one of many crucial methodological laboratories" sociolinguists needed constant access to

but it was "social history" from which they would "derive [their] worthiest hypotheses, [their] most fundamental substantive problems, [their] basic orientation toward society" (Fishman 1972a:217).

3.5.1 The study of language contact had "now reached a point, however, where the lack of unitary attitude w[ould] retard further theoretical development" (Di Pietro 1970:15). While the contribution of linguists, psychologists, sociologists and anthropologists had "been essential to the growth of the study of language contact" it was now time to move toward a theory "acceptable to all disciplines concerned" (1970:16). "The positing of a body of universals" would be "an important contribution to this end" (1970:16). Consequently, a number of universals were proposed with respect to language contact situations such as "the relative stability" of use of the languages involved, and the direction of change determined by the languages involved (Di Pietro 1970:19-20).

Bilingualism was taking on a new meaning as the linguistic situation in most "unilingual" countries in post World War II Europe had changed. Although European populations had been bilingual throughout history, it was only now recognised that the existence of minority groups led to contact not only between languages but also cultures. The new picture also included Gastarbeiter groups with considerable numbers in Germany and Sweden, arriving mainly from Southern Europe, The Balkans and Turkey. Needless to say, socio-cultural problems between groups manifested themselves in communication situations everywhere, particularly at school, work and administration (Oksaar 1972a:477, 484). Consequently, the focus of the study of bilingualism was now the group rather than the individual, and political issues such as group identification. This shift of focus could also be attributed to the fact that during the previous decades "no general agreement" had been reached "as to the degree of proficiency in two languages or the relations between them" in a bilingual individual, nor as to the techniques employed to measure bilingualism (Oksaar 1972a:480).

The *compound-coordinate* distinction was still causing problems as the Ervin and Osgood model of the previous decade had made "no provision for denotation as distinct from connotation or for emotive meaning as distinct from either"

(Macnamara 1970:28). The model had neither discussed "the problem of selecting an appropriate meaning from among the many meanings of a polysemous term" (1970:28). Likewise, Lambert's approach to compound-coordinate distinction in the 2nd decade was the target of criticism as he had shifted it "to one between early and late bilingualism" (1970:29). Was it useful to have this distinction, after all? Macnamara proposed "employing a distinction based on degree of semantic interference rather than the coordinate-compound distinction" (1970:31). However, a study of the bilingual's semantic system had to handle grammar and would not be valid if "confined to pairs of lexical items" only (1970:30).

So the focus was moving onto the processing issues in bilinguals (Ervin-Tripp 1970). So far, linguists had relied on "native speakers' judgements of grammaticality, paraphrase equivalence, and structural parallelism". The psycholinguist would bring in other criteria: "effects of memory, time to process, mistakes in repetition, and differences between types of performance" (1970:315-6). The types of performance involved were 'imitation', 'interpretation' and 'production'. These would form "the bases for making inferences about mental processes" of bilinguals". Would these processes show differences between young children and inexperienced adult learners (Ervin-Tripp 1970:316; Di Pietro 1971:9)? The questions of *lexical storage* and *lexical processing*, still very topical today, were already being explored. "Lexical storage" had to "contain information regarding selectional restrictions" of single lexemes as well as lexical units. Lexical processing, on the other hand, was still waiting for an explanation as to whether it took place before or after syntactic processing (Ervin-Tripp 1970:329). In this area, also Paradis (1978) and Opler and Albert's (1978) work should be mentioned. In an attempt to reassess Weinreich's tripartite distinction among types of bilinguals; Paradis (1978) compared bilingual aphasic speech with healthy bilingual's interference. Opler and Albert (1978), on the other hand, proposed a monitor system to replace the popular switch mechanism.

3.5.2 The term *interference* was now almost entirely outdated as Haugen himself admitted (1970:224):

The concept that bilinguals show interference in the structure of their languages is wrong because the intermediate norms do not need to be described in terms of interference at all. On this point I am in full agreement with Joshua Fishman's criticism. [...] I agree with him that 'interference' is not a good term. We should introduce the concept of intermediate norms [...].

For some researchers, however, interference was used with respect to "the mistakes that result[ed] from this process" (Di Pietro 1971:6). In accordance with the current trend, it was pointed out that the "ideas of transfer between" languages would be "deepened by the understanding of universals" (Bolinger 1971:ix). Also relevant to this argument was the question of errors although by now it was clear that not all errors resulted from "the contrasts between native language and language being learned" but were "due to factors such as memory retention and the type of instruction used" (Di Pietro 1971:7).

The measures of *integration*, Mackey (1970:202) suggested, could be taken from the *message* or from the *code*. While trying to avoid Saussure's synchronic/diachronic dichotomy, Mackey could not avoid his *langue/parole* distinction. Measuring from the message involved counting the frequencies and observing the range of occurrence of items. Measuring from the code, on the other hand, involved the bilingual's intuitive judgement on a number of tests previously proposed by Hasselmo (1969, 1970). Generally speaking, *resistance to interference* would be stronger in standardized codes compared to nonstandardized ones, and it was also affected by the degree of tolerance that varied among speech communities (Mackey 1970:198).

3.5.3 For a while there had been very little work published "about lexical borrowing" (Scotton and Okeju 1973:872) as the assumptions that the "center of interest [was] the end product, the borrowed words themselves"; and that "lexical borrowings represent[ed] mainly new items to the culture" still prevailed (1973:871). These assumptions had led researchers to consider the "subject of borrowing" "solved or even dead" (1973:871). But there was more to it. The process of borrowing was as important as the borrowed word itself and had to be considered in terms of the socio-cultural context and the group membership of the

individual speaker involved. One of Scotton and Okeju's criticisms is still with us today: "borrowings within the core vocabulary itself are also very common" as are cultural borrowings. This observation has been useful in the development of Matrix Language Frame model (see 3.8).

3.5.4 Until the third decade *codeswitching* was considered part of the performance of the imperfect bilingual who was unable to carry on a conversation in one language. It was therefore ignored by researchers until the early 1970s when Blom and Gumperz (1972) approached it as a skilled performance and introduced a distinction between *situational* and *metaphorical* switching. In this decade, codeswitching was also seen "as the resolution of th[e] linguistic tension" bilinguals suffered from (Hasselmo 1972:261). It was not haphazard but involved "ordered selection" from "a set of options" (1972:261). Yet, switches to be considered had to meet the criterion of *grammaticality* and *acceptability* (1972:265, 273). The unit of analysis in the study of code-switching would extend beyond the sentence and the investigation at discourse level would bring the role of "discourse markers" into the picture (Hasselmo 1970). Consequently, moving from sentences to discourse would lead to the distinction between *inter-* and *intrasentential* code-switching later in the decade (e.g. Poplack 1978:4).

The study of code-switching was influenced early in this decade by a model known as the *Speech Accomodation Theory* developed in reaction to the normative bias in traditional sociolinguistics (Giles 1973; Giles, Bourhis and Taylor 1977). In this model, codeswitching was analysed in terms of interlocutors' reciprocal adaptation in the course of interaction. Accordingly, their mutual relations and the way they perceived these, their motives, attitudes, perceptions and group loyalties would all form the basis for the initial choice of language as well as codeswitching further on.

This decade closed with new work (Pfaff 1979a and b; Poplack 1978, Wentz and McClure 1977) that contributed towards the long debate on *universal constraints*. As in other areas of language contact research also here the factors at play were of two types: linguistic and extra-linguistic. Proficiency in the other language or the lack of it was found to be a relevant factor, both in terms of the type and the

frequency of switches. Also the degree of formality or informality of the environment had an impact on the frequency of switching. Linguistically, there was no doubt that code-switching was "rule-governed" yet "the precise nature of the rules which govern[ed] code-switching" (Poplack 1978) was an open question. It appeared that "surface structures common to both languages [were] favoured for switches" and there would be little "syntactic conflict" (Pfaff 1979a:314).

Nonce-forms had returned. The definition was adapted to the now fashionable level of analysis, discourse: "segments which the speaker formulates anew in the course of a speech event". *Nonce-forms* contrasted with preformulated segments called linguistic routines (Hasselmo 1970:183). Such preformulated units of discourse varied in size from single morphemes, phrases and sentences to longer stretches (1970:181, 200). The switch-specific concept *frame* was also taking its place on the scene: "The frames within which they [switches] can be studied range [...] from phrases (including single words as a limiting type of phrase) to complete discourses" (1970:181). Yet, switching was still used in connection to mixing and borrowing, and the distinction did not seem very clear (e.g. Gumperz 1970). Although there was no consensus as to the degree of adaptation for an item to be considered a borrowing or switch, what the researcher observed as "the complete lack of adaptation" was generally classified as code-switching (Poplack 1978:2).

In terms of the *location of switches* it was difficult to ignore the trends outside language contact. It was the time of deep and surface structures however language contact data did not fit well into the transformations. The reason was obvious: mixed sentences did not always meet the criterion of grammaticalness and even when they did, the structures of the languages involved did not always match up. Both external and internal conditions impacted on the location of switches (Ervin-Tripp 1970:337); the former involved "the language of quotations, boundary markers for greetings, and for arrivals and departures of participants" whereas the latter was to do with "constituent boundaries".

As the study of code-switching was gaining prominence, also the notion of predictability was taking on a new role. If we were to "linguistically describe the different types of code-switching, it was first necessary to define the total

population of possible forms" (Poplack 1978:10). Could we predict which elements in discourse could not be switched or in which environments in discourse switching could not occur (1978:10)?

3.5.5 The similarity between *immigrant languages* and *pidgins* and *creoles* was still on the agenda since such features as zero copula, zero article, zero preposition, lack of inflections and reduced lexicon were observed in both cases (Pfaff 1979b:2). While the end-product differed, "the process by which they reached their present state" would provide language contact researchers with useful insights (Gilbert 1972:238). The term *mixing* was back, well-fitted to the new picture of codes and switches: *code-mixing*. How did it differ from switching and borrowing? According to Kachru (1978:28, 31) it did although by looking at his definition it was difficult to see how: "the use of one or more languages for consistent transfer of linguistic units from one language into another".

With Gumperz's sociolinguistic repertoire model in the previous decade (1964 a and b, 1967a and b), the bilingual's 'two distinct systems' had lost popularity. In the repertoire model, there was one 'single grammar' offering bilinguals all the options to select from. The impact of this view on the notion of *mixed code* was obvious, particularly in terms of the lexicon: the entire lexical resources of both languages would become available to the bilingual in contact situations (Hasselmo 1972:263). The repertoire model was welcomed although it was of little help to explain the mixed code grammar. While the idea of "the emergence of new grammatical categories" in contact situations was still popular, "the lack of structural criteria and measurement techniques" for such categories remained as a major problem (Oksaar 1972a:488-9).

3.5.6 *Contrastive analysis* was criticized on the grounds that it "owed no allegiance to any particular theory" (Bolinger 1971:viii) although it had the potential "to evaluate the postulations and claims of linguistic theory itself" (Di Pietro 1971:1). In fact, "the assumption that there are universal constraints in languages [was] basic to the implementation of Contrastive Analysis" (Di Pietro 1971:3). After all, contrastive analysis was not simply "an exercise in taxonomy"; on the contrary, "the specifics of each language reflect[ed], in some way or

another, universal linguistic properties" (Di Pietro 1971:3). The idea was that in spite of "variations in detail due to environmental factors and historical change" all languages were "alike in many essential ways" (1971:3). It was, however, "a matter of record that most of the existent contrastive analyses ha[d] not been made in terms of explicit universals"; and consequently "ha[d] not reached a high level of explanatory adequacy" (1971:4). In spite of Di Pietro's work early in the decade, it was now common to consider contrastive analysis out of fashion but recognise its contribution in "the identification of potential error sources" that "facilitate[d] the teaching process" (Danesi 1983:215). The reason why contrastive analysis had lost popularity was that it was "initially associated with audiolingualism and its theoretical progenitor, structuralism; and it ha[d] become fashionable to attack anything that smack[ed] of structuralism" (Danesi 1983:217). There were also other reasons for its discreditation such as "the tendency to apply it too inflexibly" and "its tendency to focus almost exclusively on native-language interference" (1983:218). Yet another source of objection was its emphasis on grammar without due attention to the communicative aspect of languages (1983:219).

3.5.7 After two decades of efforts to define and link *synchrony* and *diachrony*, it had become clear that the notion of *change* was elusive:

One cannot help noticing that there is a great deal of confusion in the treatment of interference and linguistic change. Is it relevant to speak of loanwords as interference? If they are loanwords they are already part of a new norm, part of the language of a group. However, interference and the idiolect of an individual belong together first - at first. We have, when describing the cases, to make a distinction between more level of phonemes, morphemes etc. and the concrete one of phones, morphs. Yet a central question, not yet satisfactorily answered is: When does interference become borrowing? (Oksaar 1972a).

Mackey had already given the problem a name: *The Synchronic Fallacy* (1970). With Labov's studies in the 2nd decade, it had become clear that languages changed through "variation and vacillation", even more so under the impact of

bilingualism. The current theories and methods, however, were not sophisticated enough to explain this linguistic complexity based on "the framework fashioned by the fictitious synchronic/diachronic dichotomy" (Mackey 1970:195). The assumption of "a fixed code or norm" that had "led the students of language contact up a blind alley" had to be abandoned.

3.5.8 The *lexicon* was starting to enjoy a better status thanks to the sets of semantic features discussed in relation to the universals debate. It was the start of a shift of "focus from the study of individual lexemes to an investigation of how semantic features are interlaced in a network" (Di Pietro 1971:112). Yet, it was a promising start: "how these universal matrices of semantic features are constructed, and what semantic primes they [...] include is, at this time, totally speculative" but "the years [...] ahead will find great strides made in this area" (Di Pietro 1971:111). Elsewhere, interesting points of a different nature were also made. For example, the observation of preformulated lexical units in switching (Hasselmo 1970); and that a loan item, depending on its position, would constrain the structure of the sentence were an indication that the lexicon was more than what it had been.

Lexical items still travelled freely "different parts of the distance from one language to the other" and "in different directions"; and when they arrived at the destination what took place was *lexical insertion* (Hasselmo 1972:264). *Compounds*, *lexical units* as well as patterns such as Verb+Noun constructions had attracted some attention. *Co-occurrence rules* were observed though not strictly in the sense the term is used today. The emphasis on the 'social meaning' was evident here, too. The point was that certain vocabulary and certain speech styles would co-occur and any deviation would result in socially marked behaviour (Ervin-Tripp 1972:233-235). Additionally, *co-selection* and any combination between items of the languages involved would be subjected to *restriction*. Ervin-Tripp's (1972:213-250) *co-occurrence* then was the interdependence of the social setting and structural alternatives and the independence between horizontal and vertical levels:

Thus lexical selection would be horizontally inconsistent in mixing an informal form into a formal sequence in "You may rest assured that His Excellency is most seriously considerin' your suggestion". A striking case of vertical selection is found in Surinam (former Dutch Guiana), whose two main languages are Dutch and the English-based creole, Sranan. To speak Dutch with a Dutch pronunciation is considered affected but to speak grammatically and lexically perfect Dutch with a Sranan pronunciation is praised. In some circles, however, to speak Sranan with a Dutch pronunciation is prestigious.

While her concern was mainly the notion of register developed in the previous decade (Halliday, McIntosh and Strevens 1964), in relation to alternation rules, the starting point of her analysis was the bilingual's *lexical selection*.

However, there was some hope for the lexicon as the common assumption that it was more susceptible to external influence than grammar was challenged (Gumperz and Wilson 1971). It was shown that the local versions of three genetically unrelated languages in northern India converged grammatically while retaining separate lexicons (Gumperz and Wilson 1971). There was a change of opinion elsewhere, too. Two decades after his detailed study of Norwegian in the US where he confidently preached that lexicon was the most-readily affected part of a language, Haugen (1973:533) seemed to have changed his mind:

The doctrine that grammar is the most tenacious and resistant part of a language and the one by which languages are classified, depends on a rather narrow definition of grammar. It is probably true that bound morphemes are less readily borrowed than free ones and that high frequency (form) words are less readily borrowed than low frequency (content) words, but these are more matters of availability than of structure.

Fishman's *domain* approach had been very successful since the previous decade and the speech forms associated with each domain were still at the centre of language contact research. The emphasis was ofcourse on the status of minority languages in domains like ceremonial events, education etc., and how research on "code-alternation" would contribute to the better understanding of bilingual urban

settings (e.g. Gumperz 1970). More precisely, the question was which of the existing "linguistic alternates [were] used when and under what social circumstances" (Gumperz 1970:132).

At the turn of the decade Mackey (1969b:364) had already drawn attention to the fact that in order to measure interference, items had to be assigned to different *codes*. But it was difficult to establish what exactly constituted each code as "codes themselves [were] not stable and evolve[d] rapidly within the same community while varying in time and space within and between individuals". So by the third decade there were a few definitions of *code* circulating among the participating disciplines.

3.5.9 The methodological option for those working within *the ethnography of speaking* tradition was to take the socio-cultural matrix of language use as the starting point and investigate its linguistic concomitants in detail. The reverse was also possible by demonstrating first the distribution of linguistic features which could not be "understood solely in terms of their internal relationships within grammar" but only when placed "within the broader socio-cultural context in which they occur[ed]" (Sankoff 1974:47-79).

3.5.10 At least the first part of the 3rd decade can be seen as a continuation of the first two decades as the emphasis on the social basis of communication remained strong while in structural terms there was growing dissatisfaction as to how the diachronic and the synchronic dimensions of change could be combined. This decade also saw an increasing interest in bilingual language processing in terms of a switch mechanism. Finally, the debate on the universal constraints started near the end of the decade.

3.6 Fourth decade:1980-1990

By this decade, the fact that language contact had become an important subfield of linguistics led Nelde (1983a:12) to coin the term *Contact Linguistics*. He defined it as "un ensemble de disciplines linguistiques" that was "plus maniable que des termes savants comme sociolinguistique, neurolinguistique, psycho-linguistique,

etc.". Such a term was better fitted at a moment when the image of bilingual communities was beginning to change (Nelde 1983a; Appel and Muysken 1987; Stubbs 1985):

First, the boundaries of particular languages cannot in reality be established. The study of variation within language, of linguistic change, and of bilingual repertoires has led sociolinguists to the conclusion that it is not possible in the real world to see languages simply as fixed structures [...] Second, the issue of who is to be counted as 'a speaker' of a given language is not always easy to resolve. [...] This problem is especially difficult in the case of linguistic minorities where the process of language shift or the creation of a mixed code is under way (Stubbs 1985:19).

In other words, it was recognised that monolingual situations were rare in many communities and the emphasis on language contact research would constitute "a mandate for politicians and educators to develop a policy for multilingual education and introduce it into the school curriculum" (Nelde 1983b:xi).

3.6.1 Language contact researchers had by now accepted the fact that bilingualism was a *multidimensional* phenomenon, embracing a broad range of methodologies and a diversity of assumptions and theories. "No single researcher" however could deal with "more than a few aspects of bilingualism" in a single study. While bilingualism appeared as a dynamic and rapidly developing field, the common view was that it was "still very much in a pre-scientific stage of development" (Baetens Beardsmore 1982:19). Over the past few decades "almost as many definitions of bilingualism" had been proposed as the "scholars investigating it" (Skutnabb-Kangas 1981:81; Baetens Beardsmore 1982). This could "partly be accounted for by the very subjective approach taken" to the study of bilingualism (Baetens Beardsmore 1982:19). The never-resolved problem of dichotomies was still on the agenda and *Simultaneous Bilingualism* was proposed for the acquisition of several languages from birth, as opposed to *Sequential* or *Successive Bilingualism* for the acquisition of the second after the first (Grosjean 1982). The other popular dichotomies was the mainly socio-psychological *Subtractive* vs. *Additive Bilingualism* (Hamers and Blanc 1989) applied to the

situations where the ethnolinguistic minority was forced to put L1 aside in order to develop communication skills in a language of higher status. The consequence would be a certain degree of subtraction of L1 and its cultural accompaniments (Hamers and Blanc 1989), and some tension between the cultures to which the two languages belonged (Baetens Beardsmore 1986:22-3). *Additive bilingualism*, in contrast, applied when the bilingual felt enriched by an additional language "with little or no concern or fear of losing their basic language. The addition could lead to a high level of skill in the second language, making it easy for them to build on the advantages of bilingualism (Hamers and Blanc 1989). There had been other dichotomies of this kind, too. For example, the influential *Compound vs Coordinate* distinction introduced by Weinreich (1953) in the first decade; the dichotomy of *balanced vs dominant* bilingualism introduced by Lambert (1955, Peal and Lambert 1962); *individual vs. societal* types of bilingualism originating from Ferguson's work, as well as the *stable vs. dynamic* types of bilingualism also originating from Ferguson's 1959 article on *Diglossia*.

The major areas of bilingualism to be concentrated on were summarised as follows (Hyltenstam and Obler 1989:2):

1. bilingualism at the various phases of the lifespan;
2. bilingualism under healthy and pathological conditions; and
3. its development and loss.

However, among the multiple dimensions of bilingualism the ones that came to the fore in this decade were psycho- and neurolinguistic as well as socio-psychological.

The question of representation, storage, organisation, access, and processing of languages in the bilingual's brain, and whether they were functionally dependent or independent attracted considerable attention (Kolers and Paradis 1980a). As summarised by Paradis (1987:1): "the basic questions in the neuropsychology of bilingualism are whether the two languages of the same subject have different

cerebral representations and whether the fact of having acquired two languages influences the cerebral organization of higher cortical functions". Until this decade the dominant position was that "all languages of a polyglot [were] subserved by the same cortical locus or loci" (1987:1). There was also "the theoretical position" according to which "all languages share[d] the same linguistic principles" and shared the same underlying principle. Previously offered explanations for the language control of a healthy bilingual included Penfield's (1959) automatic switching system would ensure that when one language was being used the other language would be switched off. Macnamara (1967), on the other hand, had proposed two separate systems, one for production and one for perception. In this model, bilinguals had control of an output switch that enabled them to select a language for speaking or writing, and an input switch that was automatically controlled by the input, i.e. the language being heard or read. According to Paradis (1980:502), however, bilinguals possessed "neither an input switch, nor an output switch, nor a translation switch, nor even a monitor system". Bilinguals rather had one set of experiential and conceptual information or one 'world knowledge' store, and two language stores, one for each language, each connected to the world-knowledge store (Paradis 1980). They would simply decide to use one language rather than another, just as a unilingual would decide to speak or remain silent or use one register rather than another. On the basis of extensive literature review, however, Perelman (1989) concluded that in healthy bilinguals it was never the case that one language was completely "turned off".

Cases of bilingual aphasia had already contributed data relevant to cerebral organization in bilinguals (e.g. Paradis 1978) and this decade would build on previous work, particularly in relation to code-switching. Paradis summarised the previous decade's work as the placement of "the assessment of bilingual aphasia in the context of experimental and clinical studies in the neurolinguistic aspects of bilingualism" (1987:ix). It was observed that while normal bilinguals were able to keep their languages apart or to mix them at will, this ability was lost in aphasic patients. Importantly, differences in the recovery patterns of aphasic patients led to the reconsideration of influences from the context of acquisition and use, other than those used in the 50s and 60s (i.e. coordinate, compound, subordinate) (Paradis 1983).

The recognition that it was "not sufficient to assess polyglot patients' language deficits in only one of their languages" had led to the development "of a standardized instrument to assess the patient's other languages" (1987:ix-x). *The Bilingual Aphasia Test* would test the bilingual aphasic's pathological language "at all levels of linguistic structure in all modalities of language use" as well as grammaticality judgements and translation capacities in both directions. The observation that "two languages in a bilingual patient [could] be differentially affected by brain damage" and thus had "some separate representation in the brain" (1987:3) led to the consideration of a number of hypotheses current elsewhere in linguistics, too. These were: 1. *The age hypothesis*, "according to which a language learned after puberty" would be "less lateralized because of the difference in maturational states during acquisition". 2. *The stage hypothesis*, where "the second language" would be "gradually lateralized to the left hemisphere" with the mastery of that language. 3. *The revised stage hypothesis*, "according to which the increased participation of the right hemisphere [was] limited to adults at the beginning of the acquisition of a language in a natural environment, but not through formal learning". 4. *The type of bilingualism hypothesis*, where the coordinate bilingual's languages were represented separately "with a greater participation of the right hemisphere than for compound bilinguals". 5. *The context hypothesis*, emphasizing the participation of the right hemisphere "in a second language context than in a foreign context". 6. *The modality hypothesis*, according to which the learning of a second language through reading and writing" would promote greater left hemisphere participation while acquisition by ear would promote greater right hemisphere participation. 7. *The language-specific hypothesis*, stressing that some characteristics of a given language would "foster right hemisphere participation". 8. *The structural distance hypothesis* where structurally very different languages would be "represented more separately, with greater participation of the right hemisphere than with two closely related languages" (1987:5).

3.6.2 While the traditional line of research on interference and integration in language contact situations was continued by some (e.g. Oksaar 1980), the focus of *transfer* research had now moved to the context of language learning (e.g. Gass

and Selinker 1982). The issues investigated in both contexts were very similar such as the predictability of transfers, constraints on transferability, bidirectionality as well as the issues of structural and lexical transfers. Yet they served different purposes within different frameworks. The cognitive framework now in vogue recognised the 'decision-making' capacity of the learner forcing the contrastive analysis framework out of the scene.

3.6.3 Studies of this decade hardly contributed to the investigation of *borrowing* as the basic problems, such as whether lexical borrowing was responsible for syntactic borrowing (Appel and Muysken 1987:167) and whether and how single synchronic phenomenon could be distinguished from integrated diachronic borrowing (Vallduvi 1988), still persisted. Hierarchies of borrowings had not gone out of fashion (Appel and Muysken 1987:171) yet it had become clear that it was not "possible to establish hierarchies of borrowing simply by counting elements in a corpus" nor was it possible to explain the process of borrowing through such hierarchies (1987:171). Whether single-word incorporations constituted a genuine switch, a loanword or some other category was one of the central questions of Poplack's research. She claimed that "borrowing as a process differ[ed] radically from code-switching, and failure to separate data on the two phenomena c[ould] only obscure the conditioning of each." (Poplack *et. al.* 1989). In their view, borrowing was "a much more productive process" which was "not bound by" the constraints on other phenomena, and Weinreich's notion of nonce borrowing was revived within context (e.g. Poplack, Sankoff and Miller 1988) but complicated the issue of distinguishing loanwords from code-switches even further. According to Poplack and associates (1989) the morphological and syntactic role of a nonce borrowing was equivalent to that of an established loanword and the two contrasted with code-switching. However, (Poplack, Wheeler and Westwood 1989) this distinction posed problems in a language pair as typologically distant as Finnish and English. Poplack and colleagues argued that grammatical incompatibility between two languages made switching difficult. In such contact situations, nonce borrowing instead became more prevalent.

3.6.4 This decade was the high time of the constraints debate in the study of *code-switching* that can be summarized as follows (Appel and Muysken 1987:121):

1. An early stage in which grammatical constraints specific to particular constructions were focused on.
2. A stage which has produced the classical studies in which universal constraints on code mixing were explored around 1980.
3. The present stage, which may be characterized by the search for new perspectives: What alternative mixing strategies are there and are constraints perhaps relative to a particular strategy?

Poplack's (1980) work at the beginning of the decade placed emphasis on the issues of syntactic equivalence, grammatical well-formedness, the supposed link between the degree of bilingual ability and switching. None of these arguments, however, proved fruitful in the long run. The idea of syntactic equivalence was about the use of constituents of one language at one point and those of another at another point as long as the order of these constituents was shared by the two languages. This was called the "equivalence constraint" (Sankoff and Poplack 1981). Also of importance was Poplack's Free Morpheme Constraint (1980) that did not allow the switching of bound morphemes. And for bound morphemes to be switched, the condition of phonological integration had to be met.

A number of other constraints on intrasentential code-switching were proposed by various researchers (e.g. Joshi 1984, Sridhar and Sridhar 1980, Woolford 1983). These researchers fell into two groups: those who maintained that the assignment of code-switched sentences to one of the participating languages was not necessary, and those who held it was. Their respective decisions were based on the language pair they investigated, whether they were typologically similar or not. Poplack worked with Spanish/English data and concluded that code-switched sentences stemmed from a mixture of phrase structure rules extracted from two languages. She argued that phrase structure rules of the participating languages

could be freely mixed in the construction of the tree structures of such sentences. Consequently, Sankoff and Poplack (1981) argued that language assignment of code-switched sentences was impossible and inappropriate, partly because switches took place between constituents whose order was shared by the two languages. Woolford (1983) also found it unnecessary to assign a language to code-switched sentences. However, according to Sridhar and Sridhar (1980), who worked with Kannada/English data, there was a basic language to bilingual discourse and guest elements in intrasentential code-switching would obey placement rules of the host language, regardless of their own internal structure. Joshi also claimed that in his examples of Marathi/English sentences Marathi was the *matrix language* (1984) as it was the dominant language in the community. He then derived these mixed sentences from Marathi sentence structure by establishing rules of switching. He proposed an "asymmetry constraint" according to which switching would take place only from the matrix language to the *embedded language*. Language assignment also seemed possible in the case of Japanese and English switches that took place between constituents whose order was not common to both (Nishimura 1986). Yet, Sridhar and Sridhar (1980) and Joshi (1984) claimed that language assignment would proceed only in a particular direction as opposed to bidirectional switching that took place in Japanese/English sentences (Nishimura 1986). The question of grammatical constraints also led researchers to consider a metagrammar (Sridhar and Sridhar 1980) independent of the grammars of participating languages.

The universality of constraints and some assumptions underlying them were challenged towards the end of the decade as it was argued that evidence against all the constraints were proposed so far and called for the study of "a typologically very wide range of language pairs" (Clyne 1987:762).

3.6.5 *Code-mixing* in healthy bilinguals as such provided "an important source of data on bilingual language processing" but mixing in bilingual aphasics would "offer a direct window on the mechanism of interaction between language systems" (Perecman 1989:227). A comparison of language mixing in bilingual aphasics with healthy bilinguals showed that both groups "mix[ed] languages in much the same way" yet it was difficult to "determine whether a particular case of

language mixing in aphasia observe[d] the constraints on normal language mixing" (Perecman 1989:232).

3.6.6 As mentioned above Poplack and associates (1988) claimed that the *typological characteristics* of the languages involved would have consequences on code-switching patterns. In typologically different languages, incompatible word order would result in code-switched sentences violating the patterns of the participating languages. Hence, bilinguals under these conditions would resort to nonce borrowing. Speakers of typologically similar languages, on the other hand, such as French and English in Canada (Poplack 1985) and Spanish-English in New York (Poplack 1980) would have code-switching available to them although the use of this strategy depended on the social value given to language mixing in the community.

3.6.7 In an evaluation of the developments around the notion of change since his co-authored article (Weinreich, Labov and Herzog 1968), Labov (1982) underlined the centrality of two principles as the basis of "empirical foundations": 1) *heterogeneity* as the normal condition of the speech community, distinguished from variation in that it involved "many alternate, semantically equivalent ways of saying "the same thing"; 2) grammar as "the system of communication used in social interaction" (Labov 1982:18). What stood in opposition to these principles were the assumption of *homogeneity* and the notion of *idiolect* used in synchronic description (1982:19). However, the sociolinguist's "temporal perspective" suffered from "shallowness" and "unless they obtained additional time depth from historical records of their own speech community" their interpretation would "rest on shaky ground" (1982:20). Likewise, the search for the universals of language change could not be complete without historical generalizations (1982:21). Labov (1982) however noted that the problem of causality was still far from a solution; as was the question of which internal or external forces motivated a specific change at a given time and place. Later in the decade also Thomason and Kaufmann (1988:13) observed that "Language contact research ha[d] so far failed to produce a theoretical model" that could predict "what types of conduct-induced changes w[ould] occur when".

During 1980s, the investigation of language *Loss/attrition* would develop into a subfield of language contact (e.g. Lambert and Freed 1982, Weltens, de Bot and Van Els 1986). This development embraced healthy and pathological language loss as well as the factors influencing the process. Various approaches have been employed: psycho- and neurological (e.g. Sharwood Smith 1983), sociolinguistic (e.g. (Dorian 1981) and linguistic (e.g. Schmidt 1985). What were the determinants of loss? Case studies of guest workers in Europe had mainly concentrated on the movement from an L1 to an L2 environment as well as the lapse of time and some demographic aspects such as age, gender, educational level, and social class of the migrant. Two dimensions interested language contact researchers in particular, that also involved the question of intra- and intergenerational loss (Van Els and Weltens 1989):

1. Loss of native languages by immigrants (loss of L1 in an L2 environment).
2. Second language loss by ageing migrants (loss of L2 in an L2 environment).

3.6.8 Interest in the psycho- and neurolinguistic issues was starting to change the status of the *lexicon* in linguistics in general and in language contact in particular. An unresolved issue in contact situations was that the bilingual speaker would often come up with new words. However, Levelt showed that the generation of new words was equally common to monolingual speakers although "the degree of spontaneous new-word formation during normal speech varie[d] drastically between languages" (1989:185). This was an important insight. That is, in terms of new word formation there was a considerable difference between for example an English speaker and a Turkish speaker, Turkish being an agglutinative language. Supporting previous research by Hankamer (1989), Levelt claimed that speakers of agglutinative languages would have a "strongly developed processing component dedicated to lexical encoding, which produce[d] new words as output" (1989:186). Importantly, Kolers and Gonzalez (1980) pointed out that the translatability of words was not necessarily evidence of commonness of mental representation. That is, it was not the words that belonged to common or separate stores or the bilingual who kept them separate or together but it was the experience that was responsible for representation.

3.6.9 In terms of the *methodology*, many of the bilingual memory tests under laboratory conditions were "flawed by inadequate control of the stimulus materials" and their outcome had "little generality for the use of natural language" (Kolers and Paradis 1980a:293). Similarly, language mixing experiments, "where bilingual subjects were required to read or speak linguistically mixed words", were not entirely successful as mixing appeared more frequently under natural conditions (1980a:297). This was an "important methodological problem in the empirical study of code mixing within sentences" also for another reason: language contact researchers had not yet worked out how "a case of mixing" should be distinguished from "word borrowing" (Appel and Muysken 1987:121).

Studies of language loss also suffered from the lack of longitudinal data that would measure language behaviour or competency at one point of time, then repeating that measure at another point of time, ideally on the same subjects. Consequently, most studies of loss were either cross-sectional measuring language behaviour of a group of individuals at one point of time (Lambert 1989).

However, we were not totally without "progress in methods of sampling the speech community, gathering data, measuring change and analyzing the factors that influence change" (Labov 1982:32). The problem with the random sample interviews previously used was that there was no "direct information about the social networks" surrounding "the individual". An important contribution in this area was Milroy's *Language and Social Networks* (1980) that quantitatively investigated the linguistic differences between social and cultural groups in different areas of Belfast in terms of social network structure. This study provided a new methodological dimension to language contact research as data collection through social networks became popular (e.g. Poplack, Wheeler and Westwood 1989).

3.6.10 The developments in this decade can be summed up around a number of basic themes in accordance with the popular cognitive framework of the time: the cerebral aspect of the bilingual's linguistic capacity and the loss of the bilingual's languages under normal and pathological conditions; the significance of code-

switching in terms of the bilingual's control of languages; and the syntactic constraints on code-switching as well as the universality of the proposed constraints.

3.7 Fifth decade: 1990-present

It can be seen from the preceding sections (from 3.2 to 3.6) that there are certain approaches in the study of language contact, or a mixture of these, that have by now become conventionalized. An example is to approach bilingualism from the *individual* or *group* perspective placed somewhere along the bilingual continuum (e.g. Silva-Corvalán 1995b:4). Along these lines, Myers-Scotton (1993b) has proposed a model of codeswitching in terms of the representation of communicative intention. According to the speaker's intention, a negotiation of self would follow in relation to other participants in a conversation. In these approaches, the status of the bilingual situation is also often emphasised as to whether the language use is *stable* or *dynamic*. This sort of argument is usually linked to policy making and the political status of ethnic languages in particular settings (see Milroy and Muysken 1995 for a review). Another common point of emphasis is whether the focus of research is on *structural* or less linguistic aspects of contact (Baetens Beardsmore 1990:1), making an impact on data collection methods. As Silva-Corvalán (1995b:4) summarises:

the linguist's major concern is the structural relationship between the languages in contact, and the manner and extent to which this contact affects the system of the languages involved

as well as, she adds,

the impact of extralinguistic factors.

These two concerns of the linguist hardly reconcile in the study of language change, particularly in contact situations. There are attempts to use an ensemble of various concepts, approaches and techniques so far employed in language

contact research along with those from outside language contact (e.g. Backus 1996). However, it is interesting to note that after years of solid effort to combine sociological, linguistic, psychological and anthropological approaches, the 'integrated approach' is still to be offered.

3.7.1 The question of whether languages change due to internal or external factors has hardly reached a definitive answer in linguistics. Naturally, language contact and bilingualism have always been at the centre of the causality debate, and studies, today, try to reconcile the two:

Studies of language contact and bilingualism give evidence that cognitive, social, and linguistic factors contribute in complex and interactive ways to changes of forms and meanings, supporting a view of language structure to a large extent constrained by cognitive and interactional forces (Silva-Corvalán 1995b:11).

According to Myers-Scotton (1993a:212) contact-induced change results in a turnover of the matrix language. In this process two related language contact phenomena, namely code-switching and borrowing play a major role. It is now "generally accepted that the bilingual is not two monolinguals in one person, but a unique speaker-hearer using one language, the other language, or both together depending on the interlocutor, situation, topic, etc." (Grosjean 1995:261). Consequently, current psycholinguistic research concentrates on the processing of switches and borrowings in terms of "language modes" (1995:261).

3.7.2 *Integration* continues to be a big issue in language contact, today. For example Backus (1996:98-114) still offers an analysis of Turkish-Dutch mixed data along the lines of insertion and alternation. He also looks at the cases of insertion in terms of separate levels of grammar, as has been done for four decades, as morphological, syntactic, semantic and so on. This preference is due to the influence of earlier works as well as the principles laid down in Myers-Scotton (1993a) and Johanson (1993), that have given further support to the still fashionable continuum idea (Backus 1996:383-386). The centrality of the integration debate in Language Contact is critically examined in this thesis.

3.7.3 While lexical borrowing has received a great deal of attention it has not been studied systematically due to the fact that it has been seen as a cultural phenomenon rather than linguistic and that the standard tools of structural linguistic analysis were not suitable to study the lexicon (Van Hout and Muysken 1994). Questions involved in the study of lexical borrowing included (Van Hout and Muysken 1994) degree of adaptation and the types of borrowing. Van Hout and Muysken questioned the validity of establishing hierarchies of word categories in the study of borrowability and proposed other criteria for quantitative measurement. While most studies of borrowing have focused on the main syntactic categories (e.g. nouns, verbs, etc.) and some semantic fields, morphosyntactic categories have recently begun to attract more attention (Van Hout and Muysken 1994).

3.7.4 The accumulation of studies on *code-switching* over the previous decades had led to the current shared view that

code-switching has in fact proved itself to be a litmus test for different interfaces between linguistic, psychological and sociological theories concerning multilingualism (Lüdi 1990:3).

Consequently, a series of meetings were organised at the beginning of the decade (*European Science Foundation Scientific network on Code-Switching and Language Contact*) with the aim of reviewing the available concepts, constraints and models as well as the impact and consequences of code-switching on the bilingual individual and society. It was felt that there was no "need to consider new data" but it was time to examine the current topics "with reference to the copious data which already exist[ed]" (Outline, ESF Basel meeting 1990:7). It was also felt that "a general model was needed which would allow an integration of linguistic constraints, psycholinguistic constraints and socio-pragmatic constraints" and such a model was not beyond the range of European research within the next decade" (Lüdi 1990:3). The question of whether we should "distinguish between predominantly linguistic preoccupations and predominantly social goals in code-switching research" was debated as usual but still left open as "relations between codeswitching patterns" of the individual and the society

would "have implications for linguistic and social theory" (Lüdi 1990:2). As a result of ESF network meetings, that helped reshape many ideas, hypotheses and theories; new models emerged (e.g. Milroy and Muysken 1995; Myers-Scotton 1993a and b). Not surprisingly, "a strong need was expressed to find criteria to interpret correctly language mixing phenomena" that still proved to be difficult after four decades of language contact research. "The notion of the base language", proposed in the second decade, was now considered useful in "speaker-based" studies although it was acknowledged that while it served "analytical purposes", it could not always be identified (Baetens Beardsmore 1990:2). At the Basel meeting, Muysken (1990:15) pointed out that the current constraints debate was "characterised by an attempt to relate two observations":

1. in different contact situations different switch patterns are found;
2. the differences are related, at least in part, to typological characteristics of the language involved.

3.7.5 Towards the end of the previous decade, the question of how to characterize *language mixtures* had been given a new look (Thomason and Kaufman 1988). In this decade, Thomason (1995) building on her earlier research, concluded that the "mixed languages" came about as a result of "unusual social circumstances" rather than linguistic processes. Pidgins and creoles - "which develop always in the absence of full bilingualism", and "in the presence of more than two languages" - have been widely studied as mixed languages. Thomason (1995:16) made a distinction between mixed languages as pidgins and creoles and those arising "in two-language contact situations under conditions of full, or at least extensive, bilingualism". In such situations *time* is "the most obvious non-linguistic correlate" of a further distinction between *persistent* and *new* ethnic groups. While in persistent ethnic groups mixed languages emerge slowly, new ethnic groups develop such mixtures rapidly 'sometimes within a single generation' (1995:17).

3.7.6 The notion of *equivalence*, another target of criticism in this thesis, has not lost its popularity in the study of language contact since Weinreich (1953). As Muysken (1995:192-193) states:

The guiding assumption is that equivalence between the grammars of two languages facilitates bilingual usage, be it second language learning, lexical borrowing, or code-switching. There can be equivalence of categories (lexical elements, phonemes, phrase structure nodes, morpho-syntactic features) or of relations between categories, in structuralist terms. The latter are either syntagmatic (e.g. word-order or agreement rules) or paradigmatic (equivalent oppositions).

The notion of equivalence is bound to stay with us as it is a convincing way of explaining where and when a switch is possible. Muysken (1995:196), in fact, concludes that "switching is possible under equivalence". Also Poplack and Meechan (1995) explore switching where equivalence occurs between the grammars of participating languages as opposed to sites where it does not. This is done, as usual, around the central theme of borrowing-code-switching distinction.

3.7.7 In their preface to *Maintenance and Loss of Minority Languages* (Fase, Jaspaert and Kroon 1992), de Bot and Huebner summarised the development of this subfield of language contact since Fishman's *Language Loyalty in the United States* (1966) as follows: "The last two decades have seen the emergence of a whole field of research" on the "factors that appear to play a role in the maintenance and shift of minority languages" and "how these factors interact in various language contact situations" (de Bot and Huebner 1992:v). However, they recognised the fact that the study of maintenance and shift changed according to "current sociopolitical thinking" and was based on "generally short-term ideologies" (1992:vi). They also acknowledged that attempts to "fill in the matrix of contextual factors effecting maintenance and shift" were "lacking a dimension that [was] able to capture diachronic change among the contextual factors themselves" (1992 v-vi). Fase, Jaspaert and Kroon (1992:4) divided studies in this area into two groups according to the choice of "facts" "used as an index for maintenance or disappearance": those focusing on language use and those focusing on language *proficiency*.

3.7.8 Classification of loans is also still popular in the study of language contact and this seems even more justifiable when such classifications can be, in some way, linked to syntactic change. For example, Silva-Corvalán (1995a) considers the following classes of loan material: single-word loans, single-word calques, multiple-word calques of various types and lexico-syntactic calques. She, then, places these across the bilingual continuum and concludes that multiple-word calques open "the door to syntactic changes in the secondary language" (1995a:265).

3.7.9 Data-collection methods used in language contact today, such as semi-structured and open-ended interviews, various forms of questionnaires and so on have existed since the first decade. It is, however, acknowledged that, different methods of data collection lead to "apparent contradictions [in] the results" (Lüdi 1990:3) and that researchers should distinguish between "grammatical" and "pragmatic" foci of research (Baetens Beardsmore 1990:1-2):

a distinction [should] be made between two major approaches, (a) where the focus is primarily grammatical, and (b) where the focus is more pragmatic [...] Certain difficulties arise out of a failure to distinguish between these two possibilities. Some researchers concentrate on theory-oriented approaches which may or may not use quantitative techniques, yet often the work may shift from a concentration on form to issues which are participant-orientated as the analysis progresses. Hence there need not be mutual exclusivity between pragmatic approaches and quantifiable data as a reflection of predictive factors.

Whether the focus is grammatical or pragmatic, the type of data used in language contact research remains spoken. One of the major premises of this thesis has been that written data should also be considered in the study of bilingual settings.

3.7.10 On the basis of the overviews of the preceding four decades, the conclusion of this section has been that the current issues in the study of language contact (e.g. Silva-Corvalán 1994) are not really new. However, a shift of focus seems to be taking place, although very slowly, in relation to the lexicon in

language contact. This shift was summarised in the concluding remarks made by Larsen-Freeman at the AILA 1996 congress (AILA Review 12:87):

A great deal of attention at the congress was given to lexis. Now I realize that here in Europe many of you have maintained an interest in lexis over the years; however, it was my impression that at this congress, lexis received much more attention than in recent times. Consistent with the trend that I have just pointed out, some of the attention given to lexis came in the form of a blurring of the boundaries between lexical and grammatical elements motivated by the awareness that syntax and lexis cannot be described independently. This is not a new awareness, of course. Halliday has proposed that we speak of 'lexico-grammatical units' rather than attempting to treat either lexis or grammar independently. However, at this congress, there were many other refrains of this theme, especially with the invoking of Sinclair's idiom principle, Pawley and Syder's lexicalized sentence stems, and Nattinger and DeCarrico's form-function composites. Even Chomsky, while still maintaining the distinction between syntax and lexis in his minimalist position, asserts a much more prominent role for the lexicon and claims that knowledge of a language is universal grammar and a language-specific lexicon; thus, language acquisition is in essence a matter of determining lexical idiosyncrasies.

While the lexicon has received considerable attention in the study of language contact, the dynamic view of lexicogrammatical relations provided by the recent corpus research has not played any role in the field.

3.8 To Myers-Scotton or not to Myers-Scotton?

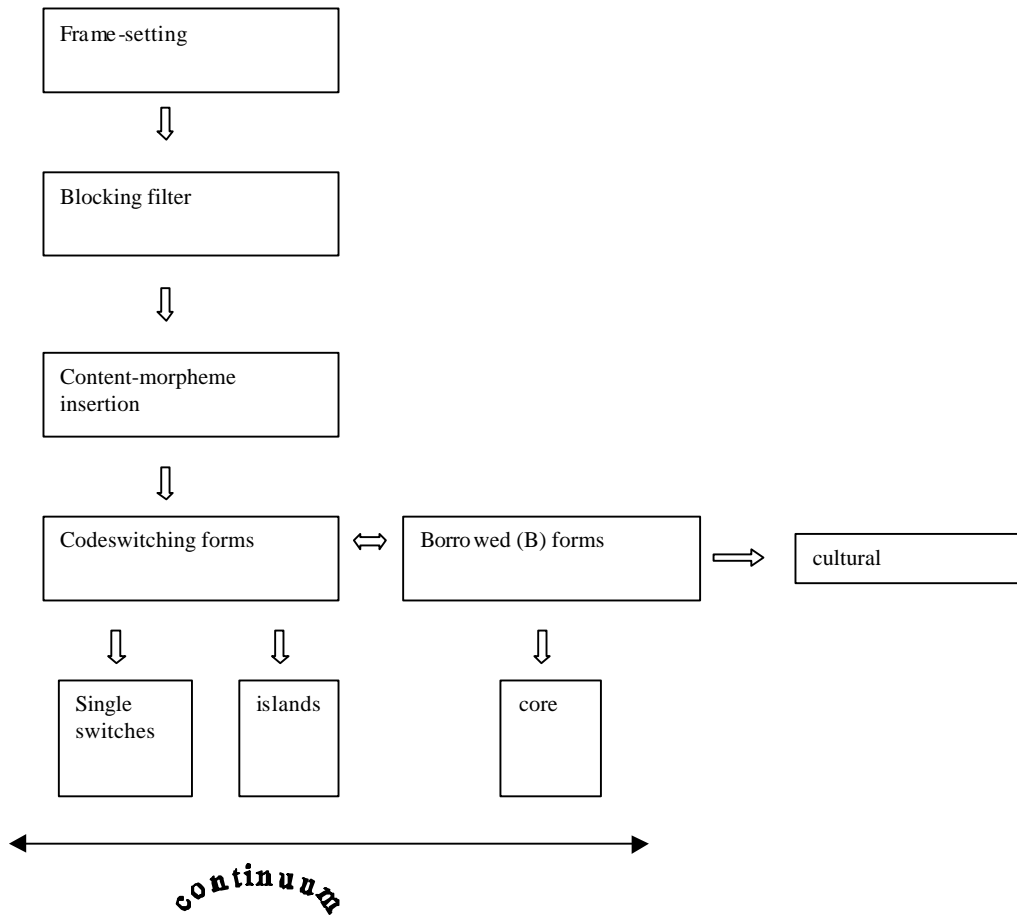
At present, in Language Contact, the major theoretical model is *the Matrix Language Frame Model* (1993a) and associated with it, *the Turnover hypothesis*. It came out at the same time as *Social Motivations* (1993b), as the outcome of Myers-Scotton's life's work. These two volumes came out separately in an attempt to distinguish the social and linguistic levels of analysis, and confirm that

an integrated approach does not always work. I deal with the myth of "integrated framework" in chapter 10.

In the MLF model Myers-Scotton still makes an effort to combine system-based and speaker-based approaches. That is, the changes in the language systems and the way the language user produces them are explained. So both the language as a system and the speaker as the producer and user of that system come into the picture. She draws on recent research in Psycholinguistics and explains codeswitching in terms of psycholinguistic stress. Her main concern, however, remains structural, although the claim is a lexically-based model. In her explanation, codeswitching utterances conform to the grammars of the participating languages though they are subject to the constraints of her MLF model (1993a:2). At the centre of Myers-Scotton's model lies the hierarchy, the language that plays a major role, the *matrix language* and the one that plays a minor role, the *embedded language*. This idea, as many others, dates back to the writings of the first decade researchers: "The real question is whether a given stretch of speech is to be assigned to one language or the other" (Haugen 1956:39). Also the notion of *matrix* makes a powerful return with Myers-Scotton's model. It was used in the first decade (Pike 1956), although with reference to the social context. However, it is surprising to find such familiar wording as "the matrix in which they are embedded" (Pike 1956).

This is one of the controversial issues in language contact, and in particular in Myers-Scotton's model, as there have been and still are strong reactions to the assignment of a base language to switches. Previously, Sankoff and Poplack (1981) argued that language assignment of code-switched sentences is impossible and inappropriate, partly because switches take place between constituents whose order is shared by the two languages. More recently, Auer (1997⁴) questioned the need to establish a base language. Schematically, Myers-Scotton's model, that proposes a code-switching basis for language contact phenomena, can be summarised as follows:

⁴ This issue was also debated at 1st International Symposium on Bilingualism, Vigo, Spain (21-25 October 1997).



While Myers-Scotton rejects, to a certain extent, the distinction between borrowed forms and codeswitched forms, the concern for a differentiation between cultural and core borrowings remains (1992, 1993a). Poplack (1980, 1987; Sankoff, Poplack and Vanniarajan 1990) has maintained for the last two decades that there is a clear distinction between borrowing and codeswitching. Even if such a distinction can be justified, from the point of view of the lexical approach, where the central idea is the co-selection of lexical items, whether these forms represent borrowing or code-switching is of little interest. In other words, the main concern remains their environment, rather than their status in the language. Similarly, the assumption that B forms are part of the mental lexicon and CS forms are not, is very difficult to test. However, MLF model presupposes a relationship between borrowings and switches, and in order to be able to do this, core borrowings, those that already exist in the language, and cultural borrowings need to be distinguished. Also in other approaches there is objection to such a distinction. For example, Duranti⁵ (1997), within an anthropological framework, rejects the distinction between cultural and core borrowings on the basis of his Samoan examples. In his view, what may be a cultural borrowing at the time of first contact gradually loses its cultural content and becomes a mainstream lexical item.

My objection to such a distinction is based on a recent development in *Translation Studies*. The view that core borrowings are used in L1 in spite of their existence in that language assumes one-to-one equivalence between languages. This has been the dominant view in translation, as well, and the prominence of this view dates back to the first decade when for example Haugen (1956:43-44) systematised the establishment of equivalents between languages under the label *Interlingual Identification*.

However, within an approach that starts from the non-equivalence between languages, rather than equivalence, such a distinction loses its significance (see later in the section of translation strategies). Similarly, a lexical item may exist in

⁵ This was also debated at 1st International Symposium on Bilingualism, Vigo, Spain (21-25 October 1997).

both languages but not all meanings of that item will be necessarily available in both languages. In other words, the transferred meaning may be new to the host language.

Myers-Scotton uses a frequency metric as a criterion to distinguish between borrowed forms and codeswitches. This is in agreement with current corpus research methodology. Her cut-off point is 3 as in other corpus studies (eg Bank of English) and items with fewer occurrences are considered switches. However, the notion of frequency is not without problems when the corpus in hand has been compiled in a bilingual setting such as Turkish in Australia. For example, if we accept this criterion to establish what is a borrowing and what is a switch, when we come across a word such as *delivery*, that occurs 7 times in Ozturk Corpus, will be considered a borrowing in Australian Turkish and will be excluded from Myers-Scotton's analysis. However, the exclusion of such an interesting example from the data means losing a good deal of valuable information that would provide useful insights into the internal operations we want to discover. The solution I propose is rather the opposite: to concentrate on co-occurrence. This point is elaborated in chapters 8 and 9. While Myers-Scotton also concludes that regardless of the differences among them, various forms of Embedded Language (EL) material appearing in codeswitching discourse arise from related processes. However, her explanation has a psycholinguistic base.

3.9 Summary

This chapter has proposed a novel type of literature review, that has been organised in terms of five generation's work on language contact rather than placing the emphasis on the so-called language contact phenomena.

CHAPTER 4: CORPUS RESEARCH-AN OVERVIEW

My research has been influenced by the developments in corpus linguistics conducted particularly with reference to lexicography over the last 20 years. For the design and implementation of my corpus, I relied mainly on the corpus work carried out by Sinclair and colleagues in relation to the COBUILD project (see publications from 1987 onwards). It is, however, important to understand how corpus research has taken its present shape and how it can contribute to language contact, in the sense it is used in lexical research today. While various computational approaches have been adopted in language contact before (see Poplack and Sankoff, and Joshi among others), the focus has remained traditionally on sentence analysis and with the narrow focus of distinguishing between borrowing and codeswitching and identifying nonce loans. Although, corpus-based language contact studies can be encountered today, in particular in the area of historical linguistics, such studies are still few.

The use of a corpus for linguistic analysis can be seen as an outcome of the parallel developments that have taken place in various areas of study since 1950s. The common point of interest of researchers in Artificial Intelligence, Computational Linguistics, Theoretical Linguistics, Machine Translation, Lexicography and so on, has been the study of natural language, though with a more peripheral status in some than others. The following section gives a summary of the developments in the fields concerned and their impact on corpus-based research. Section 4.2 outlines the major areas of research in Corpus Linguistics, and the work of Sinclair and Francis whose thinking is central in the present thesis. Section 4.3 looks at the future of corpus studies, section 4.4 summarises the current state of corpus research in Australia and Turkey and finally section 4.5 relates migrant settings to corpus building.

4.1 Background

The convergence of computer science and linguistics in the 1960s, largely stimulated by a desire for *Machine Translation* (see later), led to the development of a new interdisciplinary field called *Computational Linguistics* (CL) whose concern was the use of computers to understand and process human language. Prior to this development, in the 1950s, *Artificial Intelligence* (AI) had been born out of a similar convergence between computer science, psychology and logic, and subsequently influenced the methods used in Computational Linguistics (Stillings 1995, Öim 1989, Leech 1987).

AI started with the aim of developing concepts and methods for simulating various intellectual abilities generally performed by humans on computers such as playing chess, proving theorems, writing music and so on. In the early periods of AI, language processing problems were not much interest to AI-researchers. However, in the early 1970s language understanding moved to the centre of AI research and influential texts such as Terry Winograd's SHRDLU "program that understands natural language" stressed the importance of incorporating problem solving methods into language understanding models (Öim 1989). The central issues in AI i.e. representation, reasoning and recognition have been central also in *Natural Language Processing* (NLP), and NLP researchers realised that AI could offer solutions to these problems. NLP research was carried out first at sentence level and later in limited context, however, little work was done on longer texts until the 1980s (Grosz et al 1986).

In fact, the development of the corpus-based paradigm within the field of Computational Linguistics did not happen immediately. Computational Linguistics also started with a wide variety of aims similar to AI although the development of specific practical systems involving natural language has always been a central area of interest. Three classes of applications, namely *Machine Translation*, *Information Retrieval* and *Man-Machine Interfaces* have been central in the development of CL (Grishman 1986). Work on Machine Translation (MT) started in the 1950s when empirical and statistical methods were popular, and the problems encountered by MT researchers stimulated work both in linguistics and CL, including the development of some of the earliest parsers. Extensive work on MT in the early 1960s resulted in the realisation that fully-automatic high-quality

translation was not possible without fundamental work on text 'understanding', and thus in a cutback in funding (Hays 1976). However, monolingual and bilingual lexical and terminological collections have continued to be used in MT research and the importance of large corpora for the development of grammars and lexica has been recognised (Atkins *et al* 1994). *Automatic Information Retrieval* (IR) from natural language texts also had little immediate success but the idea of extracting the relevant text from a corpus in response to a query stimulated research in knowledge representation. Man-machine Interfaces, on the other hand, have been a successful area of research within CL since the early 1970s. The development of interactive systems that required syntactically and semantically simpler input was particularly needed to allow the user without computer specialisation to interact with systems such as databases (Grishman 1986, Ritchie *et al* 1992).

Theoretical framework for such research was provided by *Transformational Grammar* developed as a basis for MT and led to the emergence of the early parsers. Understandably, small corpora have been sufficient for the purposes of such theoretical work. A corpus-based quantitative approach to the study of language was ignored, especially in English-speaking countries regardless of its ongoing use in such fields as academic lexicography and descriptive linguistics. In the late 1970s CL research began to influence the generative-transformational school and emphasised the need for large textual corpora to explore the so-far-neglected lexicon (Armstrong 1994, Atkins *et al* 1994, Sgall 1989, Grishman 1986).

With the shift of emphasis in the late 1980s and early 1990s from language system to language use, it became obvious that the data extracted from corpora were more complex than was described by the rule-based systems. For example, the traditional parsing technology ignored collocations and word associations as they were too difficult to capture using rule-based systems. More data-intensive methods, on the other hand, offered an expensive alternative, especially for those working in universities in the 1970s (Church and Mercer 1994). Industrial laboratories such as *Texas Instruments* were already engaged in corpus creation in the 1980s and a number of such corpora have since been released gradually for the

use of the research community. The technological advances in publishing also made the use of electronic texts available for academic research, and electronic textual data from a wide range of sources such as newspapers to parliamentary proceedings can now be acquired by researchers (Armstrong-Warwick 1994, Atkins *et al* 1994, Church and Mercer 1994).

Many important corpora existed long before the computer was invented. These corpora were mainly of three types: Lexicographical, dialectological and grammatical (Francis 1992). Large sets of textual data were used in European lexicography as early as the 17th century (Dictionary of the *Accademia della Crusca* 1612) and large corpora were already in use by the end of 19th century to establish word frequencies (Atkins *et al* 1994). However, the era of computer corpus started with the Brown Corpus compiled by W Nelson Francis and H Kuèera in 1963-64, who saw the advantages of making corpora machine-readable for academic research. Officially known as the *Brown University Standard Corpus* of Present-day American English, the Brown corpus consisted of 500 texts of American English drawn from diverse genres, all printed in 1961. It was made available for academic research in the form of magnetic tape first without grammatical tagging, and more recently in the tagged form (Francis and Kuèera 1982), and is now distributed by *the International Computer Archive of Modern English* (ICAME). Its counterpart for British English was the Lancaster-Oslo-Bergen corpus (LOB) with 500 texts of British English. More recently, also the tagged version of the Lancaster/Oslo/Bergen (LOB) corpus was welcomed as a milestone in English corpus linguistics and would open up new areas of research (Meijs 1987a).

The sociolinguistic emphasis on data and the realisation that a corpus would enable linguists to make more objective statements about the language as opposed to the theoretical linguist's introspection and emphasis on ideal language, has been the main reason for the increase of corpus-based projects. Various other reasons for corpus use can also be mentioned: verifiability, diachronic studies, the correlation between relative frequency and register, test-bed for theoretical claims and so on (Svartvik 1992a). Freely available corpora distributed through international bodies

such as ICAME have also increased the popularity of corpus-based research enabling researchers to access language material otherwise difficult to obtain.

Many texts converted to machine-readable form over the past two decades by researchers in humanities have also begun to be collected, organised and stored in data banks at various institutions, and made available for academic research. Data banks of natural language corpora are numerous although the size and coverage vary enormously amongst these. The construction of general language archives have been in progress at various centres including *The Oxford Text Archive*, *Cambridge Literary and Linguistic Computing Centre*, *Humanities Research Centre at Brigham Young University*, *Novosibirsk University Computational Linguistics Centre* then in the USSR and so on (see Allen 1989 for a detailed list). In fact, access to corpora is not so much a problem these days as the quality of the texts accessed.

While corpus creation at various institutions is in progress, the development and improvement of computational tools for the exploration of corpora continues to be a major area of research, calling for closer collaboration between computer scientists and linguistics. There is a growing interest in corpus exploitation tools such as bilingual concordancers and parallel text alignment also in MT for the use of translators (Church and Mercer 1994, Johansson *et al* 1994). Today, numerous corpus exploitation tools are available on the market (see detailed surveys by Schulze *et al* 1994, Christ 1996).

4.2 Major issues in corpus research

Major areas of research within the field of Corpus Linguistics can be examined in three groups (e.g. Oostdijk and de Haan 1994, Svartvik 1992a, Meijs 1987a):

1. Corpus design

2. Corpus Annotation

- a. encoding

- b. tagging
- c. parsing
- e. databases

3. Linguistic exploration of the data

4.2.1 Corpus design

Corpus design considerations differ for different groups of users depending on their research goals. However, the principal aspects of corpus creation and the major decisions to be made can be identified regardless of user needs. Corpus design is mainly concerned with the sampling of texts. A hot issue in corpus design is that of a "balanced corpus" (Sinclair 1987, Atkins *et al* 1992, Biber 1993, see Crowdy 1993 for spoken corpus design). There is no consensus amongst researchers as to the size and representativeness⁶, and many corpus builders select, what seems to them, intuitively relevant without justifying their selection.

While corpus projects usually seek 'to categorise texts according to their primary discourse function without too much precision (Sinclair 1987:12; Rissanen *et al* 1987), Biber's extensive work on the typology of English texts (1989, 1990, 1993) has shown that a thorough definition of the target population and decisions concerning the method of sampling are more important considerations (Biber 1993). Also Gellerstam (1992), considering the ways corpora are used, concludes that representativeness is not always very important. In terms of population definition (Biber 1993, Engwall 1992, Atkins *et al* 1992), the corpus builder needs a sound 'sampling frame', an itemised listing of population members. An adequate sampling frame can then be used to obtain a representative sample. Biber (1993), however, emphasises the need to consider also the non-linguistic characteristics of the target population. Especially in the case of spoken texts, the sampling frame will prove to be more problematic since a typology of spoken texts does not exist. The issue of text length has also been addressed by Biber (1990) in statistical terms. However, as Church and Mercer (1994) point out, there is a difference between the

⁶ I am grateful to Sue Atkins, Patrick Hanks, Pam Peters and Oliver Christ for sharing their views with me on these issues.

two sides of the Atlantic as American industrial laboratories. For example, IBM and AT&T tend to favour quantity (very large corpora) while European dictionary publishers tend to favour quality (smaller corpora with more attention paid to representativeness of many genres).

A major constraint on the development of large corpus projects is the use of copyright material. In the light of rapid technological development in computing and publishing, copyright legislation is being revised. However, the following considerations are usually made in relation to copyright and corpora: is the text protected by copyright, will the payments be offered?, will the corpus be used for commercial purposes?, will the agreement be renewable?, and so on (Atkins *et al* 1992).

Data capture is the first step to corpus creation and optical character readers (OCR) have long been used in corpus projects. OCR first began to be used in the early 1980s, and one of the earliest machines used was Kurzweil Data Entry machine (KDEM). Many humanities computing centres and projects acquired a KDEM or its successor the Kurzweil 4000 (CETH Newsletter 1995). Typical OCR systems today consist of a desktop scanner attached to a PC or MacIntosh, and appropriate software on the computer, that performs the recognition. The text is converted to electronic form with minimal retention of text layout, and font information is usually lost. However, the information on the document layout can be overlaid on the OCR conversion of the scanned text by imposing the *Standard Generalised Markup Language* (SGML), the Text Encoding Initiative (TEI) or other mark-up systems (Foott 1993).

4.2.2 Corpus annotation

A corpus can be accessed in three different forms: raw, annotated or analysed (McNaught 1993). Leech (1987, 1993) defines corpus annotation as the practice of adding interpretative (especially linguistic) information to an existing corpus of spoken and/or written language, by some kind of coding attached to, or interspersed with, the electronic representation of the language material itself. He distinguishes

between various linguistic levels of annotation so far carried out on English corpora:

<u>Linguistic Level</u>	<u>Annotation</u>
Orthographic below)	SGML/TEI (see section 1
Phonetic/Phonemic	Widespread
Prosodic	Two prosodically annotated corpora exist
Part of Speech (ie <i>grammatical</i>) tagging	The most widespread type of annotation
Syntactic (i.e. parsing)	Rapidly becoming more widespread
Semantic	Some
Pragmatic/discourse	Little but developing

4.2.2.1 Encoding (Orthographic Annotation)

The acquisition of texts is the first step to corpus-based research, however, unless the contents are represented in a standardised form, the corpus will be of little use to the research community. While various encoding schemes had been in use until the late 1980s, major international projects such as *Text Encoding Initiative*, have since succeeded in the formulation and dissemination of a set of guidelines to be used in the preparation of machine-readable texts (Sperberg-McQueen and Burnard 1990, Leech 1993, Hockey and Walker 1993).

Depending on the user's needs, a text in electronic form without structural mark-up can be very difficult to use for a number of reasons. It is difficult to cite a passage in it since there is no information on pagination or run a text analysis program without being able to identify where the words appear in the text. The Text Encoding Initiative has addressed such issues in the construction of TEI header that includes information on file description, encoding description, profile description and revision description (Mallery 1995). Recently, also scanners have begun to be

used for encoding at the recognition stage. While they have made a significant contribution to tasks such as part-of-speech tagging and sense disambiguation, annotated corpora are expensive and generally unavailable for languages other than English. Research is moving in the direction of self-organising methods that may eventually remove the need for annotated corpora.

4.2.2.2 Tagging

Part-of-Speech (POS) Tagging is the process of assigning word-class information to the words in a text. Tagging programs based on morphological analysers can work stochastically (e.g. Corpus Linguistics Research, University of Birmingham) or can be rule-based (AI approach). Stochastic tagging calculates the most likely usage of a word by selecting one of its possible part-of-speech interpretations in a particular context.

Example: TAG "Live off the land" --> live + {Verb} off + {Prep} the+{Art} land+{noun}

(From Leech 1987)

Tagging then involves the resolution of ambiguities in context as in the case of *live* (adjective) and *land* (verb) belonging to several classes, and it is also referred to as part-of-speech *disambiguation*. Both the probability of the word belonging to a certain word-class and the probability of the word-class occurring at the specified position in the text are taken into account. Probabilistic tagging systems, however, cannot achieve 100% correctness without human intervention (Sgall 1989).

Information provided by tagging is useful for further linguistic study that is, for analysing the syntactic structure of the text's sentences or for statistical work such as counting the distribution of the different word classes in text corpora. Taggers have also found immediate application in lexicography as the first step to lemmatization.

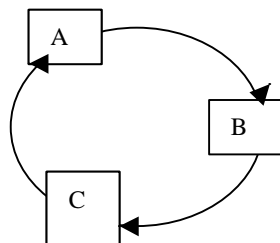
4.3.2.3 Parsing

Parsing has been one of the concerns of CL since the early days and a number of research groups have concentrated on the syntactic analysis of corpora since the early 1970s, e.g. the TOSCA group in Nijmegen (Aarts and Van den Heuvel 1985). Tagging and parsing are closely related and POS tagging can be seen as a basic step to any parser. Parsing systems replace a string of characters with a category label that contributes to the elucidation of the meaning of the text (Karlsson 1992, Sinclair 1992). Parsing a sentence in order to build a syntactic representation for it is of course barely an end in itself. The syntactic structure will usually serve as input to some further processing, typically semantically-oriented (Bennett *et al* 1986).

4.3.2.4 Databases

A database is a representation of information designed to make data entry and retrieval easy. It serves as a repository for large amounts of data and the organisation it imposes provides easy access for researchers. Databases have long been standard repositories in phonetic research, but they are finding increasing use not only in phonology, morphology, syntax, historical linguistics and dialectology but also in areas of applied linguistics such as lexicography and computer-assisted language learning.

Any system for processing large amounts of NL text data must make use of a substantial linguistic database, and the performance of an NLP system depends largely on the quality of the databases it uses. Linguistic databases, on the other hand, must be corpus-dependent since they need corpora-derived frequency data, indispensable for the enhancement of probabilistic systems.



A: process new data; B: enhance databases; C: enhance NLP systems

(From Leech 1987)

Before we go on to the linguistic exploration of corpora, it should be mentioned that this study does not address all the issues of language engineering covered in this chapter. While familiarity with the issues is essential to understand the way the corpus is compiled and organized the emphasis is on the theoretical and descriptive aspects of corpus research as applied to the field of language contact.

4.2.3 Linguistic exploration of corpora

Sinclair (1992) distinguishes between specific and general purposes of corpus analysis. Specific purposes appear when a client with a specific need is interested only in the end result. An example of this can be a spelling checker and the client is interested in its performance rather than its theoretical construct. However, such a device depends on corpus analysis to function properly. General purposes, on the other hand, "are those where the task is so complex that we have to rely on the application of linguistic principles" as in the case of MT, speech recognition and so on.

Applications of corpus analysis in the study of language can be found in Biber (1990) in the investigation of variation. It has also been used to study specialised collections such as legal and medical texts (Biber and Finegan 1994). Corpus material is also exploited at spoken discourse level in Altenberg (1987) on the Tone Unit and its identification, Aijmer (1987) and Stenstrom (1987) on the use of various discourse clues. Biber (1992) also studied the use of anaphora and other referring expressions at discourse level in the Brown corpus. Lexically oriented corpus analysis has been very common over the past two decades due to the revolution in lexicographical methods (Sinclair 1991, Sinclair 1987a, Kjellmer 1987, de Haan 1987).

It is appropriate at this point to take a closer look at Sinclair's work, whose thinking has had a major impact on this thesis.

4.2.3.1 Sinclair

As a result of a major computational corpus project⁷, and a new research methodology, (Sinclair *et al* 1987a), Sinclair has challenged the traditional view on lexis, and placed it in the centre of linguistic analysis. Such an approach handles texts with two principles of interpretation (Sinclair 1991:109-110):

(1) The open choice principle

The open-choice principle is a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed (a word or a phrase or a clause), a large range of choice opens up and the only restraint is grammaticalness. This is probably the normal way of seeing and describing language. It is often called a 'slot-and-filler' model, envisaging texts as a series of slots that have to be filled from a lexicon which satisfies local restraints. At each slot, virtually any word can occur [...] Any segmental approach to description which deals with progressive choices is of this type. Any tree structure shows clearly: the nodes on the tree are the choice points. Virtually, all grammars are constructed on the open-choice principle.

While the language we use is mainly constrained by the nature of our world and the choices of register, there is still far too much opportunity for choice. Consequently, another principle is necessary to account for the restraints that are not captured by the open-choice model.

(2) The Idiom principle

The language user has available to him/her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. To some extent, this may reflect the recurrence of similar situations in human affairs, it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time

⁷ The Bank of English, Birmingham.

conversation. The principle of idiom has been noted by many writers on language, but its importance has been largely neglected.

While language contact researchers realise that the open-choice principle does not fully operate in the context of immigration and the migrant has fewer choices to fill in the slots, the fact that the choices made may be based on another principle does not receive any attention. I shall return to these principles in the final section on corpus analysis.

Sinclair has also questioned the operational value of 'word' in lexical analysis and brought the concept of *lemmatization* to researchers' attention (Sinclair 1991:68-9). Traditionally, the concept lemma would not include semantically related forms and a distinction would be drawn between various forms of a word. Recent corpus research has shown that one form of a lemma usually has a higher frequency than the other forms. I shall return to lemmas in the final section. The notion of lemma here is used in the lexical sense that embraces the related forms of a word as opposed to psycholinguistic notion of lemma (Myers-Scotton 1993a after Levelt 1989).

Another relevant issue discussed by Sinclair (1984, 1988) is the question of *naturalness* versus *well-formedness*. Although, the point he makes concerns another area of language research, specifically teaching, language contact research can also benefit from this discussion. Sinclair is critical of the line of linguistic research that has mainly used well-formed sentences in order to illustrate the workings of a language at the expense of real and natural examples. In language contact research, the negative perception of transfer expressed as 'deviations from the norms' confirms the preoccupation with well-formedness. The fact that the so-called 'deviant forms' are the natural instances of language use has been ignored, and these have been analysed with reference to L1 or L2 rather than in their own right.

An equally important aspect of the lexical research conducted by Sinclair and colleagues is the data-driven approach to linguistic analysis as opposed to data-based approaches (Clear et al 1996). In the *data-driven* approach the linguist

investigates the corpus with an open mind to discover how language really works as opposed to the *data-based* approach in which the linguist first establishes the model and then investigates the corpus to find natural examples to fit into that model. The data, therefore, "is the major informant, providing the raw information" needed for the description of linguistic phenomena (Francis 1993). Importantly, in this approach, the lexicon is considered empty at the start. It subsequently builds up through usage (Sinclair 1996a). The implications of this view for language contact research are discussed below.

4.2.3.2 Francis

The present study draws heavily on Gill Francis's work on lexicogrammar that derives from her detailed observation of the Bank of English corpora (1991, 1993), and provides invaluable insights into the workings of language as does Sinclair's work summarised above. The main premise of her approach is that words have their own grammar that includes usage, colligation, collocation, and other patterns (1991). Francis puts the common supposition that 'grammar and lexis are distinct systems' to the test by investigating the distribution of nominal groups across clauses (1991).

If nouns, in accordance with the traditional view, are capable of occupying any functional slot at any time, then a sample of randomly selected nominals should function sometimes as Subject, sometimes as Object, sometimes as Complement and so on. She observes that not all nouns fit into functional slots freely but "are more frequently associated with some grammatical constituents than with others" (1991). Her analysis challenges the view that "nominal groups can realise different grammatical categories and remain basically unchanged", and thus places the emphasis on the syntagmatic environment as opposed to the paradigmatic axis of choice. Francis's observation, in turn, leads her to examine the relationship between the different senses of a noun and different grammatical categories. She finds that the different senses of the nouns under investigation "are associated with certain strong syntagmatic patterns or co-selection of lexical items". At the end of her detailed analysis of various nominal groups, Francis

comes to the conclusion that "we should stop seeing the verb as being the centre of the clause, and somehow responsible for its whole structure" (1991).

The main point of a lexicogrammar is that "each item or structure is used in its own right, rather than as compared with members of the same or a contrasting paradigm" (Francis 1993). Traditional grammars, including mixed code grammars, have not considered the role of lexis in their search for structural restrictions. "In many cases there is no mention at all of the fact that a structure may be lexically restricted other than by considerations of word-class and related factors" (1993). To illustrate the point, Francis chooses a grammatical structure that is lexically restricted. She examines the behaviour of it when used as the object of a verb followed by an adjective or noun group. She finds that this structure occurs with a very limited range of verbs, mostly *find* and *make* and with a very limited range of adjectives, *difficult*, *hard*, *easy* and *clear* as in:

.....they often find it difficult to.....
.....many will find it harder than usual.....
.....finds it easy to make friends.....
.....I made it clear that I was not angry.....

Francis (1993) points out that such facts about lexico-syntactic interdependence and their functions become obvious through corpus search. She also underlines the fact that these structures are "a blend of colligation and collocation, where the collocational possibilities involve not individual words, but semantic sets of words and phrases" (Francis 1993).

The implications of this view for language contact research are also discussed in Chapter 7.

Stubbs (1996:40) summarises the central conclusions of the work carried out by Sinclair and his colleagues on lexicogrammar as follows:

1. Any grammatical structure restricts the lexis that occurs in it; and conversely, any lexical item can be specified in terms of the structures in which it occurs.

2. Such restrictions are typically not absolute, but clear tendencies: grammar is inherently probabilistic.

3. Meaning is not constant across the inflected forms of a lemma.

4. Every sense or meaning of a word has its own grammar: each meaning is associated with a distinct formal patterning. Form and meaning are inseparable.

5. Words are systematically co-selected: the normal use of language is to select more than one word at a time.

6. Since paradigmatic choices are not made independently of position in syntagmatic chain, the relation between paradigmatic and syntagmatic has to be rethought.

7. Traditional word-classes and syntactic units also have to be rethought.

8. Native speakers have only limited intuitions about such statistical tendencies. Grammars based on intuitive data will imply more freedom of combination than is in fact possible.

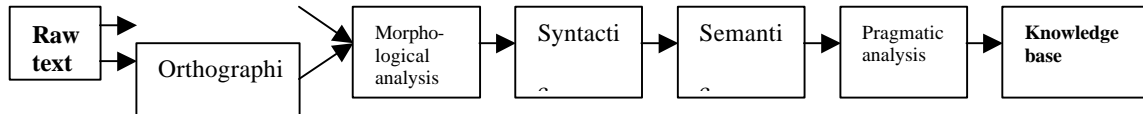
4.3 The future of corpus research

Currently large corpora occupy a central position in linguistic research:

today the development of large reference corpora is one of the main priorities in the European Commission Fourth Framework for Information Technology (IT). What is referred to as "Language Engineering" - the application of engineering techniques to linguistics, in other words the computational approach to language - is a key concept in the development of efficient multilingual communication in Europe (Tognini-Bonelli 1996a:2).

and Leech (1987) summarises the progress in corpus research as follows:

Acoustic
analysis



While most of the work has been done in areas (3) and (4), research in computer text comprehension is directing itself to semantic analysis especially in the area of sense disambiguation. Tagging at different levels is also needed as well as a flexible system of analysis that can move between the levels. Currently, corpus research is in search of many answers that will shape its future (Allen 1992):

- what is a corpus in theory and practice?
- what is language in a corpus perspective?
- to what extent is formalisation possible?
- how to design a corpus or a language bank?
- how to ensure representativeness of samples?
- how to cope with the dynamics required?

Bates and Weischedel (1993), on the other hand, predict that the lexicon and speech processing will receive more emphasis in the corpus-based NLP research. While one-million-word standard size corpora are being replaced by 100-million words as in the case of the *British National Corpus*, Svartvik (1992) predicts that this will be the standard size of future corpora. Data collection efforts over the last three decades have contributed to the improvement of hard and software. This in turn made the creation of larger corpora possible. With the advances in computer technology, international exchange of on-line textual resources and electronic transfer will be facilitated even more.

Sinclair (1992) stresses the need to "devise methods of analysis that prioritise information about language that we can derive from the corpus, and not the vindication of models." A word of warning against blind corpus-dependence also comes from experts in the field (eg Quirk 1992, Svartvik 1992, Johansson and Stenström 1991): intuition should not be neglected at the expense of corpus analysis. In fact, an example of putting corpus analysis and the native speaker's

intuition together can be found in Greenbaum (1984) who used elicitation tests to complement corpus-based analysis.

According to Quirk (1992) there is a certain tension between "(a) those who want to know as much as possible about language [...] and (b) those who want to know as much as possible about what the computer can do". While both approaches are equally valid, and "potentially complementary", they cannot be seen as two ways of achieving the same goal. "How we choose between them, or where we put our emphasis, will affect public policy, the support of funding, and hence the direction and future of our research."

4.4 Corpus studies in Australia and Turkey

A computerised corpus of Australian English did not start until mid 1980s. The unavailability of standard corpora in Australia such as the American and British ones stimulated the initiation of a two-stage project: a) creation of a sample corpus of a million words modelled on Brown and LOB corpora, b) creation of a much larger monitor corpus that will comprise whole texts. Preliminary results based on the Australian corpus can be found in Collins and Peters (1988). As stated by the linguists involved, the aim of the project is to contribute to the "understanding of the relationship between Australian English and the standard English of Britain and the United States" and to "provide many kinds of dialect comparisons" as well as "stylistic comparisons" (Collins and Peters 1988). Since the start of the Australian corpus initiative, a number of machine readable corpus projects have been undertaken at various Australian universities. Computational corpus research with lexicographical emphasis is also underway at the Macquarie Dictionary Centre, Macquarie University.

Computational analysis of Turkish morphology has already been studied by Hankamer (1985, 1988, 1989, and 1993). Similarly, computational applications for Finnish that has affixational morphology similar to that of Turkish, have been developed by Koskenniemi (1983) and provided useful models for computational corpus research on Turkish. In 1993 a NATO-sponsored NLP

project started at Bilkent University, Ankara and the developments include a morphological analyser, a parser, a spelling checker and a text tagger (Oflazer and Bozğahin 1994). However, as the quotation above from Quirk specifies, the current corpus research in Turkey remains along the lines of computer engineering rather than corpus-driven lexical research.

4.5 Corpus-building for language contact research and the uniqueness of textual resources in migrant settings

Very few corpus-based studies have been reported in the area of language contact research so far and they have mainly concentrated on L2 acquisition. Biber *et al* (1994) used corpus analysis to examine the development of discourse competence and register awareness of the adult learners of English. More recently Biber also compared the dimensions of variation of spoken and written registers in English, Nukulaelae Tuvaluan, Korean and Somali through corpus-based analysis (Biber 1994). Similarly, Lux and Grabe (1991) used corpus-based analysis to compare the compositions of university students, written in Ecuadorian Spanish and English. Also in Canada, the acquisition of French as a second language has been investigated using a corpus-based approach (Van der Linden 1994). In this study promoted by *the Office de la Langue Française* the researchers used five different corpora including data collected from the Portuguese community in Canada, as well as other mixed migrant groups. The areas covered in this corpus-based analysis were the acquisition and use of reflexive pronouns, anaphoric use of personal pronouns, the use of auxiliary verbs and the use of text markers. Previously Poplack (1990) and more recently Myers-Scotton (1993a:13) used a computer-based concordance program for an analysis of codeswitching instances in her Swahili-English corpus. Also de Bot and Clyne (1994) mention a concordancing program to report on the L1 loss experience of the Dutch community in Australia although they do not display any concordance lines, nor give any information about the software. This is the general trend in language contact circles.

Corpus-based approach and computer applications are still not widely-used in language contact research although as Biber states, with "the explosion in the

availability of on-line corpora and computational research tools, analyses and applications of corpus-based work should become increasingly common over the coming years" (Biber *et al* 1994).

Ethnic press and information leaflets translated into the migrant's language are valuable sources for corpus research in the study of language contact. Community newspapers document migrants' use of L1 over time and remain to be the only written source available although they have not been widely used in language contact research⁸. Similarly, information leaflets put out by the government and public institutions, in spite of their ephemeral character; deserve better treatment in the study of language contact as translation and contact are closely related. Thus, Turkish Community newspapers and information leaflets printed in Turkish in Australia provide valuable material for such a study (see chapter 7 for details).

4.6 Summary

This chapter has outlined the major developments in the field of computational corpus research since the compilation of the Brown corpus in 1960 and the impact of corpus research in many areas of study. Current work on machine-readable corpus and tools in Turkey and Australia has also been surveyed. In particular, John Sinclair's work on the analysis of the lexicon has been emphasised.

⁸ Haugen (1973:513-514) refers to an early study of German immigrant newspapers by Wacker who concentrated on archaic words and word forms compared to the infusion of English.

CHAPTER 5: The Contribution of Translation Studies

5.1 Migrants and Translation

According to Delisle (1988:19) “bilingualism and translation are closely related: both are products of the contact between languages that occurs when different linguistic groups communicate with each other”. However, both areas suffer as they cut across a number of disciplines “without really belonging to any of them”, and “not being fully understood”.

Migrants obtain a major part of the information about the host country and its culture through translation. Most documents that contain vital information for the migrant in such areas as social security, health, legal services and so on, are made available in a variety of languages other than English for non-English speaking migrants in Australia, and in other countries with regular migrant intake. While translated information plays a crucial role in the life of the non-English speaking migrant in Australia, and anywhere else for that matter, translated texts are never included in language contact research. Interestingly, Baker (1993) points out that translation is not taken seriously in linguistics and related disciplines such as literary theory, cultural and communication studies and so on, and we can add Language Contact to this list, as well.

Baker (1993) also states that according to a recent survey many corpus builders in Europe exclude translated texts from their corpora presumably on the grounds that translated texts are not representative and they might distort our view of the 'real' language. However, with the recent shift of emphasis in translation studies from *source* and *target* texts to the nature of the *translated* text, this negative view has begun to change (see later for similarities between translation and transfer). Also the developments in corpus research and computational technology over the last three decades are currently showing their impact on translation studies (Sinclair *et al* 1996).

Translation comes into the context of immigration also from the point of view of "psychological stress" previously discussed by Myers-Scotton (1993) in relation to code-switching. It has been pointed out that "translating is coextensive with bilingualism" and its emergence as a skill is "coextensive with the ability to establish similarities and differences across languages" (Toury 1995:248). In an immigrant setting then, it can be assumed that with the development of bilingualism, translating ability starts developing, as well. According to Oksaar (1976:303):

The translator has, of course, a somewhat different approach to the dynamic field of language from a bilingual in his everyday interaction, disregarding the fact that he is a bilingual himself. During the process of translating, however, he will often face the same problems as to the exactitude of the message and the means of the language as the bilingual.

Both the migrant and the translator then are constantly under pressure to come up with a *translation equivalent* of new words. However, non-equivalence across languages is a well-known problem. Naturally, the choice of a suitable equivalent in a given context depends on various factors but it also requires the development of various strategies. These strategies and the question of non-equivalence will be discussed in 5.3. One important difference between professional translators and the average bilingual speaker is the "awareness of switching". As Haugen (1953:65) put it "Speakers will often be quite unaware that they are switching back and forth; they are accustomed to having bilingual speakers before them, and know that whichever language they use, they will be understood." although some of his informants were indeed aware of this mode.

Delisle (1988:28, Chapters 2 and 3) defines the characteristics of a successful translator in terms of such cognitive processes as the ability to draw analogies, the ability to interpret texts through analysis and logical reasoning, and the ability to reformulate and re-verbalise. Although little aware of these processes, the migrant develop the same abilities while developing some form of bilingualism in the host society. Clearly, the qualities of a professional translator are not there, but the processes seem to be the same.

5.2 Similarities between studies of Language Contact and Translation

As summarised in the section above, the majority of translation research has been concerned with the relationship between target and source texts and the question of equivalence. Similarly, Language Contact research has mainly concerned itself with the relationship between L1 and L2 although the emphasis has been on the structural make-up of its outcome as opposed to the preoccupation with meaning in translation. Both areas of study, however, have paid little attention to the inner dynamics of the final product in its own right. It is also interesting to note that the principles of bilingual behaviour and the constraints under which it operates have been central to both language contact and translation research but a satisfactory theoretical model has not been offered yet. A turning point, however, seems very near, as pointed out by Baker (1993):

It is my belief that the time is now ripe for a major redefinition of the scope and aims of translation studies, and that we are about to witness a turning point in the history of the discipline. I would like to argue that this turning point will come as a direct consequence of access to large corpora of both original and translated texts, and of the development of specific methods and tools for interrogating such corpora in ways which are appropriate to the needs of translation scholars.

Unfortunately, contextual interpretation of data (ie the verbal context) has not played an important role in Language Contact since the emphasis on the social context in which L1+L2 interaction takes place has removed the focus of the study away from the usage for four decades now (starting with Fishman's work). In Translation, on the other hand, there has been very strong emphasis on meaning, and usage has not been taken into consideration, also because large corpora made up of same types of texts have not come into translation research until recently (see Sinclair *et al* 1996 for a review).

The words *corpus* and *corpora* figure "prominently in the literature on translation", and in Language Contact for that matter, however, "they do not refer to the same kind of corpora" as in computational corpus research. "Corpora in translation

studies have so far been very modest affairs" (Baker 1993) as they have also in Language Contact research, although for different reasons. So far, Language Contact research has dealt exclusively with spoken texts, and until relatively recently, searched manually for instances of transfer.

From the point of view of Language Contact research, the outcome of a translation activity, typically includes transfer, which may also be observed in the form of a switch. In fact, the use of loans in translation is one of the common strategies used by translators, as Baker (1992) demonstrates. Patterns that emerge under such circumstances, it has been observed (Baker 1993), cannot be explained in terms of the participating languages. It is at this point that the similarity with Language Contact research becomes striking and only through computational corpus research that such patterns become evident (see section below on mixed codes). In both cases, there is emphasis on the change of context, to which Fishman refers as 'dislocation' and Tognini-Bonelli as 'displaced situationality' (1996b).

Yet another common aspect shared by Language Contact and Translation is the search for universals. In Language contact the search has concentrated on syntactic constraints whereas in Translation Studies the search for universals has focused on textual features (see Baker 1993 for a review).

5.3 The 'equivalence syndrome'

The notion of "translation equivalence" has been central in language contact and in translation. As early as the first decade, Osgood, Suci and Tannenbaum (1957:154) were concerned with the equivalence question in different types of bilinguals:

compound bilinguals have learned two languages in such a way that translation-equivalent signs are associated with a single set of meanings (e.g., ordinary language courses in schools, vocabulary lists, etc.); coordinate bilinguals have learned two languages in such a way that translation-equivalent signs are associated with a double set of

somewhat different meanings (eg., one language at home, the other at work, or one as a child, the other as an adult, etc.)

Also Lambert, Havelka and Crosby (1958) found that 'co-ordinates' distinguished translation equivalents better than 'compounds', but the hypothesis that they would be poorer translators was not confirmed.

While the myth of translation equivalent has been dominant in the field of translation, it has been questioned. For example, Delisle (1988:45-47) discusses English 'sympathetic' and French 'compréhensif' and why bilingual dictionaries "fail to qualify as translated works" as dictionary definitions are rarely complete. In his words, language equivalence should be understood in terms of a "comparison of potentials, because linguistic signs are, by nature, indeterminate".

Frawley (1984) postulates a third code which, although a "derivative of the matrix information and target parameters", would "take over and establish itself as a valid code" (1984:169). Accordingly, the emergence of a new code in itself is enough reason to question the notion of equivalence. This supports the view that "the establishment of a third text as a fully individuated unit with its own logic naturally leads to questions of good and bad translation and radical and moderate recodification" (1984:172) (see Chap 8).

As the title of this section illustrates (Snell-Hornby 1988:13) the notion of equivalence in translation can be seen as a similar problem to what Mackey (1970) called 'synchronic fallacy' in language contact. The centrality of the concept of translation equivalence in what Snell-Hornby calls "the linguistically oriented schools of translation theory" "shifted the focus of translation theory away from" the other concerns (1988:15). However, as Snell-Hornby (1988:22) concludes, the concept of translation equivalence is "unsuitable as a basic concept in translation theory":

the term equivalence, apart from being imprecise and ill-defined (even after a heated debate of over twenty years) presents an illusion of symmetry between languages which hardly exists beyond the level of

vague approximations and which distorts the basic problems of translation.

5.4 Translation strategies in the absence of "equivalence"

It is known that translators develop strategies for coping with specific types of translation problems that arise particularly in relation to non-equivalence among languages. The problem of non-equivalence has been investigated at different levels (Baker 1992; Barnett et al 1994 in relation to Machine Translation). Non-equivalence at word level and collocational level are of immediate interest to this study. Let us now look at these types of non-equivalence and the strategies used by professional translators when there is a non-equivalence problem to be solved. Language contact research has so far concentrated on the non-equivalence of culture-specific concepts. As the list below shows clearly, there are many other types of non-equivalence that may lead to the development of mixing strategies. The examples below are taken from Baker (1992) and have been supplemented with examples from other sources (eg. Barnett *et al* 1994). As can be seen, these problems of non-equivalence are not necessarily grammatical or in the word order of the participating languages.

Common problems of non-equivalence at word level

(a) Culture-specific concepts.

These may be abstract or concrete concepts. One such example given by Baker is the word *privacy* in English, 'notoriously difficult to translate into other languages' (1992:21). Within the context of immigration to Australia, one such word for Turkish migrants is *benefit/s* as can be seen from the concordance lines below:

ts236 001fa:1	kayýt olunuz hospital	benefits	association manchester unity
ts279 002fa:2	health care veyá	benefit	kartýna sahipseniz kýþlýk
ts279 002fa:2	health care veyá	benefit	kartýnýzy göstererek indirim
ts281 007fa:1	veya 'rates of payment	benefits'	adlý broþirleri isteyiniz
yor294 006fa:1	arayýnýz pharmaceutical	benefits	scheme department of

WOM 001:1	Unemployment	benefit	işsizlik ödeneđi Federal
yv756 004x:13	Yardımyı Kartı Health	Benefit	Card gibi kartlarınyz varsa
yv806 011x:8	Programı Pharmaceutical	Benefits	Scheme PBS kapsamına giren

As the concordance lines above illustrate, this word has different right and left collocates as in *unemployment benefit*, *hospital benefits*, *benefit card*, *pharmaceutical benefits scheme* and so on. It is not possible to translate all the uses of *benefit* into Turkish using one word.

(b) The source-language concept is not lexicalized in the target language.

An example to this type of non-equivalence, the German word *Schadenfreude* ('pleasure in someone else's misfortune'), comes from Barnett *et. al.* (1994). There is no word in English that matches the concept of *Schadenfreude* in German, "nor is there a word for a nearby (up, down, or sideways) concept. In this case, a complete restructuring of the sentence may be necessary (Barnett et al 1994).

(c) The source-language word is semantically complex.

Baker (1992:22) gives *arruaqio* in Brazilian as an example that means 'clearing the ground under coffee trees of rubbish and piling it in the middle of the row in order to aid in the recovery of beans dropped during harvesting.'

(d) The source and target languages make different distinctions in meaning.

Baker's (1992) example to this type of non-equivalence comes from Indonesian, that is the distinction made between going out in the rain with the knowledge that it is raining (*kehujanan*) and going out in the rain without the knowledge that it is raining (*hujan-hujan*).

A similar but less exotic example is also given by Barnett *et al* (1994). In German a distinction is made between knowing a person (Ich *kenne* ihn) and knowing a fact (Ich *weiss* es).

(e) The target language lacks a superordinate.

Baker's example comes from Russian which has no equivalent for the word *facilities* (equipment, building, services etc) but several specific words for types of facilities: *sredstva peredvizheniya* (means of transport); *naem* (loan); *neobkhodimye pomescheniya* (essential accommodation); *neobkhodimoe oborudovanie* (essential equipment).

Similarly, Barnett *et al* (1994) compares English *vegetable* and Japanese *yasai* that includes greens but leaves out carrots and so on.

(f) The target language lacks a hyponym.

Baker (1992) finds that the hyponyms of *article* in English *feature, survey, report, critique, commentary, review* and so on have no precise equivalents in other languages.

(g) Differences in physical or interpersonal perspective.

Baker lists common verb pairs in English such as *come/go, take/bring, arrive/depart*; and the verb *give* in Japanese has six equivalents depending on who gives to whom as in *yaru, ageru, morau, kureru, itadaku, kudasaru*.

(h) Differences in expressive meaning.

In languages words or lexical units may have the same propositional meaning but differ in expressive load (e.g. neutral vs emotionally loaded). Baker compares *wife battering* in English with Japanese *tataku* (to beat) that requires a modifier such as 'savagely' or 'ruthlessly'.

(i) Differences in form.

Productive mechanisms in English producing pairs such as *employer/employee, trainer/trainee, payer/payee* do not exist in Arabic. Such mechanisms are useful to

obtain certain effects in text and the translator needs to come up with a different solution.

(j) Differences in frequency and purpose of using specific forms.

Baker compares the *-ing* form in English that has equivalents in German and Scandinavian languages but is used less frequently to bind clauses.

(k) The use of loan words in the source text.

This point is of particular interest to language contact research as different cultures react differently to the use of loanwords in text. For example Baker mentions *au fait*, *chic*, *alfresco* used solely in English text for their prestige value and add sophistication to the text. This however is lost in translation.

Cases of non-equivalence resulting in the development of strategies by professional translators can be summarised as follows:

(a) translation by a more general word

(b) translation by a more neutral/less expressive word

(c) translation by cultural substitution

(d) translation using a loan word or loan word plus explanation

(e) translation by paraphrase using a related word

(f) translation by paraphrase using unrelated words

(g) translation by omission.

Also of interest to language contact researchers is non-equivalence at collocational level. Baker (1992:53-4) notes that "even when there appears to be a close match

between collocational patterns in two languages, they may not carry the same meaning. For example, *to run a car* in English means 'to own, use, and be able to maintain a car financially' and in Greek, the literal translation for *run a car* is 'speeding'. Hence, non-equivalence at collocational level may result in:

1. the engrossing effect of source text patterning (see 8.1.5.1)
2. misinterpreting the meaning of a source-language collocation
3. the tension between accuracy and naturalness
4. culture-specific collocations
5. marked collocations in the source text.

The emergence such of strategies has also been referred to as a process of automatization: "[T]hey may in fact develop automatized ways of handling specific problems, even a series of fixed solutions which are mobilized whenever a certain problem occurs" (Toury 1995:251). The use of loanwords in language can be explained with reference to such strategies developed under pressure:

such shortcuts seem to form an important part of a translator's acquired ability **to cope with problems in real-life situations, involving, e.g. time pressure, growing fatigue, incomplete knowledge, and much more.** From a certain point on, it may even compensate for incomplete and/or unbalanced bilingualism, which is one possible verification of the claim that an increase in translation skill and language proficiency do not necessarily run completely parallel (Toury 1995:252). [emphasis added].

This, I believe, can be related to the way Myers-Scotton explains codeswitching i.e. psychological pressure. The emergence of new collocations formed with L1 and L2 elements such as in register belgesi or delivery yapýlýr can be seen as a 'fixed solution' mobilized in certain contexts. Another such strategy is to use *denilen* 'called' and the Turkish translation after the foreign word has appeared in the text. This formulation can be seen in the concordance lines below:

ts338 003fa:1....concession card	denilen	tenzilaty kartynyzý tabýmanýz
ts338 003fa:1....aylýk ve multitrip	denilen	biletler de mevcuttur þayet
HEA 003:5...minipill	denilen	hap

WOR 011:7...'take away'	denilen	hazýr yemek çeþitleri satan
yv245 006x:1...konu "Negative Gearing	denilen	fakat Türkçe karþýlýðýný
yv275 003Ac:4...Mr Hepburn register	denilen	bu kayýt defterine aþaðýda
yv543 011x:7...Appeals Tribunal	denilen	kuruluþ var ya itirazda
yv796 003Ac:6...bölgesindeki reef	denilen	kayalýk bölgeler sularýn
yv806 011x:9...Record Form	denilen	bir belgeye iþlenmesi
yv806 014x:20...kiþiler STARRTS	denilen	iþkence ve Travma

The 'impact of environmental feedback', according to which a translator revises his/her production, has also been emphasized (Toury 1995:250). This idea is relevant to the language used in community newspapers. Community's response to certain constructions may work as an external monitoring device for the editors.

5.5 Summary

This chapter has dealt with the similarities between Language Contact and Translation Studies in terms of their search for explanations in the participating languages, bilingual behaviour, the question of equivalence and the common strategies developed by bilinguals and translators under stressful conditions. It has been claimed that both areas would benefit from computational corpus research.

CHAPTER 6: Turks in Australia

6.1 Turks in Australia

Currently, migrants from as many as 160 different countries live in Australia and four out of ten Australians are migrants or the children of migrants (Borg 1998). The Turkish community has officially been part of this multicultural mosaic for the last thirty years. However, a small number of Turkey-born persons had already been living in Australia before assisted migration began. The section below looks at the Turkish community in Australia before and after 1967, the year in which the Turkish and Australian governments signed the bilateral agreement. In the second part of this chapter, attention is paid to language issues and how language needs are met.

6.1.1 Before 1967

As part of Australia's post-Second World War immigration program, a bilateral agreement was signed with the Turkish government in 1967, and the first families to come from Turkey under the scheme arrived in 1968. Prior to this agreement, the Turkish-speaking population in Australia was 2476. These were mainly Turkish Cypriots who held British passports (Manderson 1988, Ýçduygu 1991:84). From 1950s onwards, Turkish Cypriots were allowed to migrate to Australia "because of their service in the British Armed Forces during the Second World War", and before the bilateral agreement was signed there were over 10,000 Cyprus-born persons, 16 per cent of whom were ethnic Turks (Ýçduygu 1991:84-85). However, this was the era of "White Australia" policy that aimed at a "homogeneous English-speaking Anglo-Saxon culture" and very few facilities were provided for new migrants (Romaine 1991). Consequently, Cypriot Turks arriving in this period were assimilated into the mainstream Anglo-Australian society.

Unfortunately, census data concerning the number of Turks in Australia in 19th century are not precise. Manderson (1988) reports that the Turkish community in Australia counted 20 in 1891, and had grown to 322 by 1911. After the First World War this number declined and in 1947 there were 252 Turks in Australia.

There were also other Turkish nationals resident in Australia who were not ethnically Turkish:

In addition to this small Turkish-born population, other Turkish nationals were accepted as residents and naturalised as Australians, including people from present-day Turkey, from elsewhere in the Ottoman Empire (particularly Egypt), and from other regions that included the United Kingdom, South Africa, the Americas, Asia and the South-West Pacific [...] For the most part they were Jews, Armenians, Georgians and especially Greeks who had been born in, or were nationals of, Turkey (Manderson 1988).

These numbers, however, may not be exact due to the fact that censuses before 1986 did not take ethnicity into account and migrants were classified by birthplace. Detailed information on ethnic origins of various migrant groups became available with the 1986 census. For example, Batrouney (1995) claims that amongst those who came to Australia "as part of the great Turkish migration in the sixties and seventies" was a considerable number of Kurds and they "were not willing to [...] identify themselves as Kurds". The exact number of Cypriot Turks is also difficult to obtain, although unofficial sources record 20,000. Unfortunately, information released by the Australian Bureau of Statistics (ABS) relating to Turkey as a separate category did not become available until 1980 (Elley 1985:39).

Large-scale migration from Turkey had not been considered until immigration figures from Western and Southern Europe started to decline in 1960s. Previously, the numbers of Turkish-born migrants were small in relation to the overall migrant intake:

Table 1

year	assisted settlers	unassisted settlers	total settlers	settlers departing	total gain	Turkish Settler Arrivals as a proportion of all Settler arrivals (per cent)
1959-60	4	52	56	2	54	0.05
60-61	5	65	70	2	68	0.06
61-62	0	73	73	3	70	0.09

62-63	6	87	93	5	88	0.09
63-64	6	165	171	6	165	0.14
64-65	3	175	178	8	170	0.13
65-66	22	247	269	7	262	0.19
66-67	40	276	316	11	305	0.23
67-68	190	322	512	7	505	0.37
68-69	2088	186	2274	13	2261	1.29

Source: Survey of Turkish Workers. Department of Immigration, Canberra.

The bilateral agreement was signed in 1967 and between late 1960s and mid 1970s migration from Turkey to Australia reached its peak. By 1986, the Turkish community in Australia had grown into a 'middle-sized immigrant group' that was the "twenty-third largest among the 38 birthplace groups whose population exceeded 10,000" (Y cduygu 1991:77). Today, migration from Turkey has been reduced to less than one thousand per year, and "in terms of net total migration has been close to zero since 1981" (Price 1989:22). The table below shows the increase of the Turkish population since 1901.

Table 2

Year	Number	Annual Growth Rate
1901	200	-
1911	322	4.8
1921	185	-5.5
1933	261	3.4
1947	252	-0.8
1954	1036	20.2
1961	1544	5.7
1966	2476	9.4
1971	11589	30.9
1976	19355	10.3
1981	24314	4.6
1986	24529	0.2
1991	27770	

Source: Y cduygu 1991 and BIMPR Turkish Community Profile

In the mid 1960s, the Australian government was concerned to maintain a high level of intake of immigrants, and to recruit from a range of nations in order to prevent the overgrowth of any particular ethnic group. Migration agreements with Spain and Italy had already lapsed and not been renewed. This decline of migration from Western and Southern Europe forced the Australian Department of Immigration to look further to the east of Europe. By 1965 Australia had established an immigration office in Beirut. At around this time the success of Turkish guestworkers in Europe had drawn the attention of Australian authorities to Turkey as a potential source of migrants (Manderson 1988; Elley 1985:56). Immigration from Turkey was also "a significant step in the development of a more heterogeneous immigration program and one that was less concerned with ethnicity, religion and cultural background as considerations in migrant selection" (Manderson 1988).

The Turkish government, on the other hand, was looking for a new destination for "an estimated 900,000 workers" who "were on waiting lists to go abroad, although fewer than one per cent of them were admitted by European countries" (Y cduygu 1994, 1991:74). Labour recruitment agreements with West Germany, Austria, The Netherlands, Belgium, France and Sweden had already reached the saturation point, and Australia would be the next to take large numbers of migrants.

The decision to start assisted migration of Turkish families however, "caused debate and controversy in Australia. This controversy could be seen to have its origins not only in the vision of an assimilated Australian nation, but also in deep rooted stereotypes of Turkey and the Turkish people" (Elley 1985:50)⁹. Commentators on immigration matters, such as Jupp, feared that the arrival of Turkish immigrants could cause conflict with the already well-established Greek community: "the injection of Turkish migrants might [...] disturb the peace of the highly concentrated Greek communities of the major cities" (Jupp 1966:21). Jupp argued against Turkish migration also on the grounds that these were non-

⁹ For a critical examination of the 'persisting negative image of Turkey and the Turks in Australia' see Elley (1985).

Christian people; there was not an established Turkish community in Australia; their customs and traditions were different and their education levels were low:

The Turks in particular would present a grave communication problem and would contain a higher proportion of illiterates than any previous group encouraged to come (1966:21-22).

Moreover, the tendency was to associate Turks with Asia rather than Europe and the general opinion was against the extension of large-scale migration to Turks and other peoples from the Middle East that would have been followed by further migration from Asia (Manderson 1988).

Even those who supported Turkish migration could not help emphasising the differences between Turks and Australians. Bryant (quoted in Manderson 1988), for example, maintained that

the only affinity between ourselves and most Turks is the colour of our skins. In religion, in political view, in social habit, in attitude to their women - the subjection of their women - in their skills and in their language, they have no affinity whatsoever with us [yet] they come into the community and do not cause any real disruption.

Along with some politicians, the *Returned Soldiers League* also supported the agreement with Turkey, regarding Turks as white Europeans.

6.2 After 1967

European countries such as Germany, Austria, France and Belgium started accepting male Turkish guestworkers on a temporary basis in the early 1960s. Family migration was not encouraged and the men were expected to return to Turkey when their contract came to an end. Contrary to what then appeared to be a temporary arrangement in Europe, "the Australian government intended that the Turkish migrants stay in this country on a permanent basis, contrary to the guest-worker schemes operating in Europe at the time". Obviously, this "security of status, coupled with greater financial rewards than those possible in Turkey, was

an attractive proposition" not only for the men but also for the women (Keçeli 1992). However, as shown in various studies (Bağrı́n and Bağrı́n 1993, Ýçduygu 1991, Elley 1985), most Turkish migrants did not consider themselves to be permanent settlers in Australia and had initially planned to return to Turkey after a few years.

In spite of the announcements made by the Australian government that emphasised 'gradual rather than the immediate flow of large number of Turkish workers to Australia', 10427 Turks had arrived between 1968 and 1971, allowing very little time for the development of adequate community services and organisations to meet their particular needs' (Manderson 1988). The Turkish migration started in 1968 on the basis of assisted migration program that ended in 1974. Ýçduygu (1994, 1991) defined three distinct periods of migration from Turkey to Australia each with different group characteristics: 1968-74; 1975-80 and 1981 onwards. The first group of migrants were mainly "village-born, young couples (with young children) with limited formal education and little experience of working in an industrialised setting" (Ýçduygu 1994). As they admitted in various studies themselves, these migrants did not come to Australia with the intention of staying permanently, in most cases they were not even aware of their permanent resident status. The second group of migrants were mainly accepted to join their families and were not different from the first group in terms of their socio-economic status. The third group of migrants, on the other hand, were mainly 'city-born' and 'single university graduates' who were 'relatively younger than the earlier migrants' (Ýçduygu 1994).

Those who arrived in late 1960s and early 1970s found jobs in unskilled and semi-skilled manufacturing sectors. In the mid-70s, however, the manufacturing sector was badly affected by the economic recession and most of the Turkish migrants faced a period of unemployment (Inglis et al. 1992, Isaac 1988). Elley (1988) reports that at one stage 8 husbands out of her 20 informant families were out of work, and in spite of constant search were unable to find a job over a period of two or three years. Deteriorating economic conditions put considerable pressure on the wives who also took up work in factories. Such unskilled factory work was generally performed by migrants from various backgrounds, who shared the

inability to converse fluently in English. As a result there was little opportunity for the Turkish migrant women to develop networks in the work environment (Keçeli 1992, Çevik 1996). Consequently, traditional communication networks, centered around female kinship and friends disappeared, and "missing links with traditional female networks meant that the women inevitably underwent deep psychological traumas including a sense of isolation, confusion and loss of identity and worth" (Keçeli 1992). Only a few among the Turkish migrant women, who had educational qualifications, previous work experience and English speaking ability, managed to find work outside the manufacturing sector such as teaching (Elley 1988). However, the assessment of English speaking ability for employment could also be affected easily. For example, a Turkish woman who took the same language assessment test twice, first with and later without a headscarf, was given considerably higher scores in the second test¹⁰.

Elley's informants stated that in the early days after their arrival there were no Turkish-run businesses where they could shop, very few Turkish-speaking interpreters, no Turkish newspapers, no Turkish radio programmes and no community network (Elley 1985), that led to isolation at the early stage of migration:

Turkish migrants generally came to this country as nuclear families and from various regions of Turkey. Due to their relatively small numbers, as compared with other migrant communities, they settled in isolated pockets spread across Australian cities. More often than not, the population in any given Melbourne suburb, represented diverse geographical Turkish origins and customs, and at times social norms and behavioural patterns. Such diversity exacerbated the sense of loss occasioned by leaving behind the traditional female networks and familiar village environment. Feelings of reticence and suspicion as regards forming new networks, resulted. Accordingly, any newly formed associations generally failed to provide the traditional source of emotional and physical support reciprocities, information easily comprehended and avenues for female influence and power. The husbands' participation in such associations merely emphasised such losses (Keçeli 1992).

¹⁰ Miriam Faine, personal communication (1995).

The lack of a community network led to another major problem experienced by the early groups of Turkish migrants, that of childcare. Childcare centres with Turkish speaking staff were not available until late 1970s and many Turkish migrants had to send young children back to Turkey to overcome these problems. A decade after the arrival, between 1978 and 1983, a few childcare centres were opened in those areas of Melbourne with high Turkish concentration. These centres, funded by the federal government, employed English and Turkish speaking staff (Elley 1988).

These difficulties faced by the early migrants forced the Australian government to tighten migration selection procedures. By 1972 services to Turkish migrants had increased including a pre-embarkation orientation course and counselling staff (Manderson 1988). When it ceased in 1974 assisted migration had brought 19,000 Turks to Australia who were mainly employed as factory labourers. After a decade, Price (1989: 54) found that Turks were amongst the few migrant groups who "were consistently having the hardest time in mid 1980s." Turks were seen as incapable of doing much better due to high levels of unemployment (27 %), low income, lack of appropriate skills and qualifications as well as poor English-speaking ability. Only 4 per cent of Turkish men were found to earn over \$26,000 p.a. Commentators such as Jupp shared the same view and claimed that "this intake [of Turkish people] produced many social problems, including high rates of illiteracy and unemployment and low levels of citizenship" (Jupp 1991:78). Ýçduygu's (1994) detailed study of Turkish migrants in Melbourne explored the reasons for such low levels of citizenship and downward occupational mobility. He argued that

the majority of Turkish migrants (particularly in the early period of settlement) saw their stay in Australia as a temporary sojourn, they did not have a long-term interest in their work and this affected their attitude towards their occupational status and progress. In essence they did not think they would be around long enough to make themselves occupationally and socially comfortable in their work. When their settlement intentions shifted to permanent stay, they had already spent several years in unskilled, dangerous, unhealthy and

unsatisfactory jobs, and a large proportion had even quit the paid workforce. After an average of twelve years in Australia, their occupational gains were less than modest. Aspirations to return to Turkey in improved economic circumstances or (later, after accepting the fact that they were unlikely to return to Turkey) to occupy a relatively high economic or social position in Australia had not been fulfilled.

The very early members of the second generation also followed the steps of their parents. Elley (1993) found that they left school early, worked in unskilled or semi-skilled occupations and also married early. They also suffered from the lack of an established Turkish community and appropriate resources. In addition to these factors, their parents were also unfamiliar with the host society and "reluctant to allow their children to mix with non-Turkish persons" (Ýçduygu 1994). As they intended to return to Turkey in the shortest possible time, they did not make an effort to socialize and 'be incorporated into the new society' (Ýçduygu 1994). Elley (1985) reported that the children who "entered school soon after arrival [...] had difficulty in learning English or that the schools they entered were not prepared for what had been a sudden influx of Turkish migrant." Turkish migrants were seen particularly disadvantaged and the problems they encountered were attributed to low levels of English, low level of education, origins in rural Turkey and religion (Inglis et al. 1992). Turkish children were also perceived as members of a particularly disadvantaged migrant group and suffered from teachers' stereotypes at school (Inglis and Manderson 1988). Various reports showed that "teachers were not adequately prepared to work in schools of this kind [of high migrant density] and there was much indifference and even intolerance towards migrant children" (quoted in Price and Martin 1976:58). Akçelik and Pappas (1988) also confirm that cross-cultural awareness courses, highlighted by 1978 Galbally Report, for professionals working with immigrants, have only helped strengthen such stereotypes.

6.2 Language Needs

This section investigates two language issues with reference to Turkish migrants:

- a) The place of Turkish amongst LOTEs in Australia;
- b) English language proficiency among the Turkish migrants.

a) Turkish amongst LOTEs

By the end of 1960s, "the level of public and official discussion of the question [of the education of NESB children] was uninformed and there existed no coherent, comprehensive statement of philosophy, policy or practice" (Price and Martin 1976:47). There had been little support for community languages and few teachers with the knowledge of migrant languages. Understandably, community languages were maintained in the home domain and in those domains determined by the ethnic communities themselves (e.g. ethnic social life, religion, ethnic press, ethnic part-time schools, certain aspects of the transactional domain - continental shops, cafés, guest houses (Clyne 1991). As part of multiculturalism policy in 1970s, there was some degree of recognition that migrant languages were a potential force within the otherwise declining field of language education (Ozolins 1993:85). By mid-1970s there was also considerable pressure from various ethnic groups on schools and universities to run courses in their respective languages. As a result of such pressures several committees were established and "the one most directly concerned" was "the Committee on the Teaching of Migrant Languages in Schools, established 1974" (Price and Martin 1976:48). Over the period 1972-78 the Working Party gave priority to the development of a number of LOTEs that gradually gained a more permanently structured place in the education system and tertiary education. Turkish was one of the languages that had been given a place in the period of 1976-8 (Ozolins 1993:132-133).

However, a survey of secondary school teachers¹¹ found that for many migrant languages including Turkish, there was a considerable shortage of qualified teachers (Ozolins 1993:138). Not only migrant "teachers trained in non-English speaking countries" had "found it extremely difficult to gain registration in Australia" but there had also been no "opportunities for retraining" (Price and Martin 1976:53). In 1975 the Turkish government sent five primary school

¹¹ This survey was conducted by the Committee on the Teaching of Migrant Languages in Schools (CTMLS) which was established in 1974.

teachers to teach Turkish to the children of Turkish migrants in parts of Sydney and Melbourne with very high concentrations of Turkish migrants. This was a stimulating move that led to the formation of an association of Turkish teachers qualified in Turkey. In 1983 the Ethnic Turkish Teachers' Association of Victoria appointed some 30 teachers to 16 ethnic schools (Elley 1985:178).

Since 1983, ten primary schools have established Turkish Community and Bilingual programs as part of the Victorian government's specific education initiative (Sayar 1988). In spite of an increase in the number of Turkish language programs in government primary schools, there has been a decrease of such programs in secondary schools. In compensation, attendance in Saturday Turkish classes offered by the Victorian School of Languages has been increasing, and Turkish has been one of the 6 languages with the highest enrolments.

Table 3

	Government School Students at the VSL						Total
	Year 7	Year 8	Year 9	Year10	Year11	Year12	
Vietnamese	79	118	124	151	286	352	1111
Turkish	154	145	147	151	173	167	937
Chinese 105	105	91	94	115	228	196	829
Greek	51	72	73	55	69	62	382
French	30	39	34	41	91	98	333
Arabic	35	43	50	38	65	79	310

Source: LOTE in Government Schools Report 1994

Sayar attributes this increase to the students' intention of presenting for the VCE Turkish examination. In fact, the number of VCE enrolments for Turkish has increased from 84 in 1986 to 187 in 1994 (LOTE Report 1994:37). There are also Turkish community organisations that offer courses to some 1500 primary school age children in Victoria. Inglis and Manderson (1988) attribute this demand to the fact that Turkish parents recognise the importance of bilingualism, even though they may lack it themselves, as an important prerequisite to mobility in Australia for their children.

A considerable increase in the participation in the mainstream schools has also been observed from 3 in 1991 to 89 in 1993 (LOTE Report 1994:38). Unfortunately, the LOTE Strategy Plan required by the Commonwealth Government's Language and Literacy Policy in 1991 in relation to Commonwealth funding to the states, led to the implementation of a new LOTE Strategy Plan by the Victorian Directorate of School Education and Ministerial Advisory Council on LOTE. It was established in 1993 that the Directorate should concentrate its efforts on eight languages in mainstream schools. According to the plan, there are 4 groups of LOTE in Victoria to receive support from the Directorate depending on their category. These groups are as follows: *Key languages in mainstream schools* (Chinese, French, German, Indonesian, Italian, Japanese, Modern Greek and Vietnamese), *Languages for priority development* (Arabic, Korean, Russian, Spanish and Thai), *Languages of particular community significance* (Auslan, Croatian, Hebrew, Khmer, Koorie Languages, Macedonian, Maltese, Serbian and Turkish), *Other languages* (Latvian, Ukrainian and so on). The reason why Turkish has been classified as *a language of particular community significance* has been explained as follows: "some languages have particular importance in a given geographical area. These are languages with significant numbers of speakers in local communities resulting in viable enrolments in a given school or cluster of schools, ie Turkish in Upfield" (LOTE Strategy Plan 1993:5).

Peköz (1993) observes that instruction through the medium of English, with Turkish taught as an additional first language is the commonest situation for Turkish learners. Unfortunately, younger Turkish learners do not get enough of the Turkish language and culture in this way. According to Peköz, serious identity crisis problems experienced by the second-generation Turkish Australian youth could be solved through a more efficient bilingual education system that caters for linguistic and cultural needs.

b) English language proficiency

Language is still our number one problem. I can understand a fair bit of the newspaper, and I can mostly make myself understood. However, I can't express the depth of my feelings and ideas in

English, and this sdepresses me a lot. All ethnic groups seem to experience this depression, this inability to participate fully. It's the only drawback of living here [...] (Bağarýn and Bağarýn 1993:16)

The description given above by one of the first generation Turkish migrants can be generalised to most of those who arrived in Australia to enter the workforce in the late 1960s and early 1970s.

This was confirmed in a study conducted with a group of 50 Turkish elderly in Melbourne by Deasey et al. (1988) who found that very few of their informants could communicate in English and the Turkish community still needed bilingual doctors, community workers and hospital interpreters. Also Elley's (1988) informants lacked English proficiency, and the English words they had learnt and used in their Turkish conversation were work-related such as *plain-machine*, *overlock*, *overtime*, *part-time*, *boss* and *compo*. Yađmur (1993) found that the lack of English speaking ability amongst the first-generation Turks in Australia was the major factor influencing the isolation of the Turkish community. Migrant arrivals of 60s and 70s included Turkish as well as Italian, Greek and Yugoslav *guest workers* who had gained experience in industrialised countries such as Germany. These ex-guest workers found employment in the Australian branches of German firms and used German rather than English to communicate.

The second generation, on the other hand, is better educated (Ýçduygu 1991, 1994, Elley 1993, Inglis 1993) and thus better integrated into the mainstream culture. However, researchers have found that (Yađmur 1993, Price 1989) 'out-group marriage' is rare, amongst Turks in Australia, even in the second generation. In fact, Price (1989:22) has observed that "the in-marriage proportion for Turkish-born women has increased from 75.6 per cent, to 81.8 per cent, 1981-7; which suggests that Turkish girls, who arrived as children in the late sixties and early seventies, have been reared in a close community system with little mixture with non-Turks." This factor, while contributing to isolation, might have also contributed to the maintenance of Turkish as a community language in Australia. Another reason for high level of language maintenance is the tendency of the Turkish migrants to settle mainly in New South Wales and Victoria, the two states

with heavily concentrated Turkish settlements (Y cduygu 1991:88). In fact, 1986 census showed that 94 per cent of the Turkish population lived in NSW (43%) and Victoria (51%) (Y cduygu 1991). Y cduygu's study has identified the reasons for concentration in certain areas as "proximity to place of employment and other Turkish migrants, and the availability of cheap housing" (Y cduygu 1994).

As opposed to the first generation migrants, Elley (1993) observed that the Turkish youngsters who participated in her study switched behaviour and language constantly and consciously. She argued that the second generation accepted their dual identity, Turkish and Australian, and switched identity according to the social setting.

6.2.1 Government services, community network and resources

Migrants enjoy a number of services in their language provided by the Commonwealth and local governments. For example, The Telephone Interpreter Service (TIS) was founded in 1973 and provides general service in some 80 languages including Turkish, and operates on a full-time basis in Victoria (ABS 1992:37). Under the Access and Equity Plan developed by the Department of Social Security (DSS) a multilingual Telephone Information Service (MTI) became available in Victoria in 1986. MTI's aim was to assist with the provision of information about social security in the clients' preferred language. Turkish was among the six languages initially offered and amongst the most used in 1990 (ABS 1992:35). Some government departments such as DEET, on the other hand, do not have bilingual staff policy, nor is being bilingual and familiar with an ethnic group a part of the job selection criteria (ABS 1992:37; Ke eli and Cahill 1995). A study on access and equity in Brunswick, Victoria, an area with very high Turkish concentration, found that the participating migrant groups complained about the staff in some departments who lacked sympathy for NESB migrants and were unaware of other languages and cultures (ABS 1992:61-63). The Victorian Government also established the position of Interpreters as Assistant Professional Appointees in the Special Services Division in 1975. Turkish was amongst the languages for which interpreters were required. In order to meet such demands, Deakin University, Department of Interpreting and Translating offered

undergraduate programs to Turkish-speaking students between 1983 and 1996. On completing the three-year course, students of Turkish background were entitled to work as accredited translators and interpreters, and/or teach Turkish in bilingual programs. On average 10-12 students graduated every three years. Also RMIT offered a Turkish program that ended in 1995. Currently, Turkish is not offered at university level in Australia, and also the secondary level Turkish programs occasionally come under threat due to policy changes.

Studies of ethnic press covering Turkish newspapers (Taylor and Young 1985, Young and Taylor 1989) reported that Turks in Melbourne and Sydney ranked amongst the highest ethnic press readership category. The number of Turkish newspapers, considering the population size, is very high.

SBS Radio stations transmit Turkish programs daily that last for an hour. There are 8 hours of transmission in Melbourne and 7 hours of transmission in Sydney, per week. These transmissions are broadcast also to the other states twice weekly. There are also independent Turkish radio programs that broadcast through community radio stations such as 3ZZZ and 3CR in Victoria, and Bayrak Radyosu in NSW. There is also a community television channel that broadcasts in Sydney for a limited number of hours on certain days.

Turkish community organisations can also be found in all Australian states with the exception of Tasmania and Northern Territory. The Directory of Ethnic Community Organisations in Australia lists 3 Turkish associations in Western Australia, 29 in Victoria, 3 in South Australia, 1 in Queensland and 15 in New South Wales promoting a variety of interests ranging from cultural to religious activities.

6.2.2 Language Use

The section above has looked at language use in the Australian Turkish community mainly as reported by researchers and government institutions. This section, on the other hand, is an informal report on what the members of the community think about the present and the future of Turkish in Australia, as reflected in the

questionnaires. A 16-question survey was distributed, as part of the present study, to a number of Turkish newspaper editors in Melbourne and Sydney (see appendix for a copy of the questionnaire), and the answers related to language use are summarised here. There are references to the answers also in other chapters (see particularly chapter 7).

This section also makes reference to language use in terms of lexical preferences in the Australian Turkish newspapers (see chapter 7). The relationship between political attitudes and vocabulary has always been a popular debate in Turkey and taken as the measurement of success and/or failure of the Language Reform. This point has been discussed in detail in section 2.2.

While the editors agree that the Turkish newspapers have very wide readership, they also confirm that these readers are mainly the members of the first generation aged 30-60; and that the Turkish media in Australia hardly reaches the second and/or third generation. This is seen as a potential threat to the community's language maintenance efforts since the first generation is now aging and it is desirable to pass Turkish onto the younger generations. However, the younger generation Australian Turks prefer English as the language of communication and reading Turkish newspapers is not their top priority. As one of the editors puts it (personal communication):

[O]kurların %90'ı birinci kuşaktır. İkinci kuşak Türkiye gazeteleriyle, Türk diliyle fazla ilgilenmiyor. Onların ilgilenmesini sağlamak ise çok büyük bir iddia. Biz bunu, bildiğimiz kadarıyla gerçekleştiremeyiz. Üçüncü kuşak bildikinden daha da ilgisiz olacak. Ne kadar çalışsak bile Türkçeyi unutulacak. Bizim ağımdan ise, Türkçe gazete için pazar ortamı yavaş yavaş yok olacak.

The majority of our readers, almost all of them are first generation members. The second generation is not interested in Turkish media, Turkish language and so on. It would be too ambitious to say we can get them interested. We cannot do it with our present means. The third generation will be even less interested than the present one. No

matter how hard we try, Turkish will be gradually forgotten. As for us, the market for Turkish newspapers will slowly disappear.

In terms of daily language use, the answers vary as to the amount of Turkish and English spoken. While some use between 2-3 hours of Turkish per day, the others speak more hours of English. Since newspaper activity is hardly anybody's full-time job, language use changes according to the profession practised by the individual editor. For example, out of four community members involved in the publication of *Anadolu*, three were teachers of English who had full-time jobs during the day.

The editors who have participated in the survey acknowledge that they do not have a strict language policy, and some newspapers publish received texts without any changes while others carefully edit them. Although in terms of content there may be differences among the newspapers examined, the language used does not reveal any hidden attitudes. It is commonly debated in Turkey that the choice of Arabic and Persian loans vs TDK-coined words in written and spoken text is indicative of certain political and/or religious affiliation. One typical example is the case of *mesela* vs *örneđin* both meaning 'for example', the former being an Arabic loan. A count of these two words in the Ozturk corpus gives us the following picture:

Newspapers	mesela	örneđin
Dünya	1	0
Güneş	0	0
Turkish report	0	1
Türk Sesi	2	0
Turkish Weekly Gazette	4	0
Yeni Vatan	1	11
<u>Yorum</u>	<u>0</u>	<u>5</u>
Total	8	16

As can be seen from the counts above, the 'Turkish' word as such is used twice as many times as the loan and the distribution by newspaper is not very significant as one would expect *Türk Sesi*, the one with stronger religious tendency, to have the highest frequency of *mesela*. Also religious words such as *Allah* (32), *Kur'an* 'Koran' (11), *camii* 'mosque' (31), and their various forms display average

frequency without any significance in terms of distribution across the newspapers used.

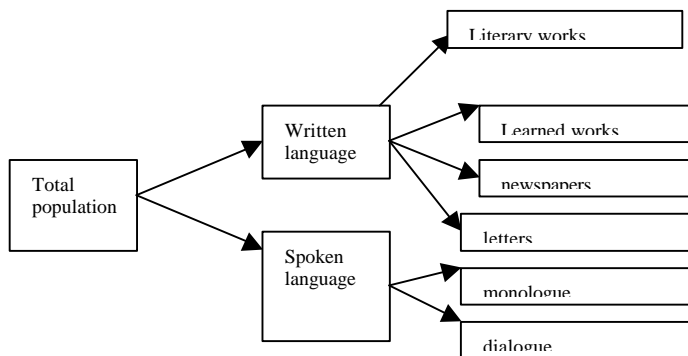
6.3 Summary

This chapter has given a short history of the Turkish community in Australia focusing in particular on the language-related problems the early members of the community encountered and the solutions offered by the Australian government and other public institutions. It has also reported on the use of Turkish as indicated by the questionnaires distributed, and the texts collected for this study.

CHAPTER 7: Methodology

7.1 Introduction

Linguistic studies involving the creation of a corpus aim at including sample texts from various domains of language use as well as different modes. Engwall (1994) subdivides the total text population as follows:



Amongst these categories that would form the corpus "for a linguistic study of general language use today focusing on vocabulary or grammar, newspaper texts clearly offer a more adequate basis than literary or specialised categories" (Engwall 1994).

Similarly, ethnic community newspapers offer a solid basis for language contact research. Various functions of ethnic community newspapers have been identified by researchers. These are preparing migrants for good citizenship in the country of settlement (Zubrzycki 1958, Gilson and Zubrzycki 1964), providing non-English speaking migrants with information in their language (Doueihy 1980, Bednall 1992) and contributing to the maintenance of the mother tongue (Petre 1995). As stated by Doueihy (1980), community newspapers, especially in the early stages of migration, are usually the only source of information available in the migrant's mother tongue and only "a migrant with no knowledge of English can appreciate the need and value of such a service". Bednall (1992:35) states that recent arrivals are more likely to depend on the ethnic press for information than on ethnic radio. His informants from

different cultural backgrounds reported that ethnic press was amongst the most common sources of initial information regarding community facilities such as Ethnic Communities Council, Telephone Interpreter Service, Migrant Resource Centres and Social Security Information Service for Migrants (1992:16-17).

Bosi (1980) states that all ethnic publications start 'with some sort of missionary zeal' (in Haugen's terms "with more devotion than financial reward" (1953:35) but may have different ends. From an organisational point of view, he distinguishes three categories of ethnic newspapers: a) the fully commercial operation (newspapers are published as a business venture and usually owned by a company); b) the semi-commercial operation (usually owned by an individual); c) the private venture (usually community organizations who feel they have a message to pass on). There have been a few attempts in the Turkish community to publish newspapers as a business venture that has failed for various reasons (eg *Tercüman* and *The Turkish Gazette*). The latest of these attempts is the *Turkish Report & Australian News*, published regularly since 1992. Bosi (1980) claims that the largest number of ethnic newspapers belong to the second category. Most of the Turkish community newspapers are mainly distributed free of charge, and survive by attracting as much advertising as possible from mainly two sources: Turkish businesses in Australia and various departments of both federal and state governments. Bednall also confirms that ethnic community newspapers are now "more actively used by service providers for particular community campaigns than they were in the past" (1992:36).

However, as stated by Bosi (1980), these publications suffer from some common ills. Firstly, they present an interesting stylistic dichotomy. "On the one hand, you have the stories clipped out of newspapers and magazines published in the respective countries of origin; on the other the stories written in Australia." These are very different from each other both in terms of content and language. Another common characteristic of these papers can be seen as "personalised journalism." According to

Doueihi (1980), the editors and staff of ethnic newspapers "often tend to express their own political views through the pages of their publications" rather than objectively report the events. Despite these factors affecting their quality, ethnic newspapers still attract high readership rates thanks to their functions mentioned above. It should also be mentioned that they exist "in a world which is far less demanding both culturally and linguistically" (Bosi 1980).

Bednall (1992:28) found that his informants' major reasons for reading ethnic community newspapers were keeping up to date with news about their country of origin, receiving news about the other members of their community, finding out about other overseas news and learning about changes in government services. Yet another reason for preferring these publications may be the low levels of English language proficiency amongst Turkish migrants. According to 1991 Census figures, 39.6 per cent of Turkish migrants reported that they did not speak English well or at all (Hugo and Maher 1995).

Doueihi (1980) estimates that "the total weekly readership of the whole ethnic press in Australia is equivalent to the total circulation of a major Australian paper such as the Sun-Herald, or the Sunday Telegraph in Sydney or a similar publication in Melbourne." However, it is difficult to obtain exact figures of circulation as well as a clear picture of the current situation of print media, since these publications have no proper audit circulation (Doueihi 1980, Petre 1995). Estimates of total readership is even more difficult if "pass on readers" are also included (Petre 1995). As "the number of old readers (who die off, become tired of reading those publications, move onto reading Australian newspapers and magazines)" diminishes, these publications gradually disappear although "[...] the demise of one is almost immediately followed by the birth of another" (Bosi 1980).

Thus, ethnic newspapers differ from daily newspapers in terms of the degree of preparation and availability (Engwall 1994). Firstly, ethnic

newspapers are published weekly, fortnightly, monthly or irregularly, and rarely on a daily basis. Both daily and ethnic newspapers have "a lower degree of preparedness than literary and scholarly works" (Engwall 1994) but due to different reasons. Publication of daily newspapers involves "many individuals working to fixed deadlines under great pressure [...]. Articles are often written quickly, then copy-edited in the newsroom. They may be drastically cut, or even dropped entirely at the last moment if more important news comes in before the paper goes to press" (Engwall 1994). In the case of ethnic newspapers, the shortage of preparation time for each text is closely connected to the shortage of staff. Many such publications are the outcome of individual or family efforts who have full-time jobs during the day. In terms of availability for research purposes, daily newspapers that use new technologies offer corpus-builders great opportunities and make machine-readable texts available to researchers. Ethnic newspapers, on the other hand, have only recently started using computers for text preparation and very few copies are available in electronic form. It is also very difficult to obtain these newspapers due to poor distribution. In terms of delimitation, ethnic newspapers, like daily newspapers and "other printed material, may be said to have relatively clear boundaries. The construction of a corpus within this category thus mainly involves the general problems of selecting texts to form a representative and reliable subpopulation. The specific problems concern the selection of sample texts [...] and the classification of these texts" (Engwall 1994). The major decisions concerning the classification of texts and the definition of subpopulations in this study are explained in 7.3.

7.2 Data collection and analysis in Language Contact

According to Stubbs (1996:4) a major question in contemporary linguistics is: what counts as data? and its implications are significant for descriptive and theoretical linguistics. This question is undoubtedly relevant to language contact. Most researchers would consider spoken

language as contact data and subscribe to one of the approaches outlined by Dabène and Moore (1995):

1. Speech samples are elicited using standardised experimental procedures from informants recruited according to strict selection criteria.
2. Informal interviews are recorded at socially marked settings (eg. school) with bilingual informants selected according to various criteria.
3. Natural language samples are collected in authentic interaction situations without intervention from outsiders. Participant observation is carried out by investigators from the same group.

Until the late 1960s language contact data centered around the function, stability, and distribution of the languages involved, in relation to their location, origin, and dominance both at individual and group levels (Mackey 1969). However, the unit of measurement was a problem. Units were often "simply measures of indices" were used to "reflect certain variables of bilingualism-dominance, skill, regression etc" (Mackey 1969).

The validity of data collection techniques such as word-association skill was also questioned and the difficulty of collecting valid samples using questionnaires or recordings was recognised. The elaboration of units of measurement such as type-token ratios or percentage of loan-words also posed problems that led researchers to turn to statistics and information theory for alternative methods of elaboration (Mackey 1969).

Selection of data collection techniques depends on the researcher's perspective. Milroy (1992) draws a distinction between *system-oriented* approaches and *speaker-oriented* approaches. System-oriented research focuses on language as a changing system in contact situations regardless of its speakers. Speaker-oriented research, on the other hand, deals with language in terms of its speakers. In both cases, however conversational

data have been dominant, as in pragmatic-discoursal line of research that focuses on sequential features of conversations or macro-sociological approach (eg. Gumperz) that uses conversations (though transcribed in a less detailed way), and emphasizes participants' post hoc explanations of bilingual language behaviour. Formal studies of language contact, on the other hand, deal with idealised or edited data to which syntactic theories are applied (e.g. Joshi 1984; Di Sciullo *et al.* 1986). Quantitative data and computational analysis have also been used by Sankoff and Poplack (1981). At the analysis stage the data are usually interpreted in terms of one of the phenomena along a continuum i.e. code-switching, borrowing, shift, attrition and finally language death. Naturally, these phenomena are controversial and overlap considerably (see Chap 3 on the development of methodological options).

As opposed to the conventional data collection and analysis procedures in Language Contact, this study brings together two major sources of data as well as data analysis techniques that have not been considered previously. These sources and the analysis are explained in this section. In order to be able to construct the history of the newspaper activity and gather information on the individual newspapers, which formed the basis of this study, the newspaper editors have been asked to complete a questionnaire of 16 questions. Details of the questionnaire are enclosed in the appendices.

7.3 Sampling

As in any corpus building project, it is important to keep the corpus 'balanced' by including texts from the community newspapers that advocate different political and religious ideologies. However, a fundamental constraint has been the limited life cycle, limited circulation and the lack of archives of these newspapers. As confirmed by Taylor and Young (1989) independent circulation figures of the Turkish newspapers are not readily available. Lists of Turkish community newspapers were

obtained from the Turkish embassy and consulates in Australia, ethnic press directories and media guides, public libraries as well as the editors themselves and a clear picture of the Turkish press since the publication of the first Turkish community newspaper in Australia has been attempted. The table 6 below shows the published newspapers throughout 1970-95. Although a variety of sources have been consulted in order to compile the list, it should be mentioned that these sources are not always reliable. For instance, a newspaper that has long closed down may still be included in one of the major media guides or publication details such as frequency of publication may be misleading.

As can be seen from the list below, there is a large number of newspapers that represent different views and have high readership in different sections of the community. For example, *İslam'ın Sesi* (The Voice of Islam), was reported to be the most popular newspaper amongst Young and Taylor's 100 informants in Melbourne, who mostly had lower levels of education and worked as factory workers.

After consulting the census figures on community size, it was decided to concentrate on a period of 15 years: 1980-1995. This decision was based on two premises. The first one was the fact that it was extremely difficult to reconstruct the history of Turkish community newspapers in Australia prior to 1980. Regular newspaper publishing activity can be said to have started in late 1970s and early 1980s. The second reason was that after assisted migration had ceased in 1974 there had been a decrease in the number of arrivals and the main reason for immigration was family reunion. The biggest intake, after a period of low migration figures was in 1980.

Table 4 below shows the number of arrivals per year from the signing of the bilateral agreement to present, and table 5 shows the net increase in the number of Turkish-born persons in Australia between 1976 and 1981 censuses.

Table 4

Turkey-born settler arrivals

Years 1967-68 to 1994-95

Financial year	Settler Arrivals	Financial year	Settler Arrivals
1967-68	512	1981-82	836
1968-69	2,274	1982-83	520
1969-70	3,832	1983-84	764
1970-71	3,809	1984-85	757
1971-72	2,438	1985-86	994
1972-73	2,272	1986-87	1,153
1973-74	3,188	1987-88	1,382
1974-75	1,744	1988-89	1,227
1975-76	1,062	1989-90	882
1976-77	1,199	1990-91	1,090
1977-78	1,159	1991-92	888
1978-79	743	1992-93	625
1979-80	830	1993-94	556
1980-81	1,002	1994-95	718

*Source: Bureau of Immigration, Multicultural and Population Research (BIMPR)
Turkish Community Profile*

Table 5

Turkey-born population of Australia

1966-1991

1966	2,476
1971	11,589
1976	19,355
1981	24,314
1986	24,515
1991	27,220

Source: BIMPR Turkish Community Profile

Table 6: Turkish Newspapers by publication date

1970	Anadolu, Türk Sesi, Australian Crescent
1971	Anadolu, Türk Sesi
1972	Anadolu, Türk Sesi
1973	Türk Sesi
1974	Anadolu, Türk Sesi, Ay Yıldız
1975	Anadolu, Türk Sesi, Ay Yıldız, Yeni Vatan
1976	Anadolu, Türk Sesi, Ay Yıldız, Yeni Vatan
1977	Türk Sesi, Yeni Vatan, Güneş Dayanışma, Yorum
1978	Türk Sesi, Yeni Vatan, Güneş Dayanışma, Yorum
1979	Arkadaş Dost Sesi, Emek, Güneş Haber Bülteni, İslamın Sesi, Merhaba, Sydney Express, Yeni Türkiye, Türk Sesi
1980	Emek, Merhaba, Türk Sesi, Türk Dayanışma, Yeni Nesil, Yeni Türkiye, Yeni Vatan, Yorum
1981	Bosfor, Emek, Güneş İslamın Sesi, Merhaba, Türk Dayanışma, Türk Sesi, Yeni Dünya, Yeni Nesil, Yeni Türkiye, Yeni Vatan, Yorum
1982	Çadır, Emek, Güneş Halkın Sesi, İslamın Sesi, Kaynak, Merhaba, Sydney Express, Tercüman, Türk Dayanışma, Türk Sesi, Yeni Dünya, Yeni Nesil, Yeni Türkiye, Yeni Vatan, Yorum
1983	Emek, Güneş İslamın Sesi, Kaynak, Sydney Express, Tele Haber, Tercüman, Türk Dayanışma, Türk Sesi, Yeni Dünya, Yeni Türkiye, Yeni Vatan, Yorum
1984	Emek, Güneş İslamın Sesi, Kaynak, Olay, Sydney Express, Tele Haber, Tercüman, Türk Dayanışma, Türk Sesi, Yeni Dünya, Yeni Türkiye, Yeni Vatan, Yorum
1985	Güneş Sydney Express, Tele Haber, Türk Sesi, Yeni Dünya, Yeni Vatan, Yorum
1986	Güneş Sydney Express, Tele Haber, Türk Sesi, Yeni Dünya, Yeni Vatan, Yorum
1987	Dünya, Emek, Güneş İslamın Sesi, Kaynak, Olay, Sydney Express, Tele Haber, Türk Dayanışma, Türk Sesi, The Turkish Gazette, Yeni Vatan, Yorum
1988	Dünya, Emek, Güneş Kaynak, Sydney Express, Tele Haber, Türk Sesi, The Turkish Gazette, Yeni Vatan
1989	Dünya, Emek, Göçmen Toplum, Güneş Kaynak, Tele Haber, Türk Sesi, The Turkish Gazette, Yeni Vatan, Yorum
1990	Dünya, Emek, Güneş İslamın Sesi, Kaynak, Kıbrıs, Tele Haber, Türk Dayanışma, Türk Sesi, Turkish Weekly Gazette, Yeni Vatan, Yorum
1991	Dünya, Güneş Tele Haber, Türk Sesi, Turkish Weekly Gazette, Yeni Vatan, Yorum
1992	Dünya, Güneş Tele Haber, Türk Sesi, Yeni Vatan, Yorum
1993	Dayanışma, Dünya, Güneş Tele Haber, Türk Sesi, Turkish Weekly Gazette, Turkish Report, Yeni Vatan, Yorum
1994	Dünya, Güneş Türk Sesi, Turkish Report, Yeni Vatan, Yorum
1995	Dayanışma, Dünya, Güneş Toplum Sesi, Turkish Report, Yeni Vatan, Yorum

SOURCES: AARDS Consumer and Outdoor, The Australian Media Directory, B & T Advertising, Marketing and Media Year Book, List of Ethnic Newspapers Circulating in Melbourne, Ethnic Press, A Guide to Ethnic Media, Margaret Gee's Media Guide, Guide to Ethnic Media in Victoria, The Ethnic Media Guide.

Samples from various newspapers representing different political and religious ideologies have been included in the corpus as their language could differ in terms of vocabulary. It can be seen from the list above that three papers have remained in circulation for a long period of time. Regular publication history would enable the examination of language change over 15 years although it was not possible to obtain the same number of texts for each year. The number of texts per year and per newspaper is shown below:

texts: number per year

year 1 (1980-1984)	118
year 2 (1985-1989)	207
year 3 (1990-1994)	305
year 4 (1995)	371

text source total number of texts

Newspapers

Yeni Vatan	322
Yorum	293
Dünya	25
Turkish Report	105
Güneş	60
Türk Sesi	78
Turkish Weekly Gazette	65
Turkish Community Assembly	1

Newsletter

The following papers were selected for the main corpus material:

1. Türk Sesi (Melbourne 1970-1993)

2. *Yeni Vatan* (Sydney 1975-1995)

3. *Yorum* (Sydney 1979-1995)

Sample texts from the following newspapers have also been included in the study to supplement the main corpus:

4. *Dünya* (Sydney 1990-1995)

5. *Turkish Report* (Melbourne 1992-1995)

6. *Turkish Gazette* (Melbourne 1990-1993)

7. *Güneş* (Melbourne 1977-1995)

After the list was made, the editors were contacted in order to obtain the archives of the newspapers selected. Their collaboration was essential for the loan of the issues identified as these newspapers have limited circulation and are not readily available in public libraries. The editors were also asked to complete a questionnaire to obtain information about their language policy, L1 and L2 use and so on (see appendix).

Yeni Vatan has been published since 1975 however the editor changed in 1980. *Yeni Vatan's* editor made his complete archive of 15 years available for this study for an extensive period of time.

Yorum has been received regularly by the Monash University library since 1979 and was arranged with the library that the issues needed could be taken out.

Turkish Report has been published since 1992 on a weekly basis and its 3-year-archive has been obtained as well as the most recent issues in machine-readable form.

Güneş has been published since 1977, first fortnightly, then monthly and its publication has gradually become irregular. Back issues have been obtained from the editor.

Türk Sesi was the first Turkish community newspaper to be published in Australia, interestingly enough, by a Turkish Cypriot. It started as a weekly newspaper in November 1970, gradually became a monthly and was later irregularly published until it ceased publication in 1993. Therefore, only years 1980, 1985 and 1990 were included in the corpus. The editor of *Türk Sesi* provided a random selection of issues for the years selected.

Dünya was first published in 1981 as *Yeni Dünya* and in 1990 the editor changed as did the name. It is a fortnightly publication.

Turkish Weekly Gazette was active in Melbourne the early 1990s and had a relatively short life span of 3 years. The texts from this newspaper were also obtained in electronic form.

7.3.1 Ethics Clearance and Copyright

The research proposal was approved by The Standing Committee on Ethics in Research on Humans of Monash University (12 September 1995/project approval no 167/95). Upon approval, a letter seeking permission for the use of the newspaper texts (see appendix) was sent to the editors concerned. Copyright is usually one of the most serious constraints on the development of large text corpora. In this case, the Turkish community has been very supportive and the editors have been willing to make the material available for inspection and use for an extensive period of time.

7.3.2 Content Categories

Although newspapers indicate the types of text by subject headlines and column titles for groups of articles, such headlines and titles usually differ between newspapers and sometimes even between issues of the same paper. It was therefore necessary to analyse the content of the sample texts regardless of the category assigned by the individual newspaper:

"The heterogeneity of the texts in a daily newspaper should be treated so as to allow different subpopulations to be defined, in the same way as different genres are distinguished in the literary category. Together with decisions on the inclusion and exclusion of advertisements, these are the problems that must be solved when newspaper material is used in text corpora" (Engwall 1994).

Zubrzycki (1958) found that ethnic community newspapers typically had "five major divisions: news of the country of settlement, world news, home-country news, group life and interests, editorial features." Later, Gilson and Zubrzycki (1967) used seven different content categories in their study of the content analysis of community newspapers. These content categories were as follows:

1 - Homeland; 2 - European and other foreign affairs; 3- Australian Affairs; 4 - Group activity in Australia; 5- Legal Information; 6 - Advertisements; 7 - Miscellaneous.

However, a careful examination of ethnic newspapers, as well as previous research, reveals that these newspapers mainly reprint articles from overseas newspapers. Although keeping up with mainland newspapers helps "readers maintain and develop vocabulary and structures, often introducing them to neologisms reflecting socio-economic, political and technological change in the country of origin" (Clyne 1991:145), it is obvious that material taken from newspapers in Turkey could not be used for the purposes of this project. The content categories were thus revised as follows:

1. Australian Affairs

- (a) Migration
- (b) Political and trade union matters
- (c) Other, including general information

2. Group activity in Australia

- (a) Social life of migrant groups
- (b) Religious matters
- (c) Sport

3. Letters to the Editor

4. Advertisements

- (a) Government
- (b) Turkish Business in Australia
- (c) Community announcements

5. Miscellaneous

6. Editorial

7. Regular features (including Legal Information and Health)

The other category of texts that has contributed to the creation of the Ozturk corpus is the information leaflets. These leaflets are regularly published by the Australian federal or state governments, public institutions and community organisations in various community languages. These leaflets are originally written in English and later translated into the community language requested, in this case, Turkish. All together, 52 leaflets have been used that can be put into seven broad categories in terms of content:

1. Children
2. Community Services
3. Domestic Violence
4. Health
5. Human Rights
6. Women
7. Work

(See appendix for details of texts)

Once the categories were established, it was then a question of identifying a range of texts that fitted the selection criteria. Since the selected texts are not excessively lengthy, they have been used in unabridged form. The number of leaflets from each category is shown below:

Leaflets

Community Services	7
Domestic Violence	4
Children	9
Work	11
Women	6
Human Rights	7
Health	8

For the purposes of this project, the notion of text has been limited to the printed newspaper texts written in an immigrant setting and information leaflets printed in Turkish in order to provide the migrants with information on various matters.

7.3.3 Time span and size

While the aim of the project was to document the instances of L1 + L2 contact, and the patterns emerging, the samples were selected in order to enable a comparison of the various time periods of the language. The years selected were as follows:

- 1980
- 1985
- 1990
- 1995

For each year selected, every fifth issue was taken. However, in some years, the frequency of ethnic newspaper publication can change and a weekly newspaper may be published fortnightly, monthly or vice versa. In these cases, every tenth issue was used in order to take account of the number of issues per year. This rationale was applicable only to Türk Sesi, Yeni Vatan and Yorum thanks to their regular publishing activity. Texts from the other papers and information leaflets are mostly from the latest period (1990-1995).

Corpus projects in general start by setting a target number of texts to be scanned and processed. However, the unique nature of ethnic newspapers does not allow such a pre-set target, as it is difficult to know in advance how many of the texts in a single issue will be copied from the mainland newspapers and how many are authentic migrant production. The collection of texts was therefore carried out by including every text in the newspaper except for texts copied from newspapers in Turkey. These were easy to identify, as they were cut and pasted directly from the original, and had a different font. In total, one thousand texts were used (see appendix for details).

7.4 Hardware and Software

The coding, indexing and retrieval systems required for corpus processing demand large disk storage and processing power. Three basic tools are needed for a corpus project to be completed. These are

Word frequency – Software that produces lists of word types and their frequency counts in the corpus.

Concordancing - Text retrieval and indexing software specifically designed for linguistic analysis. The concordancer is a powerful tool that effectively retrieves information from the corpus by placing each token of the node word in the middle of a line of text and the remainder of the line displays the immediate preceding and following context of the node. The

set of characters at the beginning of each concordance line helps identify the location of each occurrence. The biggest advantage of working with this tool is that it brings out patterns in the data that become immediately obvious to the researcher.

Interactive searching - Flexibility in search and display of patterns.

In addition to the above-mentioned tools, if texts selected to build the corpus cannot be obtained in electronic form, an *optical character reader* (scanner) is also needed for the sample texts to be captured (Atkins et al 1992).

7.4.1 Data Capture

Since the Turkish Community newspapers have only recently started using Desk Top Publishing facilities, very few texts could be obtained in electronic form. Therefore, the texts were scanned with Hewlett Packard ScanJet 3p Scanner dpi (Black and White) and the software used was OmnipagePro version 6. However, community newspapers vary in print quality and the poorly captured image of texts with frequent orthographic errors were misinterpreted during processing. The texts were thus photocopied prior to scanning that improved the electronic image to a certain extent. Some 200 texts were acquired electronically¹², and where everything else failed, the texts were entered manually. In order for the computer to recognise the non-standard characters (ç, ð, ý, ö, ħ ü), a Turkish version of Windows 3.1 was also loaded onto the computer. When the texts were cleaned up, they were coded according to the pre-set coding scheme.

7.4.2 Data Processing

¹² I am grateful to Mr Levent Efe and Mr Yüksel Yýlmaz for making the texts available in machine-readable form.

The coding scheme included the basic information in the following sequence:

- the title of the newspaper in abbreviated form,
- the issue number
- text number (also indicates how many texts have been taken from a given issue)
- content category

eg. **yor55 001H**

(Refers to the editorial of the 55th issue of Yorum).

Content categories listed below were assigned letters in capitals. Lower case refers to subcategories.

eg. **yv235 002Ba**

(refers to a text about a social event in which the Turkish Community was involved.)

The newspaper corpus was divided into a number of subcorpora as outlined below (see appendix for details). While the main corpus includes the entire collection of news stories and other features per newspaper per each selected year (1980, 1985, 1990, 1995), specific software has also made a comparison of the subcorpora such as the following:

- news stories
- editorials
- letters to the editor
- advertisements

The corpus of translated texts (ie information leaflets), on the other hand, was also available separately as well as a part of the main corpus. This subcorpus also included the 1st category of advertisements put out by the government agencies originally written in English and appeared in the Turkish community newspapers in translated form.

A break up of the text types in terms of numbers is as follows:

Text type	number of texts
x (regular features)	137
h (editorial)	39
g (miscellaneous)	32
e (letters to the editor)	22
fa (government announcements)	80
fb (Turkish businesses)	252
fc (community announcements)	69
ba (community life in Australia)	82
bb (religious matters)	12
bc (sport)	27
aa (Migration matters)	25
ab (Australian politics)	44
ac (general information)	100
ad (matters of interest to Turkey and Australia)	27
information leaflets	52

The reason for keeping these categories was that some text types might be more prone to borrowing (Clyne 1991:146). Keeping the material separately, therefore, enabled both independent and cumulative search of the patterns under investigation.

Next step was to obtain the word lists and the concordances. SIL's *Text Analysis* program (TA) (Costin and Beadle 1995) was used to obtain the

word lists and concordances. The TA program runs in MS-DOS and the non-standard Turkish characters were also created for MS-DOS environment. A variety of other programs were also used (see below) to manipulate the word lists and concordances to the printable version. The texts were further coded for TA to recognise the references and extract just the text. The main codes are as follows:

\ref - appears in front of the newspaper title and issue number

\code - appears in front of the item number

\tx - appears in front of each paragraph of text

As the analysis progressed a number of other codes were introduced to facilitate the search:

\c - word count

\y - year of publishing

\tt - text type

\mx - mixed patterns

\e - English words

Word lists were created as follows:

1. The *Word* files were saved as *text* files.
2. The Word List (WDL) program from Text Analysis (TA) was used on each year's text to produce an Standard Format Markers (SFM) file containing the unique words with their counts and references.
3. Shoebox 4.01 (Buseman *et al* 1998) was used to print the word lists.
4. Shoebox 4.01 was used to extract various subcorpora files from the main corpus for comparisons.
5. LexTools was used to create wordlist comparisons. The output files were converted to SFM format using SIL's Consistent Changes program then Shoebox 4.01 was used to view the word list comparisons.

The concordance lines were produced as follows:

SIL's Interactive Concordance (IC) program was used on the combined files to produce the concordances. This required a suitable settings (or answer file) to be created using the words from the wordlist generated by WDL. The context span was 40 characters. Standard Format (SF) converter was used to format the output of the IC program.

In order to be able to identify the patterns emerging, hard copies of the word lists and concordances were prepared. In linguistic analysis, data interpretation is usually centered around the acceptability of the forms used in speech and/or writing (see also Toury 1995 and Sinclair 1989, 1991a). However, native speakers' acceptability judgements are mainly based on intuition. Interestingly, corpus research has shown that native speakers' intuitions do not always reflect the way language works. The next chapter deals with this aspect of use.

7.5 About written language

Written language has not received much attention in the study of language contact as researchers have failed to see it as "an extension of the spoken language" (Halliday 1987). Culler (1987) points out that "[w]hen discussions of language claim to focus on speech and set aside writing as unimportant, what they in fact do is set aside certain features of language or aspects of its functioning." Thus, any theory of knowledge, and of learning, must encompass the grammars of both written and spoken language as these are complementary (Halliday 1987). One aspect that is worth mentioning in particular with reference to frequency is the difference in *lexical density* between written and spoken texts. Halliday describes lexical density as "a measure of the density of information in any passage of text, according to how tightly the lexical items (content words) have been packed into a grammatical structure (Halliday and Martin 1993:76). Accordingly, "the complexity of writing lies in its density, the packing together of lexical content, but in rather simple

grammatical frames [...]. The complexity of spoken language is more like that of a dance; it is not static and dense but mobile and intricate" (1994:xxiv; see also Halliday and Martin 1993).

While the level of lexical density varies from one piece of discourse to the next, the general tendency is that in informal spoken discourse the lexical density is low (2 words per clause) and in more formal and written discourse it is higher (4 to 6 words per clause) (Halliday and Martin 1993:76). Written language also displays "a strong tendency to encode this lexical content in a nominal form: in head nouns, other items (nouns and adjectives) in the nominal group, and nominalised clauses. It is these nominal structures that give the clause its enormous elasticity" (1993:75).

In relation to this study, it can be claimed that written data from migrants may be different from spoken data, although the language of newspaper publishing activity in immigrant settings allows for a variety of registers ranging from formal to very informal (see the section on community newspapers 7.2) and the lexical density may be lower in some cases.

An area of study that deals systematically with *written texts* is *Translation Studies* and it is interesting to note that there are many similarities between translation studies and the study of language contact. A common question to both is how the bilingual's language works. While spoken language has so far been accepted as the direct evidence of the bilinguals' repertoire, their translations should also be considered as the indirect evidence (Sinclair et al. 1996:174) of their language use. According to Baker (1993) translated texts are not taken into consideration because

[i]nstead of exploring features of translated texts as our object of study, we are still trying either to justify them or dismiss them by reference to their originals.

This is very similar to what has been happening in the study of mixed codes, that is, instead of exploring features of mixed codes as our object of study, we are still trying either to justify them or dismiss them by reference to L1 and L2. This issue has also been touched upon by Frawley (1984):

radical innovations occur when the third code begins to "break away" from both the matrix and target codes. As the new code establishes its own rules (its own redundancy), the dependency of that code decreases and the possibility of new knowledge from that code increases. [...] In Eco's terminology, the input into the new code -from the matrix and target- is transformed according to the structural necessities of the new code, which renders the translation wholly accountable neither to the matrix nor to the target.

A common linguistic device that increases the possibilities of expression in text is *nominalization*. However, this aspect of language has not been sufficiently stressed in language contact and translation, particularly with reference to its role in handling the problems of expression. The following section deals with the function of nominalization in texts.

7.5.1 Nominalization

According to Halliday and Martin (1993:2-21), nominalization was originally a feature of scientific writing that enabled grammatical metaphors and has gradually become the norm also in other types of discourse. Importantly, it should be analysed as an essential resource in discourse construction rather than an arbitrary feature of language (1993:61). The question of whether languages are inherently nominalised or simply used by people that way is still to be answered, yet it seems quite common across languages (see for example Halliday and Martin 1993:124-132 on English and Chinese).

Nominalization and the construction of nominal groups offer a rich alternative to verbal groups that "contain only one lexical element: the verb itself. Other lexical material may be expressed in adverbial groups; but these are very limited in scope" (Halliday 1985:73). In contrast, the head of the nominal groups may be preceded and/or followed by various modifiers and this brings the elasticity users need for richer expression.

However, these constructions are inadequately treated in grammar books particularly in relation to the function of certain elements involved (Sinclair 1991a:91). To quote Sinclair's example, a construction such as 'the enthusiastic collaboration of auctioneers', will be explained in rather simple terms as a clause transformed into a nominal group (through noun derivation from verbs) without paying any attention to the function of *of*. Rather than treating nominal constructions in terms of derivation, we should rely on collocation (Sinclair 1991a:92). This point is significant also for Language Contact research as the role of function words in such constructions tends to be ignored.

According to Halliday, (1985:72-73) in English the two sources of pressure towards nominalisation are "the structure of the clause and the structure of the nominal group". The pressure tending towards nominalised forms of expression is so strong that "even things that are not expressed as nouns have to behave like nouns in order to gain their appropriate status in the thematic and information structure" (1985:74).

7.5.2 Delexicalization

Delexicalization has not been studied extensively as a linguistic phenomenon, and limited literature is available on a limited number of delexicalized nouns and prepositions (e.g. Sinclair 1989), intensifiers (Partington 1993), and verbs and adjectives (Sinclair and Renouf 1988, Sinclair et al. 1996). The notion of delexicalization is central to the study of collocation as the delexical word acts in conjunction with other words

and shares their meaning (Sinclair and Renouf 1988, Partington 1993). For example, Tognini-Bonelli (1996a and b), after a detailed analysis of the adjectives 'real' and 'actual' and the adverb 'actually', comes to the conclusion that the adjectives and the adverb have lost their meaning of 'existing in the reality' but through co-selection, have acquired a structural role at the level of discourse.

Delexicalization has also been neglected in translation, and delexical structures have not been considered in terms of equivalence in parallel corpora (Pérez-Hernandez 1996). Previously, in an investigation of Serbo-Croatian and English verb+noun collocations, Benson (1989) also pointed out that the contrast between collocations in different languages was striking but not adequately dealt with in bilingual lexicography.

Also in Language Contact these are varyingly called *auxiliary verbs* (eg. Backus 1996:236 Turkish, Tamis 1986:169 Greek), *semi-auxiliary verbs* (e.g. Silva-Corvalan 1986 Spanish), *compound verbs* (Kachru 1978 Hindi), *operators* (Romaine 1989:131 Punjabi), and so on.

Whether delexicalization is a universal feature of all languages has yet to be demonstrated however ongoing projects involving parallel corpora show that it is common across the language pairs under examination.

7.5.3 Collocation

A much debated question, more enthusiastically over the last decade, has *been what constitutes a lexical unit?* This question interests different fields such as lexicography, machine translation, second language acquisition, psycholinguistics, language teaching and so on for different reasons. This section is a summary of the various definitions put forward in the fields mentioned above. It is important to understand the notion of collocation as it will be clear in the course of this study that in the early stages a mixed code operates through the *open choice principle* and later

on through *the idiom choice principle* proposed by Sinclair (1991a) as summarised in section 4.2.3.1.

Generally speaking, the intuitive identification of lexical units has followed the psycholinguistic perspective and the computational identification has followed the textual perspective (Mitchell *et al.* 1997).

Combined approaches have also been proposed. Terminological confusion is characteristic of this area of study as well and a wide range of labels are used to refer to the same units. These will be evident in the summary below.

Previously, Pawley and Syder (1983) drew attention to *prefabricated patterns* that "form a high proportion of the fluent stretches of speech heard in everyday conversation". More recently, Nattinger and DeCarrico (1992) proposed a lexico-grammatical unit to be used as the central unit in teaching. They placed the *lexical phrase*, which they describe as 'chunks' of language of varying length "somewhere between the traditional poles of lexicon and syntax". These chunks are (1992:1)

conventionalized form/function composites that occur more frequently and have more idiomatically determined meaning than language that is put together each time. These phrases include short, relatively fixed phrases such as a _____ ago, or longer phrases or clauses such as if I X, then I Y, the _____ er X, the _____ er Y, each with a fixed, basic frame, with slots for various fillers (a year ago, a month ago, the higher X, the higher Y, the longer you wait, the sleepier you get).

An immediately visible aspect of lexical phrases is that they are subject to differing degrees of syntactic modification. For example, *for the most part* is a fixed phrase that allows no modification while others like *it's*

only in X that Y, allows a certain degree of modification (Nattinger and DeCarrico 1992:8). They connected the notion of lexical phrase to the notion of *pragmatic competence*: "although grammatical competence encompasses the knowledge of the lexical forms and their internal syntax, pragmatic competence accounts for the speaker's ability to continue to access these forms as pre-assembled chunks, ready for a given functional use in appropriate context." (1992:13).

Kjellmer (1991) in a combination of corpus and psycholinguistic approaches has come up with the following typology granting that the borderlines are fuzzy and categories overlap.

1. Fossilized phrases. one element of such a phrase will suggest the other(s) with great consistency. Unassimilated loans typically belong here eg Anno Domini, aurora borealis and nouveau riche, but more genuinely English fossilized phrases also occur, such as bubonic plague, Cocker Spaniel and be-all and end-all. Variability in such phrases is exceptional. Phrases of this kind could therefore be called 'right-and-left predictive'.

2. Semi-fossilized. one word predicts a very limited number of words. Eg. Achilles heel/Achilles tendon, by and by/by and large, magnum bottle/magnum opus etc are right predictive. There are also left-predictive ones such as go bail/grant bail/jump bail/stand bail etc.

3. A third type ... consists of sequences of words that co-occur more often than their individual frequencies would lead us to expect. One of the words in such a sequence can be said to predict the other(s), as in the previous types, but 'prediction' will have to be interpreted more loosely. One word will tend to co-occur with one or a few out of a great number of words that can

also co-occur with it. 'Tendency' is here a key concept. Such groups of words will here be called variable phrases.

As a step further to lexical units, Renouf and Sinclair (1991) proposed the notion of *collocational frameworks* that "consist of a discontinuous sequence of two words, positioned at one word remove from each other; they are therefore not grammatically self-standing; their well-formedness is dependent on what intervenes." They illustrated the notion of framework with different pairings of high frequency grammatical words such as a+?+of; be+?+to; for+?+of and so on. More interestingly, observing the occurrences of these frameworks in a huge corpus they come to the conclusion that "it is the collocational frameworks that dominate and that somehow the grammar has developed to accommodate their use".

Stubbs (1995a; 1996:44) demonstrates that fixed expressions and recurrent wordings show how the culture is expressed in lexical patterns, and Baker (1992), through corpus work, how fixed expressions are relevant to translation theory (see chapter 5).

Likewise, in machine translation, descriptive parameters are equally difficult to establish as well as their treatment in computational terms. According to Caroli (1995) this wide range of phraseolexemes (he includes idioms, multiword units, phraseological units, collocations and support verb constructions) have the common properties:

- they are composed of at least two distinct lexical elements,
- their syntactic function and/or their meaning is, if at all, only partially captured by a compositional analysis,
- their components are subject to lexical restrictions, i.e., the presence of one specific lexical elements requires the occurrence of the other lexical element in order to form the whole unit.

His descriptive parameters are based on the functional aspects of phraseological units (eg. adverbial *ab und zu*; nominal *das Auf und Ab*; verbal *schief gehen* etc.), their syntactic properties that take into account their combinatorial restrictions; their semantic properties that take into account varying degrees of compositionality; and their pragmatic aspects (eg. in relation to register).

Storrer and Schwall (1995) on the other hand propose a combination of corpus and dictionary-based methods for the analysis, generation and transfer of multiword lexemes in machine translation. They stress two properties of multiword lexemes: their status in the mental lexicon and their complex nature in contrast with simplex words as well as the aspects of non-compositionality and non-substitutability particularly of items in idiomatic expressions such as *kick the bucket*.

In spite of extensive work on collocations over the past few years, however, the concept is still fuzzy around the edges. As Van Der Meer (1998:319) summarises:

But perhaps this is exactly as it should be, given the nature of the language: some combinations will be more fixed than others because some concepts are better established than others. There can be no doubt that to perpetrate a view of New York is a free combination, whereas to perpetrate a crime had better be called a collocation. As a heuristic method, statistics will be useful to solve this problem of where collocations stop and free combinations begin.

7.6 Summary

This chapter has investigated general aspects of ethnic newspapers in Australia with specific reference to the Turkish Community newspapers as well as the information leaflets selected for this study, and highlighted the

uniqueness of these texts for language contact. Corpus compilation has been explained in terms of text selection, size and time span. Computational tools used to process the corpus have also been introduced and their function has been explained. Since the texts that make up the Ozturk corpus come from written sources, particular aspects of the written language such as nominalization and delexicalization have also been explored.

CHAPTER 8: CORPUS ANALYSIS

The analysis below deals with a number of case studies. These represent many similar patterns, which occur in the entire corpus. The reasons for selecting this set of examples become clear in the subsection titles that organise the chapter. The preceding overview of the issues (Chapters 1 and 3 in particular) has emphasised the problem areas in the analysis of language contact data, which are in Myers-Scotton's terms, *singly occurring switches*, *cultural borrowings*, and *islands*. There is also the question of Poplack's *nonce*, as well as the questions of *functors* and *delexical* verbs in the context of contact. Consequently, each of the case studies identified here illustrates one or more of these points which are handled from a lexicogrammatical perspective. At the centre of the analysis is the co-occurrence tendency of the loans with the elements of the host language.

8.1 Emerging Patterns

While language contact researchers realise that the open-choice principle (see section 4.2.3.1 on Sinclair) does not fully operate in the context of immigration, the fact that the choices made may be based on another linguistic principle does not receive much attention. The use of spoken corpora has so far given the impression that slot-filling choices are made haphazardly. This is because regularities are hard to establish when individual performances are analysed manually or semi-manually. In a machine-readable corpus, regularities are obtained by concordancing the mixed units. Lexical analysis in Language Contact has often ignored the lexical environment as well as the co-occurrence tendencies developing around the "foreign material". While the changes in the phonological, morphological and semantic make-up of loanwords has received much attention, the observation that new collocational patterns may be emerging in the process of L1 and L2 contact has not been made. Instead, taxonomic approaches to borrowings have prevailed for a long time and explained the process as lexical gap filling. The following section offers a critical view on the established categories of foreign word analysis in mixed data through nine case studies.

8.1.1 The question of "single switches"

Various types of lexical behaviour can be identified by examining the concordance lines. This type of analysis is particularly useful when dealing with what has traditionally been treated as individual loans. There are two views on the treatment of "singly occurring lexemes" such as *delivery*, *register*, *benefit*, *grants* and so on. In one view there are slots in sentences and when an individual item finds itself a slot like this, it fills the slot, and the analysis does not go any further. For example, Haugen (1956:39) describes this process as follows:

The simplest possible form of linguistic influence is that in which a single item is plucked out of one language and used in the context of another.

This is the conventional way of seeing individual loans in data discussed in the early days of language contact research. Similarly, Pike (1967) states:

Change occurs when, in a particular matrix of a sequence of units, there is an intrusion of a word from an outside system into some slot of that matrix. Suppose, for example, that in the English 'frame' *I like to eat strawberries*, a bilingual Spanish speaker replaces the word *strawberries* with the loan word *tortillas* 'corn cakes'. The accommodation of the loan into English is made possible by the structure of the sentence - which, say, leaves a subject 'slot' in its matrix, for nouns - into which the borrowed word is inserted.

In this definition, we not only see the origins of the still-dominant thinking on loans but also recognize some other familiar concepts such as matrix and frame (see 3.8 on Myers-Scotton). The majority of language contact researchers have subscribed to this view so far (e.g. Myers-Scotton 1993a; Treffers-Daller 1994). In the other view, however, the relationship is seen as one of co-selection (Kurtbøke *in press*, Kurtbøke and Potter 1998). The following section elaborates on this distinction. In order not to break the continuity of the concordance lines, translation is provided only when the node word in question is considered in full context.

8.1.1.1 *delivery*

According to the first view outlined above, the following sentence in Turkish *evlere ücretsiz _____ yapılıyor* that translates into English *as free home _____ can be arranged* has a slot and *delivery* is the gap-filler without much independence of its own. According to the other view, however, what counts is the co-selection tendencies of the loan and its collocations in the new environment. The concordance lines below show that if *delivery* were a singly occurring lexeme, it would fit into various slots across the corpus. But we find that this is not the case and it always occurs with *yap-*:

yv275 001fb:2	saat arasında	delivery yapılıyor	ruhsatlı tesisatçı licence
yv275 011fb:1	ücretsiz	delivery yapılıyor	duyuru sydney united türk
yv503 006fb:1	yere ücretsiz	delivery yapılıyor	araba parçası bulamadım diye
yv235 005fb:1	saat arasında	delivery yapılıyor	müslümanların gözü aydın
dün021 003Fb:1	edilir ve	delivery yapılıyor	North Sydney'de HELAL ET'le
yv776 001Ba:10	Food evlere	delivery yapılıyor	Helal Meat Helal Chicken Bunun

Clearly, the possibility of *yapılıyor*'s (can be arranged) selecting *delivery* is very high as this verb can select a huge number of nouns (other examples from Ozturk corpus are *satıp*- 'sale', *nakliyat*- 'transport', *tamir*- 'repair' etc.) and this does not seem particularly significant. However, if we look at the possibility of *delivery*'s selecting *yapmak*, the relationship becomes interesting. In other words, *delivery* does not select itself another verb, which, even more interestingly, is in the passive form. Thus, *delivery* and *yap-* co-select each other and other combinations do not seem likely. The next step is to see whether this relationship has been modelled on the collocates of *deliver/delivery* in Australian English. The available range of collocates for *delivery* is as follows:

As a noun in the singular form:

- *Pick up and delivery* (p2234)
- *Next day delivery* (p652)
- *Delivery all suburbs* (p653)
- *Delivery and set up service* (p716)

- *Free delivery all suburbs* (p719)

In the plural form:

- *All Interstate and Country deliveries* (p 551)

- *Deliveries to all areas* (p713)

- *Prompt Daily Deliveries* (p462)

As a verb

-*delivered to your door* (p2144)

-*We deliver* (p504)

-*Deliver all suburbs* (p202)

and with *arrange*:

- *deliveries can be arranged anywhere* (p2144)

(Source: *Yellow Pages 1997*)

As a further step, I have checked these against the evidence from the Bank of English and here is the list of most significant collocates for delivery:

collocate *frequency of occurrence*

special	454
home	433
service	381
your	282
mail	134
free	128
express	90
courier	79
food	88

The point here is that a word form rarely occurs in isolation and when it is used in another language it will inevitably select a number of collocates. The question is "how does this selection occur"? If we hypothesize that the new collocate will be the translation equivalent of its collocate in the original language we can see that this does not always work. In fact, *delivery yap* is a good example of this, as *yap-* in dictionaries appears as an equivalent of *do* and *make*, not an equivalent of *arrange*, the only verb co-occurring with *delivery* as shown in the list above.

Therefore, we need to look for an explanation elsewhere. If we approached the case of *delivery yap* from an equivalence, that is the conventional Language Contact perspective, the perfectly legitimate question that faces us would be: "The word *delivery* has an equivalent in Turkish so why does the Turkish speaker use the English word?" This is not a new question of course. It has surfaced in Myers-Scotton's work recently as the *core lexemes*, and the more generally accepted answer has been psycholinguistic:

The so-called 'unnecessary' loanwords, which were [...] strongly condemned by purists (as when Norwegian immigrants adopted English words for 'fence' or 'river', for which native words were available) [...] reflected the law of least effort. (Haugen 1973:533).

In terms of equivalence, however, what the researcher has in mind is the dictionary equivalent. Two objections arise. Firstly, recent corpus work on bilingual dictionaries shows that the lexicographer's intuition, which has for centuries played a major role in compilation, is misleading. Secondly, parallel corpora in different languages show that equivalences should be considered in relation to respective co-texts. Let us for a moment ignore these points and look at the possible translations or similar words in Turkish. As a native-speaker of Turkish I can immediately think of at least two phrases that can be used in a similar way:

- *eve teslim* (home delivery)
- *kapıdan kapıya nakliyat* (door to door delivery)

Now let us go to the corpus and check how the words *teslim* and *nakliyat*, which supposedly correspond to *delivery*, are used:

ts235 001fc:7	başkonsolosluđuna	teslim	edilecektir
ts235 001fc:8	gibi çek olarak	teslim	edilecektir
ts237 003fb:1	ocak başlarınyń geriye	teslimi	rica olunur
ts237 003fb:1	Kmart'a hemen götürüp	teslim	etmeleri rica olunur
yv235 001fb:1	gemi ile Türkiye'ye	teslimi	garanti edilir eşyalarýnýz
yv756 002fb:1	fýrsat tapusu	teslim	edilecek
yor60 001h:14	işbirlikçi vurgunculara	teslim	etmişolan kiloluk bir hiçlik
yor166 006Ba:1	Belediyesi ilgililerine	teslim	edildi
yv101 001Ba:6	yükü altýnda ruhunu	teslim	etti
yv786 002x:5	Kaderini çaresizliđe mi	teslim	etmelidir

As the occurrences above display, *teslim* frequently collocates with *etmek*, a common delexicalised verb in Turkish along with *yapmak*. However, when it occurs with *olmak*, the meaning changes to 'surrender' and there is a single occurrence of this in the corpus. In any case, none of these occurrences can replace *delivery* except for perhaps the use in yv235 001fb:1, a transport company advertisement:

Eşyalarýnýzýn 6 - 8 hafta içinde gemi ile Türkiye'ye teslimi garanti edilir
(*guaranteed ship delivery of your furniture to Turkey within 6-8 weeks*)

The point here is though, *teslim* is used with respect to removals, i.e. the difference in meaning, and it collocates with *etmek* hence the difference in the environment. In the case of *nakliyat*, on the other hand, there is no difference in terms of the verb co-selected as both *nakliyat* and *delivery* collocate with *yapmak*:

yor166 004fb:1		nakliyat işleri, şehir içi ve şehirlerarasý
yor166 004fb:1	her çeşit	nakliyat yapılýr
yv756 004fb:1	mutlu	nakliyat ve mobilya taşımacýlýđý
yv275 005fb:1	ölçü alýnýr, ücretsiz	nakliyat yapılýr

In the first three concordance lines we see that *nakliyat* is used in the sense of removals and therefore resembles *teslim*. In the fourth line that comes from the advertisement of an insulation company it is used as *delivery*. The full context goes like this:

Evinizi ve iþerinizi yaz sýcađında serin kþsođuđunda sýcak tutmak ister misiniz? Evinizin çatý arasýna ve gerekirse, duvarlara koyacađýnýz. Özel maddelerle sýcak ve sođuk havaný geçmesini engelleyerek yuvanýda ferahlýk ve ekonomi sađarýz. Temiz, ucuz ve dayanýklýmalzeme. Günü 24 saatinde emrinizdeyiz. Ücretsiz ölçüalnýz, ücretsiz nakliyat yapýlýz.

(Want to keep your home cool in summer and warm in winter? With our special roof and wall insulation material we can make your home comfortable at very low cost. Quality, low cost, long life insulation. 24 hour service. Free measure and delivery).

When such an occurrence is found it is of course easy to jump to the conclusion that *delivery yapýlýz* must have been modelled on this collocation *nakliyat yapýlýz*. To ascertain if this really is the case let us look at *delivery yapýlýz* in full context. The first two advertisements come from the same supermarket but the texts are different. The third is a restaurant advertisement and the fourth is about car spare parts:

yv235 005fb

Serdar market iki ay içinde 150 dolarlýk alþ -veriþ yapan müþterilerimize 12 adet Paþabahç çay bardađ veriyoruz marketimizde her türlü Türk gýda maddeleri taze sebze ve meyvalar Türtamek konserve ve reçelleri Anatolia Türk zeytinyađý Üker bisküvi çþitleri bayramlýk pe -re-ja kolanyasýve bayramlýk türk þekerleri Nuhun Ankara makarnasý - amasya bamyasý (telefonla sipariþ alýnýz ve isteyenlerin evlerine akþam saat 6-10 arasýnda delivery yapýlýz).

(Serdar Market - Free gift 12 tea cups for 150 dollars purchase in 2 months. Large range of Turkish food. Fresh veg and fruit. Tinned food and jams. Turkish olive

oil. Biscuits. Rose water and Turkish candies for festivities. Pasta, ocra. (Orders by phone welcome. Deliveries between 6-10pm)

yv275 001fb

Serdar market her türlü türk gıda maddeleri marketimiz 7 gün açıktır. Yazın sıcağında yemek yapma zahmetinden kurtardık sizi. Türkiye'nin hazır zeytinyağlı konserve çibitlerini sizler için getirdik. ... en ucuz fiyat, en kaliteli mal ... telefonla sipariş almayı ve isteyenlerin evlerine akşam saat 6 -10 arasında delivery yapıyoruz.

Serdar Market. Wide range of Turkish food. Open 7 days. We import Turkish tinned food to save you the trouble of cooking in the heat. Lowest prices highest quality. Orders by phone welcome. Deliveries between 6-10pm.

yv275 011fb

Ali baba original charcoal chicken and lamb. Sağlıklı, temiz ve lezzetli çünkü meşhur kömürde pişirilmektedir. Kuzu budu çirmesi, tavuk çirmesi, düğün pilavı ve çoban salatası Niçin düğün, nişan ve partilerinizde misafirlerinizi \$ 5 altında ağırlama imkanından yararlanmıyorsunuz? Ayrıca, çibitli soğuk mezeler, 15 çibit dondurma ve milk shake, ayran bulunur. Sipariş almayı. Ücretsiz delivery yapıyoruz.

(Ali baba original charcoal chicken and lamb. Healthy, clean and tasty. Charcoal grill. Lamb and chicken, pilav and salad. Cater for all functions. Under \$5 per person. Large range of dips, icecream, milkshake, ayran available. Order service. Free delivery)

yv503 006fb

Güçlü bir türk kuruluşu oto parçası her çibit otunun tüm parçaları bulunur ve istediğiniz yere ücretsiz delivery yapıyoruz. araba parçası bulamadın diye dertlenmeyin telefonla sorunuz veya bizzat gelip görün...

(A well-established Turkish spare parts dealer. Any car, free delivery any suburbs. Why worry about spare parts - just phone or visit us).

What we can see here is that out of four occurrences, three uses of *delivery yapýýr* are connected to food delivery. In Turkey there is the expression 'evlere servis' that translates as 'home delivery', when one orders food by phone. So, while on the surface this word finds several equivalents in Turkish, these words fail to meet the intended meaning. If we link this argument to the question of naturalness, summarised in section 4.2.3.1, it may not fully conform to the patterns of Turkish and English but to an Australian Turkish speaker *delivery yapýýr* is a perfectly natural unit and in fact, recalls the proposal by Barnett *et al.* (1994) of the closest natural rendition as a solution to the problem of mismatches in machine translation. This point will be elaborated in section 8.3 below.

So far we have looked at the collocational preferences of delivery in the new environment. Now, let us consider its colligational tendency, the grammatical equivalent of collocation. In all the examples of *delivery* found in the corpus, *yap-* appears in the passive form. While this preference may simply be a bias of the text types in which this collocation occurs (ie. Turkish business advertisements in the community newspapers), it is significant. Particularly, when considered in comparison with a similar tendency developed by the word *register* examined below, we can see that *yap-* is not preferred in the active form as for example in:

?delivery yaparýz

delivery yap+aorist+1st person pl

we arrange deliveries

?delivery yapýyoruz

delivery yap+pres+1st person pl

we arrange deliveries

although both forms are grammatically perfect. Also in The Bank of English texts, we encounter fewer uses of passive construction in that out of 65 co-occurrences of *delivery* and *arrange* only seven appear to be in the passive. Thus, in the mixed

code this tendency may indicate that, although *yap-* can be used with a wide range of nouns in Turkish, the sole presence of *delivery* in its environment may restrict its form. There is a similar argument in 8.1.4.1.

If we return to the opening question of this section then (Is *delivery* a single switch?) it concludes from the foregoing discussion that within the context of this study the answer is in the negative. There is further discussion on *yap-* in section 8.1.5.1 below.

8.1.1.2 Register

The second 'core lexeme' to be treated in this analysis is register. The coreness of this lexeme lies in the fact that, according to the Oxford English-Turkish bilingual dictionary, its equivalent in Turkish as a verb is *kaydetmek*, *tescil etmek*, *kütüğe geçirmek* etc., and *sicil*, *kütük*, *resmi defter*, *kayıt* and so on when used as a noun. In fact, the example below does not seem to be problematic in any way:

yv275 003Ac diyen Mr.Hepburn (*register*) denilen bu kayıt defterine işlenmesi geređi

Seen in full context this use refers to the way an employer should handle the employee records:

yv275 003Ac

Birçok işverenler çalışanları için bir kayıt sistemine sahip değildirler ve bu kayıt her türlü not defterine veya kopyalı defterlere detayla rını işlenmesinden ibarettir." diyen Mr.Hepburn (register) denilen bu kayıt defterine [...] işlenmesi geređni hatırlattı

(Mr Hepburn stated that "most employers do not have the full records of the workers they employ and such a register should contain all personal details" ...)

What seems problematic, however, is another set of concordances where *register* is used differently. As we can see in the three lines below, in terms of co-

selection, *register* behaves in a similar way as *delivery*, yet, it attracts two collocates rather than one, *kađý* or *belge* and the verb *ver-*:

yv96 001fb	verilen söze saygý	register belgesi verilir mesai harici
yv513 004fb	işleri yapýlýr	register kađýdý verilir
yv513 008fb	güleriyle prensibimizdir	register kađýdý verilir eksoz işleri

This unit seems to have two different sources of origin in Australian English. The first one may be the *Certificate of Registration* issued by VicRoads when a vehicle is purchased. The second one, instead, seems to be *the Certificate of Roadworthiness* (RWC), commonly called the green slip. It is issued after a safety assessment has been completed on a vehicle by a licensed tester. In fact, if we examine the lines above in full context we can see that all three occurrences come from *garage* advertisements and refer to second-hand cars:

yv096 001fb

sönmez motor atelyesi - anýnda bakýn - garantili onarýn...register belgesi verilir
- mesai harici tamirat - istasyonumuz sabah saat 7 den akşam saat 7 ye kadar
açkýr...

sönmez auto workshop - instant service - guaranteed repairs - RWC - after hours
service - our workshop is open from 7am to 7pm.

yv513 004fb

galaxy (auto repairs) - oto tamirhanesi - her marka ve her model taşıyıcıların
motor tamiri ve diğer her türlü tamirleri garantili olarak yapýlýr...register kađýdý
verilir.

galaxy auto repairs - car repairers - guaranteed mechanical repairs to all makes and
models - RWC.

yv513 006fb

*auburn oto tamirhanesi - diesel motorlar ve her çeşit motor tamiri ...yapılır.
haftanın 5 günü açıktır - doğruluk ve güler yüzlü prensibimizdir - register kağıdı
verilir.*

auburn auto repairers - repairs to diesel engines and all types of engines...open 5 days - honest and friendly service - RWC.

Two aspects become immediately visible. The first one is to do with the word forms available in the case of contact and the way one of these forms is singled out. We observe that:

- 1) the lemma *register* has a variety of forms in English such as *registration*, *registry*, *registered* and so on,
- 2) out of these forms only one form, *register* is used to form a new unit,
- 3) the form of the lemma used in the original L2 unit is *registration*,
- 4) hence the form of the lemma used in the mixed unit does not match the original form.

Secondly, the environment that comes to be associated with this particular form differs from its environment in L2. While the word *certificate* can be used in Turkish (*sertifika*), it does not seem to be very popular in this collocational pattern as only one such co-occurrence is found (gün091 003Fb:1):

Bu anda rodvorti **sertifikası** veremiyoruz
(presently we cannot provide roadworthy certificate)

This is understandable as *sertifika* is more closely associated with the context of education in Turkish as can be seen in the following set of concordance lines from the corpus:

WOR 005:13	olursanız AECOP Final	Sertifikaşını	almaya hak kazanırsınız
yor050 001fa:3	yüksek okul	sertifikası	hsc memur girişini
yor050 001fa:3	n s w yüksek okul	sertifikası	hsc almalıları

yor050 001fa:3	olmalarý ve n s w okul	sertifikasý	almýþolmalarý þarttýr
yor388 002fb:2	resepsiyon görevlisi	sertifikasý	yýllýk ipbilgisayar ve
yv776 012x:1	muadil bir baþka eđitim	sertifikasýna	sahip olmalýdýrlar

In fact, *okul* 'school' and *eđitim* 'education' figure frequently in its environment and its use seems limited in this respect. Two new collocates are offered instead: *kađýt* and *belge* and as can be seen in all the examples the relationship is of izafet (annexation) construction (Lewis:1967:41-42):

In English one noun may qualify another in two ways. In the first, the qualifying is put into genitive: Land's End, St. Antony's College, soldiers of the Queen. In the second, no grammatical mechanism is involved; Lane End, Oxford University, Palace guard. The two types of izafet correspond fairly closely to these two English patterns, with the difference that in both Turkish types the qualified noun takes the third-person suffix.

Hence, qualified by register, the two collocates appear as *belgesi* and *kađýtý*. *The Oxford Turkish-English dictionary* gives us the following equivalents of *kađýt*: 'paper; letter; playing-card' and the following equivalents of *belge*: 'document; certificate; report; voucher'. However the lexical environment of these nouns in the Ozturk corpus shows that these meanings are rarely the ones intended as the one related to vehicles:

DOM 004:1	evlenme cüzdaný nüfus	kađýdý	banka defterleri gibi
yv513 002Ac:2	Bunun dýþýnda nüfus	kađýdý	deniz aþýrý ülkelerden alýnan
yor269 008Ac:2	Avustralya'nýn Dođum	Kađýdýnýn	Ýngiltere'den alýnacađýný
tr147 001X:1	söylemelisiniz ve satýþ	kađýdýna	yazýlmasýna dikkat
WOM 001:13	veya vergi pullarý	kađýdý	tax stamp sheet verir
WOM 001:23	para ödüyorsunuz ücret	kađýdýnda	belirtilmemiþse Emekliye

In this set we can see that *kađýt* is used in the translations of birth certificate, pay slip, tax stamp sheet and so on. Importantly, the only strong collocation in Turkish can be seen in the first two lines, *nüfus kađýdý* that corresponds to identity card in English and the remaining senses are all new units in Turkish. In the second set below, we see the environment of *belge*:

DOM 004:1	için mahkemeye çağrý	belgesi	verecektir Duruřma tarihine
twg 034X:1	jimnastik salonu üyelik	belgesini	de bulundurmayý ihmal
WOM 001:13	sonra size bir grup	belgesi	group certificate veya vergi
WOM 001:13	sheet verir Bu grup	belgesi	ve vergi pullarý kađýdý
WOR 004:25	vatandařlýđý	belgesi	Evlenme cüzdam adýnýzý
yor060 002fa:1	iki giriřimde býrakma	belgesi	1966'ya kadar olan sýnavlar
yor060 002fa:1	dersten geçme orta okul	belgesi	1975'e kadar olan sýnavlar
yor146 007Aa:1	kiřiye yurttařlýk	belgesi	verildiđi de bildirildi \tx
yor259 006fc:1	bařlýklý bu iftira	belgesini	yazan ve yayınlayanlar
yor259 008fa:1	alýnabilir revs	belgesi	elinizden alýnmaya karřý
yor259 009fa:1	bir tartýřma	belgesi	hazýrlanacaktır daha fazla
yv285 001fb:1	yazýlý tam garanti	belgesi	verilir araba satýn alýnýr ve
yv513 004Ac:4	burada evlilik dođum	belgesi	oto ehliyeti gibi kiřisel
yv766 001fa:1	her kazanan bir bařarý	belgesi	para ve uluslararası bir

Also here the same argument holds, *belge* is used in the cases of membership certificate, group certificate, citizenship certificate, guarantee certificate and also birth certificate but again the context is not related to vehicle purchase. Its strong collocations in Turkish are *orta okul belgesi* and *garanti belgesi* and the remaining ones are new units. In other words, these two words have developed different co-occurrence tendencies from the ones we are familiar with in L1 and L2.

The co-selected verb also deserves attention. In the original English texts of registration and roadworthy certificates the following verbs are used and in the passive form:

is issued

is obtained

and also provide is used but in the active form. So *register* as a verb does not figure in these texts and therefore we cannot assume that the borrowed form is the uninflected form of the verb. In our examples, on the other hand, the verb that occurs in the environment of *register* is *ver-* that is translated as *give* in the dictionary. The behaviour of this verb in the corpus, however, shows that it also has a tendency towards delexicalization although it has hardly been acknowledged in the literature (see 8.1.5 below). Contrary to the exclusive use of passive form of *yap-* with delivery, here we have seen *ver-*, in one example, in the active form:

Bu anda rodvorti **sertifikasý** veremiyoruz
(presently we cannot provide the roadworthy certificate)

This might be to do with the differences in the use of *ver-* in written and spoken Turkish as can be seen in Pierce's (1962a) list of frequency counts of the most common twenty words in Turkish (see 8.1.5). Contrary to the most frequent delexical verbs *yapmak*, *etmek* and *olmak* that occur more commonly in the written language, *ver-* has higher frequency in the spoken Turkish. This can be seen in the concordance line above that is a printed sentence from a newspaper advertisement text, yet, it is in the spoken style as opposed to the other three lines where RWC is used with *verilir*. Also when compared to *yapýlýr* in terms of frequency, we can see a significant difference. There are 95 occurrences of *yapýlýr* whereas *verilir* occurs 47 times in the whole corpus. So the passive forms of these verbs are not distributed evenly across texts, and with respect to their lexical environment this is clearly important information. The differences in the behaviour of these verbs are further discussed in section 8.1.5.

What does this type of analysis on an emerging lexical unit such as *register belgesi verilir* tell us about language contact data? Firstly, it is misleading to treat a loanword like *register* as a singly occurring lexeme as it never occurs singly but always in this particular environment in the host language. Secondly, it is unhelpful to try to justify the emergence of the mixed unit on the basis of L1 and L2 as it does not match the patterns of either of the languages. Such mixed units, however, recall the the units of translation that "neither match the unit of meaning in the SL nor that in the TL" (Tognini-Bonelli 1996b:206). This notion has been proposed in a recent multilingual research project conducted on parallel corpora (Sinclair et al. 1996), and promises to be the base for the mixed code lexicon discussed in language contact settings. This point will be discussed further in section 9.1.

8.1.2 The question of "cultural borrowings"

A long-unresolved issue in the area of language contact is the distinction between *borrowing* and *codeswitching* that has divided language contact researchers for the last five decades. It is important to review this distinction, with respect to corpus research as well as written data. Recently, Myers-Scotton questioned this distinction (1993a) and employed new criteria to distinguish and relate the two. In her model, there are two types of borrowed forms that are *cultural* lexemes and *core* lexemes. Such a distinction is not new of course and cultural loans have always been listed to comment on the history of cultural contacts (eg. Sapir 1921:Chap IX). In her analysis, Myers-Scotton disposes of cultural loans and emphasizes the importance of the other category: "Core borrowed lexemes are taken into the language even though the recipient language already has lexemes of its own to encode the concepts or objects in question (e.g. town)" (Myers-Scotton 1993a:5). However, recent work on foreign material in monolingual and bilingual corpora (Kurtböke and Potter 1998) shows that cultural vs core distinction as a central issue does not take us far in our analysis. There are various reasons for this. Firstly, it is not always easy to separate core from cultural borrowings as once-cultural loans gradually lose their cultural content and come to be seen as core borrowings (Duranti 1997).

Secondly, a lemma has a variety of word forms. A particular form of a 'core lexeme' may be culturally significant and end up as a borrowed form in the other language. And when this happens, it is not always the most frequently-used form or the most common meaning that is borrowed. Moreover, also cultural words form collocations with words in the host language that they may not normally co-select in the language of origin. The case studies presented in 8.1.1.2 (*register*) and 8.1.2.2 (*grants*) are convincing example in this respect.

Thirdly, if the main issue is 'to reveal the internal operations of language' as Myers-Scotton claims, cultural forms likewise display patterns that are well worth observing. Thus, the fact that they are core or cultural is irrelevant as long as they provide insights into the way mixed code works. Examples of this can be found in

the Bank of English, which has currently 323 million words of running text and covers a period of six years (1990 to 1996).

A frequent single cultural borrowing is *satay* 102 occurrences in the Bank of English. It significantly collocates with *chicken* and *beef*. What is of interest to us is the flexible order it displays as in *chicken satay* and *satay chicken* or *beef satay* and *satay beef*.

Another such cultural form is *balti* that has been popular especially in and around Birmingham, UK for the last two decades. There are 217 occurrences of this cultural borrowing in the Bank of English. As in the case of *satay*, typical collocates of *balti* are *chicken* and *meat*. Also *balti* is used with these collocates in flexible order as in *meat balti* and *balti meat*, *chicken balti* and *balti chicken*.

The question here is, do such patterns conform to the rules of the *Matrix* or *Embedded Language* or neither? There is no answer to this question within Myers-Scotton's model as she claims (1992): "Cultural B forms [...] may well occur alongside Core B forms in ML+EL constituents, they would enter the ML with or without codeswitching, since they indeed do fill lexical gaps. Many of them are nouns because they stand for new objects/concepts." There she subscribes to the slot-and-filler approach identified by Sinclair (summarised in section 4.2.3.1). Unless we reconsider the place of 'cultural borrowings' with respect to their role in forming lexical units and their function with respect to the position they occupy in those units, the insight they provide will be ignored. As can be seen from the examples above, the observation that a lexical item is a cultural loan does not tell us much about the nature of our data. What we discard as a cultural loan may determine, in many cases, the shape that its environment will take and provide insights into the internal operations of the languages under examination. However, one question remains open and more corpus work is needed to find the answer. Is such use (ie. reversible word order) specific to certain genres (eg. culinary) or is it a widely observed phenomenon across corpora¹³?

¹³ I am grateful to Ramesh Krishnamurthy for these examples and very useful discussions on corpus evidence in general.

8.1.2.1 Ombudsman

Here we have ombudsman, a cultural loan in Myers-Scotton's terms, that occurs seven times in the corpus. This lexeme provides insights in a number of ways and the dictionary confirms that there is no equivalent for this word in Turkish:

ombudsman *Öel kiplerin haklarýný koruyan memur*

(Oxford English-Turkish dictionary 2nd ed. p372).

This definition can be translated back into English as 'civil servant who defends the rights of special persons'. Such a translation equivalent of course does not help us understand what exactly it means. Now let us look at the concordance lines and the way *ombudsman* is used in Turkish texts:

RIG 005:1	Commonwealth	Ombudsman Federal VatandaşSözcüsü
RIG 005:9	Commonwealth	Ombudsman's Office ile ilişki kurmak
WOM 001:18	Bürosu Müdürü	Ombudsman ile temasa geçiniz Not
yor060 005Ab:4	özerk mahkeme dýlýnda	"Ombudsman" Halka karşı ihlenen suçları
yv513 005fa:1	commonwealth	ombudsman[a] başvurmasını önerdi ve bay
yv513 005fa:1	vatandaşözcüsü'nün	ombudsman ofisinden şikayetini
yv533 004X:8	State	Ombudsman[a] bir şikayette bulunmalýsýn

Firstly, a word on the context span. Extensive corpus analysis has shown that in order to get a complete picture of the collocates of a word, and any loan for that matter, both the right and the left of the node should be considered. This point becomes even more interesting when the language involved is an SOV language like Turkish where the context span can be much larger than in English due to the differences in sentence structure and due to the flexibility in word order. In order to illustrate this point the first text below is given in unabridged form. The words that occur frequently in the environment of *ombudsman* and its proposed equivalent *vatandaşöcüsü* are highlighted in bold type:

RIG 005

Commonwealth Ombudsman (Federal Vatandaş Sözcüsü)

Avustralya'da **şikayet etme hakkınız** vardır.

Commonwealth Ombudsman (Federal Vatandaş Sözcüsü)

In Australia you have a right to complain.

(Paragraph 1)

Federal Vatandaş Sözcüsü, haklarında şikayette bulunulan Federal hükümet kurum ve kuruluşlarının haksız, yanlış mantıksız ya da **yasa dışı** hareket edip etmediklerini öğrenmek amacıyla bu şikayetleri araştırarak baskınsız bir kişidir. **Vatandaş Sözcüsü** Sosyal Güvenlik ödenekleri; Göç; Vergi; Avustralya Federal Polisi; Avustralya Gümrüğü ve diğer **birçok Federal hükümet** kurum ve kuruluşuyla ilgili **şikayetleri araştırabilir** **Hükümet** kurum ya da kuruluşunun yanlış **haksız** ya da **yasa dışı** hareket ettiği sonucuna varması durumunda Vatandaş Sözcüsü, kurum ya da kuruluşu neden olduğu sorunu çözmesi önerisinde bulunabilir. Bu, kararın geri alınması ya da değiştirilmesi, özür dileme veya özel durumlarda **haksızlığa** uğrayan kişiye tazminat ödenmesi anlamına gelebilir. **Vatandaş Sözcüsü araştırılmaları** ücretsiz ve gizlidir

The Federal Ombudsman is a person who investigates complaints about Federal government institutions as to whether they have acted unfairly, unreasonably or illegally. The Ombudsman may investigate complaints about Migration, Tax, Australian Federal Police, Australian Customs and many other Federal government institutions. If the Ombudsman comes to the conclusion that the government institution has acted unfairly or illegally, then he may advise the institution to offer a compensation. This may take the form of an apology or money refund. The investigations of the Ombudsman are free of charge and confidential.

(Paragraph 2)

Şikayet yöntemi

Avustralya'da bulunan herhangi bir **Federal Vatandaş Sözcüsü ofisine** telefon edebilir, yazabilir ya da **başvurabilirsiniz** Ayrıca, bir arkadaş ya da yakınınız **adına şikayetle bulunabilirsiniz** (Bu kişinin iznini almak koşuluyla). Bir avukat, Milletvekili ya da bir göçmen kuruluşu sizin adınıza **şikayetle bulunabilir** **Şikayetle bulunmak** amacıyla **bir Vatandaş Sözcüsü ofisine başvurduğunuzda** bir araştırma görevlisi **şikayetinizi** dinleyecek **ve Vatandaş Sözcüsünün** size nasıl yardımcı olabileceğine ilişkin açıklamalarda bulunacaktır. Araştırma görevlisi **şikayetinizin** ne olduğunu soracaktır. Ayrıca, **araştırılmaya** yardımcı olması amacıyla, sizden mektup ya da diğer belgeleri sağlamanızı da isteyebilir. **Vatandaş Sözcüsü** politikacılar, mahkeme kararları, **Eyalet** ya da **yerel hükümet** kuruluşları veya özel **başvuru** ya da firmaların eylemleriyle ilgili **şikayetler** hakkında araştırma yürütemez. **Vatandaş Sözcüsü** sorununuzu ilk önce kuruluş ile çözümleme girişiminde bulunmamanız halinde, **şikayetinizle** ilgili **araştırılmaları** yürütmeyebilir.

How to complain

You can telephone, write or apply in person to any Ombudsman office in Australia. You can also lodge a complaint on behalf of a friend or a relative (with their consent). A lawyer, member of parliament or an immigration agency may also apply on your behalf. When you lodge a complaint with the Ombudsman's Office, an assistant will listen to you and explain how they can help. The assistant will ask you about your complaint and may request further documentation to support of your case. The Ombudsman may not investigate the complaints about the State or Local government institutions, private companies or individuals. The Ombudsman may stop the investigation if you do not seek reconciliation with the institution in question, at the initial stage.

(Paragraph 3)

Tercümana gereksiniminiz var mı?

Yetenmesi halinde, **Vatandaş Sözcüsü** tercüman sağlayabilir.

Yakınızdaki **Federal Vatandaş Sözcüsü Ofisi (Commonwealth Ombudsman's Office)** ile ilişki kurmak için, Avustralya'nın her yerinden kentiçi ücreti karşılığında 1800 133 057 numaralı telefonu arayınız.

Do you need an interpreter?

If requested, the Ombudsman can provide you with an interpreter.

You can ring your nearest Ombudsman's office from anywhere in Australia at the cost of a local call: 1800 133 057.

WOM 001

Haksýz yere iþden çýkarýlmanýz gibi dava konusu olacak bir durumunuz varsa...bir avukuta danýþýnýz. Hangi **Eyalette** bulunuyorsanýz, aþaðýda adresleri **verilen Resmi þikayetler Bürosu Müdürü (Ombudsman)** ile **temasa geçiniz**.

If you are facing a situation such as unfair dismissal...seek advice from a lawyer. Contact the Ombudsman's Office in your state at the addresses below.

yor060 005Ab

"Ombudsman" (Halka karþý iþlenen suçlarý araþtırmakla görevli en yetkili devlet görevlisi) istendiði zaman polisle ilgili **araþtırmalarda bulunabilecektir**.

Ombudsman (State servant in charge of investigating crimes committed against citizens) may investigate the police upon request.

yv513 005fa

Bu sorunu arkadaşý ile tartýþtı. Bay Türkmen ona **federal vatandaş sözcüsü'ne (commonwealth ombudsman) baþ vurmaşını** önerdi...

He discussed the problem with his friend. Mr Türkmen advised him to apply to the federal ombudsman...

Vatandaş sözcüsü'nün (ombudsman) ofisinden þikayetini inceleyeceklerisöylendi ve bir süre sonra da Bay Ýleri'ye telecom'un fazladan ödemiþ olduđu tüm kirayý geri vermeyi kabul ettiđini bildirdi.

The Ombudsman's office said his complaint was being investigated and after a while informed him that Telecom accepted to return all the extra money he had paid.

yv533 004X

Hemen **Eyalet Vatandaş sözcüsü'ne (State Ombudsman)** bir þikayette bulunmalýsýn
You must immediately lodge a complaint with the State Ombudsman.

As can be seen in the underlined sentence in the first text (paragraph 1), as well as in the others, **þikayet** 'complaint', the most common word in the environment of *ombudsman*, can be found as far apart as 17 words to the right of the node. Also, positionally, we observe that **þikayet** appears to the left of the node but later on also to its right in the underlined sentence (paragraph 2). This information is important in a number of ways. From the point of view of language contact, it shows us how naive the 'singly occurring lexeme' approach is and studying loans without any consideration of their lexical environment. That a switch may occur within or across sentences does not tell us much either about the context span of a node word, typically a loan in this case. While factors such as colligation and

cohesion (Hoey 1998) will come into the picture too, the location of switches cannot be considered without reference to the location of their collocates. From the point of view of lexicographical description, particularly in terms of bilingual dictionaries, the fact that the context span can extend over a certain number of words is clearly significant. And finally, from the point of view of translation, and especially in the identification of comparable units of meaning between and among languages, the context span will play an important role.

Second point to be made in relation to this 'cultural' word in the Turkish texts is that some of its collocates are also noticeable in the Bank of English. However, they do not display the same order of frequency. In the list below there are the significant collocates for the node *ombudsman* and their frequency in the Oznews component of The Bank of English. Their frequency of occurrence in the Ozturk corpus, whether in English or as the proposed Turkish equivalent, is also indicated in the rightmost column:

	<u>oznews corpus</u>	<u>ozturk corpus</u>
the	170	-
commonwealth	26	7
office	19	5
by	18	-
banking	15	-
industry	14	-
report	13	-
telecommunications	7	-
complaints	7	15
commission	7	-
tax	6	-
investigate	6	10
federal	6	13

We have seen in the texts above that the most frequently highlighted word is *şikayet* 'complaint' found in the immediate environment of *ombudsman* 15 times in

the Turkish texts. This frequency is, of course, more significant here as the Ozturk corpus is much smaller than the Oznews subcorpus of The Bank of English. Another such word is *arabý* - 'investigate' and its various forms appearing 10 times in the Turkish texts. What is even more interesting is the fact that *arabý* - also attracts *þikayet* as for example in the cases of *þikayetleri arabý* - 'investigate complaints' in the first paragraph. This tendency shows that we can expect to see both words in the environment of *ombudsman* in the Turkish text.

There is one more reason why I have chosen *ombudsman*, as an interesting example rather than simply as a cultural word. That is its relevance to the ongoing question of *integration*. While Myers-Scotton uses the criterion of frequency to distinguish the integrated from the unintegrated forms, the problems of this approach for the language contact researcher are discussed in the preceding and following sections. There is a much more evident feature of the so-called unintegrated forms, at least very evident in the Ozturk corpus, than their frequency. That is the translation provided in brackets immediately after the word (cf. *flagging* in spoken language). It is obvious that Australian Turks have proposed the unit *vatandaş sözcüsü* (lit. citizens' spokesman) as a possible equivalent for *ombudsman*, but there are also the other options like *Resmi Þikayetler Bürosu Müdürü* 'Director of the Office of Formal Complaints' as in WOM 001 and *Halka karşı ibnen suçları araþtırmakla görevli en yetkili devlet görevlisi* 'State servant in charge of investigating crimes committed against citizens' as in yor060 005Ab. Presently the two forms co-occur in Australian Turkish texts and in the case of full integration as such, the bracketed form can be expected to disappear. But the question of integration is more of interest to the lexicographer than to the language contact researcher and this point is discussed below. There are also other similar strategies activated in the event of translation, and these are dealt with in section 5.4.

8.1.2.2 Grants

Here we have an ordinary dictionary word as it appears in the Oxford English-Turkish dictionary:

grant Ýmtiyaz; hibe; yardým; baðýþ(lama); tahsisat, ödenek; burs.

However, this word gains cultural significance in the context of immigration in Australia and becomes interesting for the purposes of language contact research. Its cultural load comes from the fact that the Australian government offers financial support to promote migrant community activities such as cultural festivals, art exhibitions, publications and similar projects. That is why this word is of particular relevance to an Australian citizen of ethnic background. When studied in full context (see below), a set of words such as *multicultural*, *migrant*, *ethnic*, *ethnic community*, NESB specifically related to the Australian setting will become immediately apparent:

yv111 003Ac

Eđitim Bakaný Mr.Fife, part-time etnik okullar iđin Commonwealth Hükümetince verilecek grantların talebe başına yıllık otuz dolar olarak hesaplanacağını söyledi.

The Minister of Education, Mr Fife, stated that the grants allocated to part-time ethnic schools will be calculated as thirty dollar per student.

yv245 003Ab

Bu altý sendikanýn birđok üyelerinin İngilizce konuşulmayan ülkelerden geldiklerini belirten Bakan, "Göçmen Kadýnlara daha çok fay da sağlayacaktır" diyerek, Grantların göçmen kadýnlar iđin ...potansiyle sahip olduğunu söyledi.

The Minister stated that most members of these six unions come from Non-English speaking countries, and that the grants have a higher potential for migrant women. He said "Migrant women will benefit more"...

yv806 003Ab

'Bu yıl Victoria Etnik Ýpler Komisyonu'nun Small Grants programýna göre Victoria'nýn çokkültürlü topluma \$ 400.000 tahsis edilmiştir. Koalisyon Hükümetimizce, çokkültürlü Her Politikasında açıkladýđ gibi, Small Grants Programý 1996-97 devresinde \$ 750.000 'a çıkarılacaktır' diyen Victoria Eyalet Bakanı Jeff Kennett, bu sonuca ulađnaktan büyük mutluluk duyduđunu belirtti.

The Victorian Premier Jeff Kennett expressed his satisfaction with the outcome and said, "This year, within the Ethnic Affairs Small Grants program, Victoria Multicultural Community has been allocated \$ 400,000. As stated by the Multicultural Affairs Policy, in the 1996-97 period of the Small Grants Program, our Coalition government will raise this to \$ 750,000".

yv806 004Ac

Etnik Ýpler Komisyonu Grantları yükselecek.

Ethnic Affairs Commission Grants to rise.

Victoria Eyaleti 'okkültürlük Komisyonu tarafından yürüden Küçük Devlet Yardımları Programý 95/96 yılı için \$ 400,000'dan 96/97 yılı için \$ 750,000'a yükseltilecektir ki bu grantlar,

toplum organizasyonlarına, özellikle yaşlılar için ve onların gelenek ve kültürlerinin sürdürülmesini teşvik etmek için verilmektedir.

Small State Grants Program administered by the Victorian Government Multicultural Commission will raise the funding from \$ 400,000 for the 95/96 period to \$ 750,000 for the 96/97 period. These grants are given to community organizations, especially for the aged and to encourage the maintenance of their culture and their customs.

...iki dil bilmeyen öğretmenlerin diğer bir dili akıcı olarak konuşup, yazmalarını grantlarla sağlanacak.

...grants will enable monolingual teachers to speak and write a second language fluently

ts335 001fa:1

Yazarların bu anda üzerinde çalıştıkları projeleri tamamlamaları için verilecek proje yardım fonlarının (project assistance grants) sayısının sınırlı olması.

Project assistance grants, which will be given to authors to complete their work in progress, are limited in number.

WOM code 001

Aborijinler için Mali Yardım Programı (Aboriginal Study Grants Scheme) Aborijinler ve Torres Strait adalarında yaşayan yerliler, okulu bitirmeden ayrıldıktan sonra öğrenimlerine devam etmek için bu programa göre mali yardım alabilirler.

Aboriginal Study Grants Scheme. As part of this scheme, Aborigines and Torres Strait Islanders may receive funding to complete their education if they have interrupted their studies.

yor156 001fa

Multicultural education grants 1986. Commonwealth hükümetinin 1986 yılında, çok kültürlü eğitimle ilgili olarak, eyalet para yardımları (okul yardımları) yasası yoluyla, her eyalete fon sağlanması beklenmektedir.

Multicultural education grants 1986. The Commonwealth Government is expected to provide funds to each state for multicultural education as part of state education funds act in 1986.

yor259 002AB

New South Wales Başbakanı Nick Greiner, 1990 yılında etnik toplumlara verilen parasal yardımları (grantları) açıkladı.

New South Wales Premier Nick Greiner announced the grants to be allocated to ethnic communities.

CHI 006

Çok gerekli olan bu değişiklikleri gerçekleştirmek için, "Supplementary Grants" diye bilinen Yardımçı Ödenekler Programı oluşturulmuştur.

In order to make these necessary changes a Supplementary Grants Program has been created.

As in the previous case study *ombudsman*, discarding a word such as grants from our analysis on the grounds that it is a cultural lexeme would mean discarding also a group of words whose group membership may provide insights. It can be said

that the identification of a lexical set as such around *grants* may indicate that certain words will play a more important role in the construction of the mixed code lexicon. Interestingly, as the lines above show, this word is used 16 times in the corpus in the plural form. Before examining this particular occurrence and its characteristics in the corpus, let us first check the common collocates of the singular and the plural forms of *grant* as well as those they do not share, in the *Oznews* component of the Bank of English:

Common collocates: '*grant*' and '*grants*'

dollar	arts
government	awarded
federal	for
million	to
council	Queensland
research	received
state	

Significant collocates - '*grant*'

a/an
help
fund
applications
made
development
under
money
aid

Significant collocates - '*grants*'

commonwealth	in
assistance	allocated
commission	were
financial	projects

community	announced
states	funding
given	totalling
through	

These lists clearly illustrate that different forms of a lemma have different collocates and we cannot assume that the singular and the plural forms of a noun will operate in the same way in text: "the use of the pluralization of a noun is often not as simple as conveying only the additional information that there are more than one of these things referred to" (Hays 1997:199). Such valuable information is lost when lexemes are treated as individual loans in data regardless of their form as for example in Myers-Scotton's list of 'singly-occurring switches' where 141 nouns are reported. However, she considers the status of the plural morpheme in the mental lexicon (1993a:132). This characteristic of *grants* deserves attention here, as well. Frequently in the texts above the plural marking on this noun is Turkish, thus giving us the form *grantlar*. While we do not come across double morphology to mark plurality as in Myers-Scotton's data (1993a:61-63) like **grantslar* in the Ozturk corpus, this combination is important as it may reveal the way the plural form of a lemma is stored in the mental lexicon of a Turkish speaker. Double morphology in the case of borrowed plurals, however, is not uncommon in Turkish as for example in *evlatlar*, an Arabic loan meaning *evlat* 'children' (sing. *veled*) marked with *-lar*, the Turkish plural. This point is discussed at length in 8.1.7 alongside a Turkish derivational morpheme attached to a switch of non-cultural kind.

If we look at the use of *grants* from the point of view translation, we find a similar situation to that of *ombudsman*. In CHI 006 above *grants* is translated as *ölenek*. However, we find that also benefit is translated as *ölenek* in

WOM 001:1 Unemployment benefit işsizlik ödeneđi

as well as *allowance* in

yv786 004x:13 Mature Age Partner Allowance alan kimseler...ödeneklerini almaya devam edebileceklerdir

This shows that there is an overlap and the differences in meaning can only be identified after the comparable units in parallel corpora have been established. There are also other proposed equivalents for *grants* such as *yardým* and *fon*. All of the translation units forming around these are used in the same way and their differences become less noticeable in the new setting with respect to the original units of meaning. While all of these alternatives (*ödenek*, *fon*, *yardým*) frequently attract *almak* 'receive' and *yararlanmak* 'to take advantage of' with the preceding noun in the ablative:

fon almak, fondan yararlanmak
 ödenek almak, ödenekten yararlanmak
 yardým almak, yardýmndan yararlanmak

there are restrictions with respect to the use of delexical verbs. While *grants*, *ödenek* and *fon* may be followed by the verb *ver-*, *yardým* in the sense of *grants* is exclusively used with *yap-*, which can again be located on the right or the left of the node depending on the structure preferred:

ts234 001fa:1 federal hükümetce **yapýlan** 'yardým fonu size para
 ts275 002fa:1 ve evli çiftler için **yapýlan** özel **yardým** tek yapıyan
 ts335 001fa:1 sýnýrlýdýr dolara kadar **yapýlacak** olan **yardým** projenin
 yor156 001fa:1 özel proje yardýmлары **yapýlacak** **yardým**larda personel
 yor161 005fa:1 zamanda çocuklar için **yapýlmakta** olan ek **yardým** da haftada

CHI 008:3 Dolar'a kadar **yardým yapýlýr** Yukarıda belirtilen en yüksek
 DOM 001:13 yardýmý ve kira **yardýmý yapar** Bu olanakları bir Homeswest
 WOM 001:1 para **yardýmý yapýlýr** Buna karşılık işverenler de
 yv111 003Ac:3 okullara para **yardýmý yapýlmasýný** öngören bir Kanun
 yv111 004AAc:3 devamý için **yardým yapmaktadır**. Orta okulun son iki

It is also worth mentioning that *ödenek* is more frequently used in the corpus than the other three words of similar nature as we can see in the frequency counts below:

ödenek - 30
 ödeneđi - 88
 ödenekler - 20

mali yardým - 34
para yardýmý - 16
parasal yardým - 2

fon - 22
fonlar - 5

While these words may occur individually, they do often appear within each other's context span and as stated above the semantic boundaries tend to become blurred. However, rather than seeing this particular aspect of the mixed units in terms of matches and mismatches between the migrant's mother tongue and the language of the host country or discard them as culture-bound items, we should accept them as the basis of the empty lexicon discussed in Chapter 9.

8.1.3 The question of "islands"

Most approaches to language contact data tend to analyse mixed units in terms of sentence constituents. In such approaches the relationship between the mixed elements are ignored and the concern remains as to which participating language is lending the structure.

This point is particularly important in Myers-Scotton's MLF model because once the matrix language is set, it is the one to prepare the sites into which EL content morphemes are inserted (1993a:137). Myers-Scotton finds no problems with ML islands:

Explaining ML islands appears to be a very straightforward matter, requiring little discussion. After all, it is the ML's set of grammatical procedures which is continuously active during the production of ML+EL constituents. The only EL activity comes when content morphemes are inserted into sites prepared by ML lemmas. Therefore, producing ML islands is simply a matter of never allowing any activity from the EL.

However, she acknowledges that they have not been systematically studied: "While ML islands have hardly been studied, and certainly not systematically, it

seems they can occur at any point in an utterance" (1993a:137). If we look closely at what Myers-Scotton calls islands, regardless of their affiliation, we can see that she is talking about strong collocations. Let us take her examples (1993a:148):

1. ML+EL constituents

u-na-m-time

2. ML islands

inaanza usiku

3.EL islands

throughout the day

Whether the first two examples are strong collocations in her research setting or not is to be determined. However, every speaker of English would agree that ‘throughout the day’ is a strong collocation in the language. L1 and L2 words do not always have to be mixed in order to be called ML + EL islands. As the case study in section 8.1.4.2 illustrates to go to the corpus with such assumptions is risky. The MLF model has been postulated in a way that enables Myers-Scotton to set aside singly occurring forms as tests for a series of hypotheses. Consequently, islands are seen as "composed of at least two lexemes/morphemes in a hierarchical relationship" and even more importantly "no single EL forms may be islands" (1993a:138). Obviously, the MLF model does not draw on lexicogrammar nor on recent corpus research with lexical emphasis.

In contrast, Sinclair's approach to lexical relationships is the recognition of "the unit of meaning" where even an island as such may be considered a single form. This is further discussed in Chapter 9.

8.1.3.1 Mature Age Allowance

Typically, we find islands such as mature age allowance in the Ozturk corpus:

yv756 003x:5

Mature Age Allowance alan kiplerin eplerine verilen

(Partners of Mature Age Allowance receivers)

Mature Age Partner Allowance gibi ödenekleri alabilmek için başvuruda bulunmuş olan kadınlar..

(Women who lodge a claim for Mature Age Partner Allowance)

yv786 004x:13

Mature Age Partner Allowance alan kimseler...ödeneklerini almaya devam edebileceklerdir

(Persons receiving Mature Age Partner Allowance...will continue to receive their allowances)

Firstly, let us consider the environment, again. If we check *mature age allowance* in the Oznews subcorpus of the Bank of English we find that the two verbs that appear by this unit are:

to claim

to qualify for

However, the Turkish verb used in this context is almak, 'to get, to receive'. This choice further supports the discussion above that assuming one-to-one correspondence between languages is risky, and establishing comparable meaning units is a better option. This is done in a contact setting such as Australia, by the migrants themselves.

The colligational aspect of the phrases above deserves as much attention. We have seen before that while collocation is the lexical preference of a word, colligation has been defined as "the grammatical and positional preferences of a word" (Hoey 1998), that is the co-selection by grammatical class.

What interests us here is that in all three examples, a participle follows the unit *Mature Age Allowance*, formed by adding *-en* to the stem (Lewis 1967:158):

-alan (receiving)

-verilen (being given)

-olan (being)

As we shall see later in section 8.1.5.1 on *yap-*, this is a highly frequent form, where this construction appears over 80 times and contributes to its progressive delexicalization.

As Lewis (1967:158) explains: "These words function as adjectives or nouns" and "exercise the same governance as the corresponding finite verb". While the examples here are limited in number, they raise several questions in relation to these two points. Firstly, section 7.5.1 on nominalization has discussed the preferences for nominal construction in languages. Whether the tendency to nominalize becomes stronger in contact situations is a question yet to be systematically investigated. However, the data in hand indicates that this may well be the case. In the examples, *alan* 'receiving' forms a relative construction where the pronoun is in the nominative:

Mature Age Allowance alan kişiler
(persons who receive Mature Age Allowance)

Mature Age Partner Allowance alan kimseler
(persons who receive Mature Age Partner Allowance)

It would be interesting to see in a larger corpus of mixed data whether the sheer presence of a unit such as the *Mature Age Allowance* imposes constraints on the structure of the phrase in which it occurs and requires a nominalization of this kind. In the corpus we find 88 occurrences of *alan*, and out of these the nominalizations around *allowance* or a similar unit of meaning count 22. In the case of *verilen* 'being given', on the other hand, we notice the dative governed by the participle

eşlerine (to the partners of)

but we also notice that *verilen* is in the passive form. There are 86 occurrences of this particular form in the corpus although only 6 of these are used in a similar way with a noun in the dative preceding *verilen*. These are found specifically in the context of *age allowance* or similar units.

On the contrary, the third participle in the set, *olan* is of course the most delexicalized one with respect to the other two verbs discussed here. In the Ozturk corpus there are 983 occurrences of *olan*, the 10th most frequent word in the entire corpus. This particular pattern where the preceding participle has the past tense marker *-miş* occurs 75 times out of the total 983 occurrences of *olan*:

Mature Age Partner Allowance gibi ödenekleri alabilmek için başvuruda bulunmuş olan kadınlar

(women who have already lodged a claim for benefits such as Mature Age Allowance)

While the literary translation of *olan* is usually given as *being* or *becoming* (Lewis 1967:158), it is obvious that this particular form of the verb *ol-* has long been delexicalised as its frequency of occurrence confirms and it is best not to translate it in this way. While these examples do not provide sufficient evidence to confirm whether the use of such nominalizations in a contact setting can be attributed to the presence of an EL island and its colligational tendency, it is a step forward from the current description of such units. In the section below I consider another such case that would misleadingly appear as an ML island within Myers-Scotton's MLF model.

8.1.3.2 Aile birleşimi

In the Ozturk Corpus there are some patterns that might easily escape attention as ordinary Turkish words or Turkish islands when examined individually. These are *aile* (family), *göç* (migration), *toplum* (community), *yetişkin* (adult) and so on. We also observe that they frequently occur with the following collocates: *aile göçü* (family migration), *aile birleşimi* (family reunion), *etnik toplum* (ethnic community), *yetişkin göçmenler* (adult migrants) and so on. My native speaker intuition tells me that I am not familiar with these collocations, nor would be an average native speaker in Turkey. Let us see what the Bank of English says in terms of their frequency and collocability with each other in English:

16,760 occurrences of family

408 occurrences of migration

2 occurrences of family+migration

328 occurrences of reunion

25 occurrences of family+reunion

only 5 occurrences in the context of immigration

9730 occurrences of community

1128 occurrences of ethnic

26 occurrences of ethnic+community

adult+migrants no match

What does this picture show us? These are not significant collocations in a huge monolingual corpus. It is obvious that they gain significance in the context of immigration. Hence, we find out more about the nature of language contact data. If we count such words individually as ML vocabulary, this will not tell us anything but it will convince us more that the ML lexemes constitute the majority. Yet, if we count them as ML islands, this will again be misleading because although they are in Turkish, they are new collocations. I have checked these informally with a number of native speakers in Turkey and they have also confirmed that they are unfamiliar with these. Consequently, certain words co-select new lexical environments in Turkish in the contact setting and ordinary words like *aile* (family) end up collocating with *göç* (migration) and *birleşim* (reunification) frequently:

yv533 002Aa:7	ve ayrýcalýklý	aile	birleşimi göçleri Ekonomimiz köt
yv766 001Aa:8	Göçmen Bakanlýðý'nýn	aile	birleşimi programýna göre
yv106 005E:7	Anne ve Babalarýn	Aile	birleşirme adý altýnda gerekli
yor393 005Aa:1	göre döneminde	aile	birleşirmesi kategorisi
yor279 001Aa:8	bu þekilde aýýkladý	Aile	göçü çerçevesinde bin insani
yor279 001Aa:9	büyük indirim oraný	aile	göçü kontenjanýnda yapýlmýþ
yor393 005Aa:1	bunlarýn bininin	aile	göçü nitelikli göç ve yerel
yor393 005Aa:1	Bunlarýn bini tercihli	aile	göçü kapsamýnda kabul edilecek.

As stated above, it is unhelpful to count them as individual ML lexemes and they should be counted as units. Clearly we cannot count them as EL lexemes or units because they are not in English. Yet, they are not established collocations in Turkish, either. If we, however, accept them as the units of the mixed code, this will then allow us to start building our Cmx lexicon from scratch. This point will be elaborated in 9.2.

8.1.4 The question of delexicalised verbs

Traditionally, three verbs in Turkish, *yapmak*, *etmek*, *olmak* have received similar treatment in the grammars and dictionaries of Turkish. While it has been recognised that these verbs hold some sort of special status in the Turkish lexicon, it is impossible to find a satisfactory description in the existing grammars and dictionaries with reference to their delexicalised function particularly with reference to nominalization (Kurtböke 1998a and b).

As emphasized so far in this thesis, corpus research has shown that the commonest meanings of words are not usually those supplied by introspection and consequently by dictionaries. It has also been demonstrated that (Sinclair 1991a:113)

[t]here is a broad general tendency for frequent words, or frequent senses of words, to have less of a clear, and independent meaning than less frequent words or senses. These meanings of frequent words are difficult to identify and explain; and, with the very frequent words, we are reduced to talking about uses rather than meanings. The tendency can be seen as a progressive delexicalization, or reduction of the distinctive contribution made by that word to the meaning. [emphasis added].

Textual evidence shows the extent to which such frequent words without independent meaning operate in the construction of normal texts (Sinclair and Renouf 1988), and it is unhelpful to attempt to analyse them grammatically (Sinclair 1991a:113). Typically, have, take, give, do and make can be listed

among the most delexicalized verbs in English (*Collins Cobuild English Grammar* 1990:147).

The primary function of make, for example, is to carry nouns like 'decision/s' [...] thereby offering the alternative phraseology 'make your own decisions' to 'decide on something' [...] and so on. Which of the two formulations to choose is obviously a strategic matter in text creation, but the delexical option is firmly there (Sinclair and Renouf 1988).

It is worth noting that although bilingual Turkish-English dictionaries give *do* and *make* as equivalents of *yapmak* and *etmek*, also in this area one-to-one correspondence between this pair of languages is difficult to establish. For example, pairs like *to decide* and *make a decision* in English do not always have an equivalent pair in Turkish. Although there is a superficial similarity, the Turkish pair *kararlamak* and *karar vermek* function differently as the independent verb form (*kararlamak*) in Turkish encapsulates *noun+reciprocal suffix+causative suffix* and acquires a different meaning (*to arrange to...which involves the participation of at least two people in the decision*). In the Ozturk Corpus, we observe that not all occurrences of *karar vermek* translate as *decide* but it also means *come to the conclusion that...* as in the concordance lines below. Interestingly, 16 out of 45 occurrences have a *nominalization+dative* suffix form immediately to the left of *karar + vermek* as in the following set of concordance lines:

hangisi olduđuna	karar vermeden	önce kullanýlan bir
.....Katýlmaya	karar verirken	hahsi ilgilerini takip
kalmayı mı istediđinize	karar vermeniz	gerekecektir Eđer
hakký olduđuna	karar verilenler	programdan
için hükümetin atmaya	karar verdiđi	adýmların özeti ađıđda
bir daire kiralamaya	karar vermiđi	Son bepyıldýr kendisi
yolun hangisi olduđuna	karar verir	Sađlýk Hizmetleri Đikayet
sakatlýđý olmadýđına	karar verdi	ve bađvurusunu reddetti

While in English the verb *decide* is simply followed by the infinitive marker *to* as a base for the following verb (e.g. *he decided to leave*), the equivalent Turkish

construction uses a nominalized verb followed by a dative case marker. Likewise, *delexical do+noun* pattern in English behaves differently from Turkish. Corpus evidence shows that the nouns *damage* (154) and *harm* (162) collocate frequently with *do*. As in the case of *decide* and *make a decision*, there are pairs *to damage* and *do damage*, *to harm* and *do harm*. But with their Turkish equivalent *zarar*, an Arabic loan, the choices are limited as it cannot function as a verb on its own. In *The Oxford English-Turkish Dictionary* the equivalent of *to damage* and *to harm* is given as *zarar vermek*. Interestingly, this information helps us discover yet another delexicalised verb in Turkish, *vermek*. This illustrates the fact that standard dictionary and grammar book information on such verbs should be reconsidered; apparently there are others with the same function. This, then, points to the fact that their collocates do not run parallel in languages such as English and Turkish and further investigation is necessary to establish the equivalents.

It should also be mentioned that there are delexical structures restricted with respect to the form of the following noun. That is, there are cases where the noun is mainly used in the plural form as for example in *do+repairs*. Interestingly, Turkish equivalent of the singular form *tamir* (repair) is used with *etmek* but the plural form *tamirat* (repairs) attracts *yapmak* and there is also a difference in meaning. This is again important information about usage unavailable to the user in dictionaries and grammar books. It has been observed that delexicalized verbs are used more frequently in written texts than in spoken ones and Pierce's list below confirms this in relation to the differences between his spoken and written Turkish corpora. His frequency counts were computed on the basis of a similar size corpus to mine. While his list does not specify which forms of these lemmas were more frequent than others, his counts are important as they clearly show the difference in the use of these verbs in Turkish.

Table 7

PIERCE'S FREQUENCY LIST

	Spoken Turkish		Written Turkish	
	Morpheme	Frequency	Morpheme	Frequency
1	demek	8,742	bir	5,589
2	bir	4,763	bu	2,170
3	bu	3,278	olmak	2,053
4	o	3,203	etmek	1,944
5	ben	2,764	ve	1,736
6	ne	2,647	demek	946
7	olmak	2,645	o	856
8	gelmek	2,372	ne	685
9	gitmek	2,372	bağ	651
10	sen	1,882	yapmak	650
11	var	1,801	için	641
12	şey	1,343	ben	607
13	almak	1,281	görmek	569
14	yapmak	1,264	gelmek	559
15	vermek	1,216	iki	558
16	ora-	1,175	vermek	546
17	yok	1,175	gibi	495
18	etmek	1,098	bulmak	486
19	bakmak	1,080	hareket	478
20	kız	1,072	almak	458

The analysis below concentrates only on *yap-* that has recently attracted considerable attention in language contact research where one of the languages in the pair happens to be Turkish (e.g. Backus 1996, Kurtböke 1998a and b, Pfaff 1998, Türker 1998). The approaches differ to a large extent yet there is consensus that it is a characteristic of diaspora Turkish.

8.1.4.1 *yap-*

Standard grammars and dictionaries of Turkish do not have very comprehensive treatments of *yap-* and there is hardly any reference to its delexicalization process discussed in 7.5.2 and 8.1.5 above. The missing information in traditional grammatical and lexicographic descriptions is the frequent nouns or nominal groups that co-occur with these verbs and carry most of the meaning. Such descriptions are usually limited to the mention of Arabic and Persian verbs borrowed into Turkish as nouns through *et-* and *ol-* constructions (Underhill 1976, Kurtböke 1996a).

The frequent co-occurrence of *yap-* with nouns and nominal phrases attract considerable attention in the present day diaspora Turkish. For example Pfaff (1990) observes the incorporation of German infinitives into Turkish through *yap-* constructions. Pfaff lists also some other 'operators' that appear in her data frequently: *etmek*, *olmak*, *atmak*, *vermek*, *almak*, *gömek*. Interestingly, Ozturk corpus shows that all of these verbs have delexical functions. Also Backus (1996) finds that *yap-* occurs frequently with Dutch nouns in his data sets. The problem with these descriptions is that either they suffer from the established categories of grammar and lexicography, as *yap-* has to fit in sometimes as an auxiliary or a main verb; or their focus is elsewhere. For example, Türker (1998) observes that *et-* is being replaced by *yap-* in Norwegian Turkish. A replacement of a slightly different kind is also observed by Silva-Corvalán in relation to *ser* and *estar* in the Spanish speaking community in the US (1994: Chap 4). My view is that we need comparable Turkish corpora of larger sizes from the mainland as well as Turkish migrant communities all over the world to be able to confirm this claim. However, within the language change scenario such diachronic interpretations are commonplace, even on the basis of limited data.

Patterns similar to those with *yap-* appear in other language contact data, too. For example, in 'emigré' Greek the use of *kano* (do, make), and *ginomai* or *yinome*, (the passive form of *kano*) with loan words is very common¹⁴:

¹⁴ I am grateful to Phillip King, EISU, The University of Birmingham for his help with *kano* and *yinome* and providing these examples.

kano ena telefonima

make a phone call

kano tzoging

do jogging, go jogging

Also Tamis (1986:169) confirms this tendency who studied Greek-English language contact in Melbourne. He observes that:

many morphosemantic transferences of verbs occur in the corpus. They are almost invariably left uninflected. The common pattern involves the use of the modern Greek verbs 'kano' (to do) for the active voice and 'ginome' (to become) for the passive auxiliaries, perhaps due to their wide semantic range (to do=to make, to construct, to create, to get, etc.) or because of their productive function in Modern Greek as auxiliaries with nouns.

Also Romaine (1989) examines a number of similar verbs in Panjabi, which she typically refers to as compound verb constructions. In her description, these constructions are formed by a a verb, noun or adjective and an 'operator':

The operators comprise a small class of simplex verbs with lexical meaning in their own right. The main ones are k rna 'to do' and hona - 'to be/become'. The basic meaning of the compound is determined by the first element and modified by the verbal operator" (1995:131).

Before a further analysis of *yap-* on the basis of the evidence from the Ozturk corpus, let us first look at the nature of the examples so far reported on the basis of Turkish data sets in and around Europe. Generally speaking, we observe the following characteristics:

- a) the data used is spoken;
- b) there seem to be two uses of *yap-* in the environment of L2 lexemes:
 - with L2 nouns
 - with uninflected L2 verbs;

c) in both cases, the number of the forms used is limited.

I will deal with each of these characteristics individually. Firstly, it is important to consider the 'spokenness' of the examples. When Pierce compared written and spoken Turkish texts, he concluded that "the number of suffixes per word [was] smaller for spoken Turkish than for the written language" (1963a and b). What is the significance of this observation with respect to the use of *yap-* in language contact settings? If Pierce's observation is correct, and there is reason to believe it is, the number of suffixes attached to the borrowed words under investigation will be limited.

Secondly, the environment of *yap-* will have an L2 noun, (see for example the Dutch nouns reported by Backus 1996:241-242), and when it does, the nominal group will carry most of the meaning, hence the delexical form explained in 7.5.2 and 8.1.5. Alternatively, its environment may contain an uninflected verb and this form needs some explanation. Let us approach this from a different angle. Turkish speakers in Germany¹⁵, for example, frequently produce sentences such as

- (1) *Banhofa gittim*
 I went to the train station
 (German *Bahnhof* 'train station' + Turk. dative -a)
- (2) *Mannim işsiz*
 my husband is unemployed
 German *Mann* 'husband' + Turk. possessive -im)

In contrast they appear to be much more hesitant about combining a borrowed stem with Turkish verbal morphology. For example, sentences like

**vize beantrag-dim*

I applied for a visa

visa + German verbal lexeme *beantrag-* + Tk past tense marker -di + 1st person marker -m.

¹⁵ I am grateful to Dr Friederike Braun for a useful discussion on this point and the examples from her data.

do not appear in German-Turkish data but forms such as

vize beantrag yaptým

I applied for a visa

visa + German verbal lexeme beantrag- + Tk delexical verb yap- + Tk past tense marker -di+ 1st person marker -m.

seem frequent and sound natural to the native speakers of Turkish in that setting. Tamis's (1986:169) examples confirm the same tendency with kano and yinome in Greek, that is in the environment of an uninflected L2 verb.

kano endzoi (enjoy) ti suljamu

(I do enjoy my work)

seligo **ginome ritair** (retire)

(in a while I will be retired)

aftos omos **deginete e'fekt** (affect)

(he is not affected)

There is an example of this kind in the Ozturk corpus, as well. In text yv806 010Ba, a reported interview, we read:

"Yeri **lease yaptık**, yapýnyna başladık. Ne isim koyalým diye düşünüyorduk "

We leased the place, started doing it up. We were thinking about what to name it"

What does a delexical verb enable the speaker to do in such cases? To create a Turkish verb phrase, complete with tense, mood, aspect, person and so on from almost any borrowed verb stem. Tentatively, we can put forward the claim that rather than using an inflected stem, the speaker will opt out for a bare stem followed by an inflected form of *yap-*. This strategy will be particularly popular in spoken Turkish where fewer suffixes operate (Pierce 1963a). In fact, as the texts that make up the Ozturk corpus are written, we do not see any examples of this pattern except for the only example above, *lease yaptık*. It is not surprising that it comes from a text of direct speech.

The third point concerns the forms of *yap-* used in these constructions. The examples so far reported give us the following forms:

German-Turkish

tauschen yapmam, wechseln yapmadým, vermischen yap[a]cak, saubermachen yapýyo[r] (**source: Pfaff 1990:137-140**).

German-Turkish

malen yapýyorlar, kaputt yapýyorlar, vordrängen yapýyorlar, gratulieren yapýyo[r], Brief schicken yapýyo[ru]m, reparieren yapýyor, schreiben yapýyor; tanzen yaptýlar, wohnen yaptý, verlaufen yaptýlar, einkaufen yaptýlar, aufmachen yaptý; Weihnachtsmannbaum yapmýþlar, Hex yapmýþ Pause yapmýþlar; gesund yapalým, Haende hoch yapsýn; puzzleri yapar; Rolle yapabilir (**source: Pfaff 1998**).

Dutch-Turkish

trimmen yapacađým, wegnemen yapacaklar; trimmen yapýlýr; vragen yaptýydým, verzinnen yaptým, schatten yaptým; beheersen yapýyorken; opruimen yaparken, overschrijden yapýnca; (**source: Backus 1996:277-278**).

Norwegian-Turkish

Kyllingsalad yapmýþtým, diskusjon yapmýþtýk; Diskutere yapýyoruz, suge yapýyorlar; avslutte yapacađýz (**source: Türker 1998**).

What is striking in the way such units are formalized by these researchers is that the representation typically takes the form of a dictionary entry like

/trimmen yap-/

where suffixation is dropped and it looks legitimate to do this, although there is no explanation as to why it is done. Accordingly, some would claim that, all the forms of *yap-* can, in theory, be used in such constructions. However, some three hundred plus forms do not occur or have not occurred yet in the available data collections. This observation further supports the claim that borrowed stems seem to combine more readily with nominal morphology than with verbal morphology, and the function of *yap-*, specifically in combination with a borrowed verb stem

appears to make it inflectable. This may be taking place more rapidly in spoken Turkish where suffixes are fewer and delexicalized verbs are more frequent. According to Halliday frequent items such as delexical verbs (*general verbs*, as he calls them) contribute little to the lexical density of texts and should therefore be considered grammatical words rather than content words (Halliday 1985:64-65), hence, the case for the grammaticalization of *yap-*.

Rather than subscribing to the existing descriptions of *yap-*, and attempting another description along the lines of established categories such as *auxiliary*, *compound verb*, *operator* and so on, I want to explore its use in relation to the lexicogrammatical concept of *clustering* first proposed by Firth and developed further by Sinclair and most notably in the work of Hays (1997), a student of Sinclair's, into computational lexical scatter software. Hays (1997:173) clusters various forms of a lemma into lexical scatter sets, by paying attention to how their meaning dictates their form. For example, he looks at the lemma *write* and finds that the lexical scatter set for the meaning "[saying something]" comprises of the following forms (1997:171-172):

It is in fact, as Campbell	writes,	a variant...
gone to welcome Ms Bhutto	writes	Ahmed Rashid.
finds himself an IRA target,	writes	Alex Renton.
"I am afraid", Monet	wrote,	"that you won't..."
Of alcohol, Fleming	wrote,	"Drink relaxed Bond."
nearly turned me out" she	wrote,	for writing esoteric

However, when he looks at the form *writing* in the Bank of English, he finds that it is not used in this construction as a verb but commonly used with the verbs *says* and *said*

Lionel Barber,	writing	from Washington, says...
Derek Brown	writing	in the Guardian says...
John Major,	writing	in the forward, said...
Mr Morris,	writing	in Marketing Magazine, said

This restricted usage of *writing* leads him to formulate the lexical scatter of *write* with the particular meaning of 'saying something' as follows:

/write/={ write, writes, wrote, written}.

This set obviously does not include the form *writing* as its meaning indicated by this syntactic form belongs elsewhere. Important criteria for lexical scatter membership, then, are (1997:173):

1. The types must be systematically related by the morphology. These are the various forms of a word, and the lexical scatter set reflects this.
2. Members of a lexical scatter set must be members of the same meaning unit, and the criterion includes usage, as well. Because of the co-selection of form and grammar, there must be a limitation of forms for various senses [...] If the types are consistently used in different collocation environments, then they should not be conflated into a single lexical scatter set.

All together there are 320 different forms of the lemma *yap-* in the Ozturk Corpus with frequencies ranging between 226 to 2. One occurrence is not considered here as "a language pattern - however defined - has to occur a minimum of twice" (Sinclair 1996b:81) but a complete list of all the forms including those occurring once can be found in the appendix . Out of 320 different forms of *yap-* 170 forms appear at least twice. The overwhelming number of these forms begs some questions. These questions are pertinent to language contact research as well as to broader issues of Turkish grammar:

1. Do all of these forms have the same meaning and the same function?
2. Can they be used interchangeably, or are there differences with respect to their environment?
3. If these forms show differences, can they be grouped according to their meaning and function?
4. Do they all show equal degrees of delexicalization (King, Kurtböke and Verde, *in progress*)?

If we follow the clustering technique elaborated by Hays (1997), the word to be represented by the lexical scatter sets would be /yapx/ with x referring to a particular meaning unit. There are two starting points for this analysis. The first one is the dictionary where the available meanings of yap- are the following (p501):

yapmak Do; make; create; give rise to; build, construct; constitute; arrange; repair; apply; set to rights; make ready

It is not clear from the dictionary entry whether these meanings are realized by *yap-* standing on its own or by its environment. However, the preceding discussion on delexicalization has shown that certain nouns or noun groups carry more of the semantic load than *yap-* itself. If we look at the cases where *yap-* is not in any significant relationship with a noun around it, we will see that these are limited to an anaphoric function and/or when it is preceded by a question word:

.....bunlarýn **nasýl yapýlacađý**...(how these will be done).

.....çocuđunuzun okulda **ne yaptýđý**...(what your child is doing at school)

.....**neler yapýldýđýný** merak ediyorsanız....(if you want to know what is being done)

.....telefon bozulursa **ne yaparým**.... (what would I do if the phone breaks down)

This indicates that the lexical sets will have to be re-established with respect to the nouns and noun groups in the environment of *yap-*, as *yap-* is not capable of realizing these meanings on its own. If we take the meanings listed in the dictionary, we would end up with the following representations although the dictionary does not specify how and when these meanings occur:

/yap1/= {do}

/yap2/= {make}

/yap3/= {create}

/yap4/= {give rise to}

/yap5/= {build, construct}

/yap6/= {constitute}

/yap7/= {arrange}

/yap8/= {repair}

/yap9/= {apply}

/yap10/= {set to rights}

/yap11/= {make ready}

The second starting point for the analysis is Hays (1997:174) who states that "the various mental categories of meaning function as both nouns and verbs" and analyses his lemmas accordingly. Here, when we leave out all the other *yap*-forms of nominal, adjectival and adverbial nature, we end up with some eighty verb forms although very few of these appear in the language contact data of European researchers:

yapýlacak - 101, yapýlýr - 94, yapýldý - 30, yaptý - 30, yapýlacaktýr - 27, yapar - 25, yapýlmaktadýr - 22, yaparlar - 20, yapacak - 20, yapýyor - 19, yapýlabilir- 19, yapýlmýþ - 15, yapmaktadýr - 12, yapýlýyor - 10, yapýlmýþtir - 10, yaptýk - 10, yapabilir - 9, yapacaktýr - 8, yapýnýz - 7, yaptým - 7, yapmýþtir - 6, yapýyoruz - 5, yaptýlar - 5, yapmalýyým - 5, yapmaktýr - 5, yapabilirsiniz - 5, yapabilirler - 5, yapabiliriz - 5, yapabilecek - 5, yapýlmaz - 4, yapýlmalýdýr - 4, yapýlmalý - 4, yapýyorum - 3, yapýyorlar - 3, yapýn - 3, yapýlýyorsan - 3, yapýlýrsa - 3, yapýlsýn - 3, yapýlmak - 3, yapýlmadý - 3, yapýlarak - 3, yapýlabilecek - 3, yaptýrýnýz - 3, yaptýrýlmamalýdýr - 3, yapmaz - 3, yapmalýyýz - 3, yapmalýdýrlar - 3, yapmalýdýr - 3, yapmalý - 3, yapmaktadýrlar - 3, yapamaz - 3, yaparak - 3, yapýlmýyor - 2, yapýlmýþsa - 2, yapýlmayacak - 2, yapýlmamýþtir - 2, yapýlamaz - 2, yaptýrýlmamalýdýr - 2, yaptýrdýlar - 2, yaptýrabilirsiniz - 2, yapsýn - 2, yapsalar - 2, yapmýþti - 2, yapmýþlar - 2, yapmýyor - 2, yapmazsanýz - 2, yapmazsa - 2, yapmam - 2, yapmalýsýnýz - 2, yapmaktayýz - 2, yapmadýðýmýz - 2, yaparsýnýz - 2, yaparsak - 2, yaparlarsa - 2, yapamýyoruz - 2, yapamayacak - 2, yapmamýþtir - 2, yapacaðýz - 2, yapacaðým - 2, yapacaksýnýz - 2, yapacaksanýz - 2, yapacaklar - 2, yapabilirim - 2,

However, there are other forms of *yap*- in the corpus counting as many as ninety and they have nominal, adjectival and adverbial functions. They are also restricted with respect to their environments. Except for three adverbial forms of *yap*- in the European data (*yapýyorken* 'while doing', *yaparken* 'while doing', *yapýnca* 'after having done') we do not see any other non-verb uses as such. The dictionary, likewise, recognizes few non-verb forms, with very unhelpful translations:

yapýlan - 226, yaptýðý - 113, yapýlacak - 101, yapan - 83, yapmak - 73, yapýlmasý - 44, yapmýþ - 41, yaptýklarý - 28, yaparak - 26, yapma - 23, yapacak - 20, yaparken - 19, yapmayý - 18, yapmaya

- 18, yapýldýđýný - 16, yaptýđýnýz - 16, yapmasý - 16, yapýlmýþ- 15, yapýp - 13, yapýldýđý - 13, yapmanýz - 13, yapacađý - 13, yapýlmasýný - 12, yapýlacađý - 12, yapacaklarý - 11, yaptýđýný - 10, yapacađýnýz - 10, yapmasýný - 9, yapmalarý - 9, yaptýđýmýz - 9, yapacađýný - 9, yapabilmek - 9, yapmadan - 8, yapýlmasýna - 7, yaptýrmak - 7, yapýlacađýný - 7, yapmanýn - 6, yapanlarýn - 5, yapanlara - 5, yapýlmadan - 5, yaptýklarýný - 5, yapmalarýný - 5, yapabilecek - 5, yapýlýp - 4, yapmamak - 4, yapmaktan - 4, yapmakta - 4, yapacaklarýný - 4, yapabileceđi - 4, yapýlmadyđý - 3, yapýldýđýndan - 3, yaptýđýndan - 3, yaptýđýnda - 3, yaptýđým - 3, yapýlabeledek - 3, yaptýrma - 3, yaptýklarýnýn - 3, yaptýklarýna - 3, yapmakla - 3, yapanlarý - 3, yapamadýđýný - 3, yapýlmaya - 2, yapýlmasýnýn - 2, yapýlmasýnda - 2, yapýlmamasýný - 2, yapýlmaka - 2, yapýldýđýna - 2, yapýlabilmesi - 2, yapýlabilen - 2, yaptýđýnýzý - 2, yaptýđýmý - 2, yaptýrmanýzý - 2, yaptýrdýđýnýz - 2, yaptýklarýyla - 2, yaptýklarýnda - 2, yaptýklarýmýz - 2, yapmasýna - 2, yapmanýzý - 2, yapmanýzda - 2, yapmamalarý - 2, yapmamamktan - 2, yapmalarýna - 2, yapacađýmýz - 2, yapacađým - 2, yapacaklarýnda - 2, yapacaklarýna - 2, yapacađý - 2, yapabilmenin - 2, yapabilen - 2 yapabileceđini - 2, yapabileceklerini - 2, yapa - 2.

As stated above, all these forms may seem probable in all the environments but there are preferences, evident in the Ozturk Corpus at least. For example, if we look at the frequently-occurring forms *yapmanýz*, *yapmasý* and *yapýlmasý*, we can see that these forms are restricted to the the environment with *gerek...* where the unit functions in a similar way to the French construction *il faut*:

yor393 008X:1 tembel olur	yapýlmasý gereken	iþin önemine hiç aldýrýþ
yor393 008X:1 acaba Evet	yapýlmasý gereken	bir þeyi þu ya da bu
yv111 001H:12 birisinin	yapmasý gerekir	niçin Melbourne'e
yor 50 008X:1 ya zekidir	yapmasý gerekli	olan iþini boynuna borç
yv235 003Aa:6 bir düþþ	yapýlmasý gerekmektedir	
yv523 009x:17 soruþturma	yapmasý gerekmektedir	
yv786 004x:14 için baþwuru	yapmalarý gerekmektedir	
RIG 001:6 konularda arlamento'nun	yapmasý gereken	yasalar veya
WOM 003:1 birkaç test daha	yapmasý gerekebilir	
WOM 004:3 görmüþbir doktorun	yapmasý gerekir	
yor259 006Aa:5 aracylýđýyla	yapmasý gerekecek	
yor393 008X:1 ya bir insan ya zekidir	yapmasý gerekli olan iþini boynuna	
yv111 001H:12 iþi ydney'den birisinin	yapmasý gerekir niçin Melbourne'e	
yv523 009x:17 üzerinde soruþturma	yapmasý gerekmektedir	
COM 004:8 seçerseniz hiç birþey	yapmanýz gerekmez	
COM 004:10 verirsiniz hiç birþey	yapmanýz gerekmemektedir	
DOM 002:11 sonraki aþamada neler	yapmanýz gerektiði bildirilecektir	
DOM 002:15 kendi düzenlemelerinizi	yapmanýz gerekecektir	

DOM 002:18 daha sonra size neler	yapmanýz gerekeceđini anlatacaktýr
DOM 004:1 jiddet karþýsýnda ne	yapmanýz gerektiđini öđrenmek
twg 036X:1 arasýnda bir tercih	yapmanýz gerektiđinde nasýl
WOM 001:1 nasýl bir eđitim	yapmanýz gerektiđini
WOR 005:24 klinik çalyþma	yapmanýz gerekecektir
WOR 006:3 yaptýrabilmek için	yapmanýz gerekenler bu broþürde
WOR 006:13 denetim altýnda pratik	yapmanýz gerekir koþullarýn ayrýntýlarý
yor398 001fa:1 bu bilgileri almak için	yapmanýz gereken tek þey bir vergi

These three forms then should be clustered together around the meaning 'obligation' into a separate lexical scatter set. Similarly, there are 94 occurrences of the form *yapýlacak*, which can easily escape attention as the future form of the verb, but only in 13 occurrences *yap-* actually has the function of a verb as in:

...bu hafta Cuma günü yapýlacak
(will be done on Friday this week)

However, in the remaining 81 occurrences, we see that *yapýlacak* functions as the qualifier of the following noun as in *yapýlacak deđiþklik* 'changes to be made', *yapýlacak þeyler* 'things to be done', *yapýlacak iþ* 'job to be done' and so on.

My last example is the past form of the verb *yap-*:

gün090 001Ab:1	kitabýn tanýtýmýný	yaptý	Ayrýca Etnik basýn ve yayýn
tr148 009Ba:1	kapanyþkonuþmasýný	yaptý	ve bu konuþma Çin televizyonunda
tr149 003Ba:1	festivalin açýlýþýný	yaptý	Bu arada kafeslerde tutulan çok
tr149 007Ba:1	Ýngilizce öđretmenliđi	yaptý	Avustralya'ya 1979'da geldiđinde
tr149 005X:1	ve iþveren arasýnda	yaptý	Ýþverenlere bugüne kadar
tr150 003Ba:7	görmüyorum" yorumunu	yaptý	\tx Festival Komitesi'nin mali
ts275 002Ba:3	bir sohbet toplantýsý	yaptý	\tx Toplantýda ve EA Melbourne
twg 042X:1	yapamadýđýný	yaptý	Milli sporcularýn baþarılarına
twg 054X:1	bulunduđu dekorasyonlar	yaptý	ufacyk çocuklar yazýn ortasýnda
yor289 007Ac:4	destekleyici çýkýþlar	yaptý	Ancak sonraki yýllarda Ýþi
yor393 010E:1	panelin yöneticiliđini	yaptý	Avustralya Türkiyeliler Derneđi
yv086 003Ba:2	alarak bir konuþma	yaptý	Sayýn Arýyak'ýn konuþmas ýndan
yv086 004Bc:2	onsekiz içinde atýþý	yaptý	ve top Faruk a çarpýp kaleyi
yv086 004Bc:2	yerden vuruþla skoru	yaptý	Rakip takým ellerine geçen

yv091 003Bb:4	Arýyak bu iþte önderlik	yaptý	\tx Günün özel misafirleri olan
yv096 001Aa:3	ile özel bir görüþme	yaptý	Yeni ofisin açýlýþý nedeni ile
yv255 005x:1	birgün bir hata	yaptý	yaþý bir hanýmýn bacadýný
yv503 002Ba:4	Toplumunda bomba tesiri	yaptý	
yv513 001Ad:2	tarihinde Canberra'da	yaptý	\tx Türkiye devlet bakaný Hüsnü
yv543 004Ba:3	üzerinde þok etkisi	yaptý	\tx Kürsüye çýkan NSW Türk
yv543 005Ba:5	dedi ve þu açýklamalarý	yaptý	\tx "Bayram Programlarýmýz için
yv756 017Bc:19	kazandý Vuruþu nolu	yaptý	Kaleci Mustafa'nýn bakýþlarý
yv756 019Bc:8	gösterdi Vuruþu Kemal	yaptý	gol \tx Deplasmanda ALBION RED
yv766 008Bb:3	ile bir toplantý	yaptý	Modern gazlý kazanlarda
yv776 001Ba:5	konuþmasýný Nihat Atalý	yaptý	Büyükelçimiz ve Baþkonsolosluk
yv776 005Ba:4	davet edildi konuþmalar	yaptý	Türkiye'de Beþyýllýk Kalkýnma
yv776 007x:1	Rumlarla iþbirliði	yaptý	Nisan 1955'te EOKA örgütünü
yv786 002x:11	sýcak bir anlamda yer	yaptý	Böylece Ýsveç kooperatifleri
yv806 009Bc:16	kazandý Vuruþu Cemal	yaptý	top nolu oyuncunun kafasýna
yv806 009Bc:22	gösterdi Atýþý Kemal	yaptý	gol \tx dak nolu oyuncunun sert

This particular form *yaptý* has a total of 30 occurrences as opposed to 113 occurrences of the participle *yaptýý*, the adjectival use where the following noun is typically one of these: *konuþma* 'speech', *araþýma* 'investigation', *açýklama* 'announcement', *gösteri* 'demonstration', *yardým* 'financial aid', *hizmet* 'service' or *toplantý* 'meeting'. While among the 30 lines of *yaptý* we do encounter some of these nouns, there are not many. In other words, these two different forms attract different environments and this tendency should be indicated in formal descriptions of grammatical and lexicographical kind. If we link this argument to the European data, one point becomes obvious. While this might be a feature of spoken data, we do not come across written forms such as *yaptýým Kyllingsalad* or *yapacaðým trimmen*. Neither in the Ozturk corpus have we had *yapýlan delivery* or *yaptýýmý lease*, although these combinations are probable. Yet, frequency counts tell us otherwise. That is, collocational and colligational tendencies of different forms do not match and these cannot be collated into one single meaning or into one formalization like /yap-/ as is usually done in grammars, dictionaries and language contact analyses. Hence, the eleven meanings of *yap-* extracted from the dictionary above need to be reconsidered. To sum up what has been said so far, when we consider the overwhelming number of non-verb forms of *yap-* with respect to their environment and function we find

differences and they are not always used interchangeably. Different forms and meanings cannot be clustered without reference to the noun or nominal group located around *yap-*. While delexicalization is a characteristic of *yap-* its non-delexical use still shows up within the environment of question words such as *ne* 'what', *nerede* 'where', *nasýl* 'how' and so on. In the context of language contact, however, the major function of delexical *yap-* seems to be to inflect the uninflected borrowed stems.

8.1.5 The question of functors

A commonly recognised distinction in the study of language is between lexical items (*content words*) and grammatical items (*function words*). Lexical words are

ITEMS (i.e. constituents of variable length) rather than words in the usual sense, because they may consist of more than one word: for example, *stand up*, *take over*, *call off*, and other phrasal verbs all function as single lexical items. They are LEXICAL because they function in lexical sets not grammatical systems: that is to say, they enter into open not closed contrasts" (Halliday 1994a:63).

Grammatical items, in contrast, enter into a closed system such as the personal pronouns. Although many items in a language belong to either lexical or grammatical class, "there are always likely to be intermediate cases. In English, prepositions and certain classes of adverb (for example MODAL adverbs like *always*, *perhaps*) are on this borderline" (Halliday 1985:63; see also Winter 1977 about Vocabulary 3). Lexical items of high frequency, "often general terms for large classes of phenomena" such as "*thing*, *people*, *way*, *do*, *make*, *get*, *have*, *go*, *good*, *many*" are

on the borderline of grammar; they often perform functions that are really grammatical - for example thing as a general noun (almost a pronoun) as in *that's a thing I could do well without*, make as a general verb; as in *you make me tired*, it makes no difference. They therefore contribute very little to the lexical density. (Halliday 1994a:64-65).

While "grammatical items tend to be considerably more frequent in occurrence" and a list of the most frequent words "in the English language will always be headed by grammatical items like *the* and *and* and *it*" (Halliday 1985:64); studies of co-occurrence show that even grammatical items should be considered with respect to their environment and their description changes on the basis of corpus evidence (e.g. Sinclair 1991a, Chap 6).

The treatment of function words in language contact data has been problematic as it has hardly been established whether they should be considered as lexical or grammatical interference. As Hasselmo (1969:129) pointed out "generally, the borrowing of function words has been treated under the rubric of "lexical interference" however, there were attempts to assign them to "structural interference" (Rayfield 1970), as well. Rayfield (1970:64), in a study of Yiddish-English contact, found that "important differences" were "brought out by treating lexical and function words separately":

1. The number of lexical items borrowed in either direction is large, but small in proportion to the total number of lexical items in either language, while the number of function words borrowed in either direction is small, but large in proportion to the number of function words in either language.

2. The dominant direction of borrowing of lexical words is clearly from English to Yiddish. Function words are borrowed to about the same degree in either direction.

In spite of examples of this type, it has been assumed for a long time that functors did not travel freely across languages. For example, Joshi (1984) argued that closed class items cannot be switched, whereas open class items can. However, counter-examples to such arguments can easily be found:

...j'aime pas ça because moi j'en connais trop des Anglais...

(Heller 1990:67)

and an examination of concordance lines from the Bank of English will reveal that there are different uses of *because* and it may be that it is switched more easily when it is followed by a personal pronoun rather than by the preposition of:

1)

indy/03 said: `The BBC didn't get it just	because	they are jolly good chaps. British lawn
oznews/16 ate of transforming my own party	because	I was complacent or smug, or thought
brspok/07 got to the age of twenty-six who	because	they hadn't made a million felt as if
brspok/07 promoted because you're good or	because	you're a woman? <F02> I've not been Y
bbc/01 she's been past over for promotion	because	she's a woman. Greg Clarke reports:
times/02 usands of locals receive benefits	because	they spent years in Britain and paid
today/06 victims keep quiet about attacks	because	they cannot face a second ordeal - the

2)

oznews/16 t was stuck with that factor but	because	of product quality. Queensland sugar
indy/03 er a defending lawyer fell ill and	because	of a crowded court agenda. None of the
oznews/16 ad `skyrocketed" in recent years	because	of what the club suspected was criminal

It would of course take larger corpora created in immigrant settings and more detailed analysis to be able to confirm such claims. However, recent corpus research has shown that grammatical words such as *and*, *the*, *of* and so on, are better understood by observing their usage in typical patterns of language than their semantic content as they have little independent meaning. Since they commonly function as components of larger units such as phrases, it is pointless to claim that they do not constitute a major part of language contact data contrasted with for example nouns. Again corpus research shows that in English, the major part of a text is formed by the frequent co-occurrence of grammatical words. This brings us to consider the non-equivalence of 'grammatical' words in Turkish and English. To be able to do this, let us consider where the functors rank among the most frequent 20 words in Pierce's spoken data of some 35 years ago as well as the 20 most frequent words in the Ozturk Corpus:

Table 8

order	Pierce's counts	gloss	Ozturk Corpus	gloss
1	şey	thing	ve	and
3	padı̇ah	king	bir	a/one
2	baba	papa	bu	this
4	lira	monetary	için	for
5	tabii	obviously	da	at/in
6	zaman	time	de	at/in
7	adam	man	ile	with
8	ađa	gentleman	daha	more
9	vallah	good lord	olarak	delexical ol- form
10	ama	but	olan	delexical ol- form
11	sabah	morning	veya	or
12	Allah	God	türk	Turk/Turkish
13	filan	and so on	çok	much/many/very
14	sefer	trip	her	every
15	yani	that is to say	gibi	as
16	taraf	side, direction	iş	work/job
17	sene	year	en	the most
18	falan	and so on	kadar	as.....as
19	dünya	world	ne	what
20	kuruş	penny	Avustralya	Australia

(Pierce 1965)

It is of course important to stress the difference between spoken and written data. Pierce's list does not seem to have many grammatical words as such, compared to the Ozturk corpus where 17 top items are functors highlighted in bold. Pierce's functors are less significant in the Ozturk corpus because they are spoken forms:

tabii 116

ama 188

filan 1

yani 87

falan 23

However, the frequency of *-da* and *-de*, requires attention as a point of non-equivalence. If we look for functors in Turkish standing on their own we will find fewer in comparison with English, but a considerable number of functors will be realized through suffixation, and the locative *-da* is one of them. The study case in 8.1.7 elaborates on this issue. The function word investigated in the next section is a self-standing functor as such. However, it is an interesting example of strong co-dependence and positional preference.

8.1.5.1 Eđer

The case study here is important in terms of how Turkish handles language contact and its implications for the setting under investigation. Thomason and Kaufman (1988:217) list a number of borrowed functors in Asia Minor Greek that was heavily pressured by Turkish. Among these there is also *eđer* (if), a Persian borrowing into Turkish itself. My aim in choosing this functor for analysis is to argue against the assumptions that functors are free from structural considerations. As the other lexical choices we make, they also show preferences of collocational and colligational kind. This point is beautifully illustrated in Sinclair's analysis of the most frequent functor in English, *of* (1991a: Chapter 6). First let us consider some of the occurrences of *eđer* in the Ozturk corpus:

CHI 001:8	Eđer	evinizde merdiven varsa oralara
CHI 001:12	Eđer	kayar veya sýkýýrsa bebek
DOM 001:16	Eđer	Homeswest kiracısý iseniz ve
DOM 002:11	Eđer	hüküm sona erdiyse ne yapmalýyým
WOM 001:1	Eđer	okuldan ayrılmayı düþünüyorsanız

We have previously defined colligation as grammatical and positional preferences of a word. What is at work here is colligation as this functor is always present in the environment of a conditional construction that operates through the verbal suffix *-(i)se* in Turkish. Very frequently, 151 times out of a total of 177 occurrences it is found in the sentence-initial position. In the remaining 26 cases,

this word is found in non-sentence-initial position although it has a strong tendency to stay close to the beginning of the sentence as in:

tr147 002Ab:1	Avustralya	eđer	cumhuriyet olursa Hükümet
yor279 001Aa:14	Bu süre sonunda	eđer	Göç programý çerçevesinde
yor403 013E:1	Derneklerin görevi	eđer	gerçekten o insan ruhsal bunalýma
yv533 001Ba:12	Bunu yaparken	eđer	iki ayrý toplum arasýnda dengeyi
yv766 012x:12	bunun da	eđer	istersek kolayca önlenebileceđini
WOM 001:17	Normal durumlarda	eđer	işlen çýkarýlmýşsanýz size yine

For the use of *eđer* we have two conditions, then. The co-presence of the conditional *ise* or its suffix form *-se* or *-sa* attached to the verb, and a sentence initial or near sentence-initial position.

In the light of this observation, let us reconsider Thomason and Kaufman's (1988:217) list of individually borrowed functors in Asia Minor Greek. The question here is: can we assume that *eđer* was borrowed without any impact on its environment? Thomason and Kaufman report that Turkish imposed changes on Asia Minor Greek also at morphological and syntactic levels, and they mention some characteristics along these lines. However, there is neither reference to the co-dependence of *eđer* and *ise* as such, nor is there any reference to the preferred position of Turkish *eđer* in Greek sentences. Yet, it is difficult to think that a grammatical word carrying such constraints along with it can be borrowed haphazardly just like another lexeme.

There is another point worth considering with respect to the frequency of *eđer* in the Ozturk corpus. In most cases in Turkish, the sheer presence of *ise* is enough to realize the conditional, which makes *eđer* redundant. This is supported by corpus evidence in that there are 338 occurrences of *ise*, without counting its *-se/sa* forms, as opposed to 177 occurrences of *eđer*. In Chapter 7, it has been explained that some of the texts that make up the corpus are translations from English. If we look at the texts in which *eđer* occurs, we can see that 101 of these come from translated brochures or government advertisements. This is related to the 'engrossing effect' discussed by Baker (1992) and summarised in chapter 5, as a common feature of translations.

8.1.6 The question of "nonce"

The observation of a single occurrence in a corpus has always been problematic for language contact research and led to heated debate as to whether it should be ignored or not. Researchers have concentrated, for the last two decades at least, on such low frequency material that forms the category of 'nonce loans'. Poplack (e.g. 1987) has discussed whether they should be classified as borrowings or codeswitches in support of the 'free-morpheme' constraint proposed by herself.

My objection to such an argument is that it is based entirely on intuitive judgement, which proves of very little help in linguistic description, as recent computational corpus research has powerfully demonstrated (e.g. Sinclair 1991a). In computational corpus research, single occurrences are seen as one of many more that might have occurred (since the corpus is merely a miniscule sample of the totality of language in use), but have not in the corpus available to the researcher. As Clear (1993) explains:

Unfortunately, we cannot determine from the corpus alone, though our linguistic intuition may lead us to make judgements, whether a single occurrence is indeed an isolated chance phenomenon or one instance from many which might have appeared in a different or larger corpus.

The main reasons for ignoring Poplack's points are twofold. Firstly, the principles of text selection can never be truly objective, "and the evidence from the corpus will be affected by" the policy of its compilers (Sinclair 1991b). (Here, text is understood as any naturally occurring piece of written and spoken language). A corpus built with language contact research in mind will similarly be affected by the compiler's direction of research. Secondly, no matter how large the corpus is, "[t]here are always going to be a large number of patterns for which there is still not enough evidence" (Sinclair 1991b).

Corpus research has shown over and over again that any corpus "is a good and valid sample of some aspects of language patterning, but not all" (Sinclair 1991b) and any corpus will be constrained by the the language of the texts that form it.

Even in a corpus of hundreds of millions of words there will be nonce occurrences and "the only answer is to collect more examples until the initial pattern is reinforced or replaced by the additional examples" (Sinclair 1991b). In other words, Poplack's nonce loans are restricted, as ours, by the size of her corpus, the language of her informants, and her policy of research. It is common practice to establish a threshold value for the frequency of occurrence and co-occurrence that discards the material observed fewer than a set number of times. It is usually taken to be 3+ (Clear 1993). However, this practice is re-examined here in relation to language contact data. Frequency of occurrence has recently been used as a criterion by Myers-Scotton (1993a) to distinguish between borrowed forms and switches. However, one should remember that the frequency of occurrence of a lexical item may be very high at the moment of coinage or at a particular moment of time. Its use may eventually decrease and the item may become less frequent over time. I will illustrate this point with two examples of different types taken from the Bank of English corpora, the first of 'cultural' nature and the second 'core'. In the Oznews corpus, a part of the Bank of English, a rather familiar collocation is the *Demidenko affair*:

<p>role in the Holocaust. <p> egg on their faces because of literary culture was, until Betty Birskys' explanation of else is cashing in on the chapter will be written soon on Wales. <p> FOR better or worse, them in my case as it did with the cosmopolitan. <p> With plans to publish books on be compiling an anthology on lately and not just about to the entire nation in</p>	<p>The Demidenko affair, combined with the the Demidenko affair, but the writer herself is the Demidenko affair, in the main almost the Demidenko affair # The # Mail Perspective Darville/ Demidenko affair. <p> HarperCollins is the Demidenko/Darville affair. <p> To date, the the Demidenko affair has put Australian the Demidenko affair. <p> When the book was the Demidenko affair, they went in the opposite the Demidenko affair. <p> The Allen and Unwin the Demidenko Affair. <p> TARGETT T the Demidenko/Darville affair. <p> Queensland Darville/ Demidenko affair how adrift they are in a the Helen Darville-Demidenko affair is to be revived with the</p>
--	---

Helen Darville who won a literary prize in 1995, and was the subject of a scandal and heated debate for most of that year. As the jury later found out, she had

written her prize-winning book under a false name (*Demidenko*), claiming equally false Ukrainian ancestry. This issue was then broadly interpreted and served as a basis for questioning the criteria in granting literary prizes, hence *the Demidenko affair*. Obviously, while the debate continued, this newly-coined collocation, as well as a number of others (e.g. *the Demidenko debate*, *the Demidenko diary*, *the Demidenko saga* and so on), were very frequently used by the Australian media and public. The frequency of occurrence of this collocation at present is nil. Another point to be made here is the restricted use of this collocation in the English-speaking world as it was pertinent to Australian English only and did not make its way into the other varieties of English.

Likewise, an example of a recently-coined 'core' lexical item in the Bank of English is *spin doctor* or lately *spinner* that originated in American English and is now widely used in the UK and Australia as can be seen in the Bank of English examples below:

indy/03	watched the way in which the Labour	spin doctors,	Alastair Campbell and Peter
brmags/10	Trees also on the bill, it's	The Spin Doctors	who are pulling the crowd. <p>
guard/13 d.	The frequent charge of too many	spin-doctors	misses the essential point that
guard/13 and	Mr Mawhinney. A Central Office	spin doctor	had leaked to a newspaper a
npr/05	heir opinions. Free of charge, these	spin doctors	massage the press and the public
times/02	gainst the Scottish Claymores. The	spin doctors	have diagnosed it `the battle of
today/06 d	answer session that followed the	spin doctors'	diagnosis of Mr Major's speech

The point here is, if our corpus material dated back beyond the last decade, we would not be able to find a single occurrence of these units. It should therefore be stressed that the period of corpus compilation is also highly relevant to the notion of frequency and should be taken into account with respect to research results. Thus, the corpus perspective on the nonce-words is best summarised in Sinclair's words (1996b:81):

In gathering and organising corpus evidence, the first focus is on repeated events rather than single occurrences. This initial state does

not mean that unique, one-off events are necessarily ignored, but rather that they cannot be evaluated in the absence of an interpretative framework provided by the repeated events.

However, this does not in any way diminish the value of rare occurrences. As Halliday (1985:65) points out, a lexical item of rather low frequency in the language contributes a great deal to the density and compares three sentences with the same proportion of lexical items (1985:65):

- *the mechanism of sex determination varies in different organisms*
- *the way the sex is decided differs with different creatures*
- *different creatures have their sex decided in different ways*

Since the last two sentences include very frequent items such as *have* and *way*, they seem less dense than the first one. Halliday (1995:66) goes on to explain yet another factor operating here. That is, "the last two examples incorporate a repetition, the item *differ/different*. Repetition also reduces the effect of density—since even if a word is intrinsically rare, its occurrence sets up the expectation that it will occur again (Halliday 1985:65). In other words, "a word of low probability carries more information". Seen from this perspective, rare occurrences assume a different value in language contact data from what they have been assigned so far. Importantly, any nonce occurrence, such as the examples below in 8.1.5.1, should be examined in relation to the totality of corpora of Turkish in contact with other languages as the similarity might be significant in terms of generalizations.

8.1.6.1 *licenceli*

The importance of this particular occurrence, regardless of its frequency that would confidently classify it as 'nonce' in the current corpus, is that it answers the following question: *What is listed in the mental lexicon of a Turkish-English bilingual?* I will take Hankamer (1989) as the starting point of my argument where he criticized the Full Listing Hypothesis (FLH) and its variants and suggested that, at least in the mental lexicon of the speakers of agglutinative languages, word recognition and parsing should operate differently from what

research on major European languages had so far demonstrated. The starting point of Hankamer's criticism was that the complex morphology of agglutinative languages such as Turkish allows words of indefinite length produced by means of iterative loops; that these forms are very common and recognised as words by native speakers of Turkish. Consequently, the main premise of Hankamer's parsing model was the recognition of the root before affix-stripping took place (1989). Morphological parsing in word recognition, Hankamer argued, should proceed from left to right, at least for agglutinative languages with rich suffixation. The reason for this strategy is that "the left-to-right recognition approach narrows the choice of possible suffixes at every step to suffixes that can combine with a stem of the current stem category" (1989:402). Even when narrowed down, still a very high number of forms can be associated with any verb or noun root that would go far beyond the storage capacity of the brain. Therefore, Hankamer concluded that "the human mind has, or is capable of acquiring, a parsing mechanism that allows the recognition and understanding of words of impressive complexity" (1989:404). What is interesting from the point of view of language contact is that from the Oznews component of The Bank of English, we get such combinations as *licenced club* and *licenced bar*. However, when an agglutinative language such as Turkish is mixed with English we end up with such forms as

1. *licenceli*

where the Turkish attributive suffix *-li* is attached to the English stem *licence* and it is obvious that the example below is modelled on the L1 and L2 forms. The other forms listed in the bilingual's mental lexicon would be and most probably are:

2. *licence+d*

3. *ruhsat+lý*

However, if *mixing* works in both directions, we should also have

4. **ruhsat+ed*

But we find that this is not the case. This point is important for parsing techniques where the lexicon of systems contains syntactic and semantic information, and depending on the tagger and the parsing technique yet to be developed it might be

possible to process mixed words such as *licenceli* in computational terms. At this point, the entry's morphological makeup and its phonological properties also become important for phonological encoding as there is the question of vowel harmony, as well. That is, we do not end up with forms such as **licencelu* although in some other cases there may be two competing forms as illustrated below.

We find similar cases in the corpus that are also reported elsewhere in data sets of Turkish in contact with other languages such as *shop+u* (the shop - accusative) and *grantlar+la* (with grants).

These cases like *licenceli*, raise the question of 'morphological integration' and lead to a consideration of the status of different types of morphology in the mental lexicon. The same question applies to *-de/-da* that presently co-exists in the data attached to the same English words:

yor55 001fa	sydney ve melbourne	'da	yapýlacaktýr eyalet danýþma
yor161 006fa	multicultural centre	'da	14 kasým tarihinden itibaren de
ts335 002fc	club reception hall	'da	toplumsal ve bayram balosu
ts236 005fa	sydney'de ve melbourne	'de	24 ekimde baþladý süreleri yayýn
yv503 005fb	emerton shopping centre	'de	çalýþýr haldeki take away shop tüm
yv295 006fb	auburn masonic hall	'de	30 ađustos balosu

Yet, we do not see forms like **öleneks*. No explanation has been offered as to why a Turkish stem would not appear with an English suffix. According to Hankamer (1989) Turkish suffixes have independent word status in the mental lexicon of Turkish speakers. If this is the case, then Turkish-English bilinguals would recognize licence+li as two separate words. This would enable them to produce a sequence of indefinite length by means of iterative loops, as they usually do with Turkish stems. Consequently, forms such as *licencelilik*, *licenceliler*, *licencelilerdenmiþ* and so on would be possible. The flexibility of nominal morphology, however, contrasts with the unavailability of German-Turkish verbal forms such as **beantragdim* examined in 8.1.4.1.

At this point two arguments can be put forward. Firstly, the lack of such forms supports the grammaticalization claims on *yap-*. The delexicalized *yap-* provides the flexibility required for the inflection of borrowed verb stems. Secondly, the difference in the way verb and noun suffixes are handled in such mixed forms may in turn indicate that verb suffixes function differently in the mental lexicon from noun suffixes, at least for the speakers of agglutinative languages. While this thesis is not a psycholinguistic study of language contact, the open-minded exploration of corpus evidence that so clearly displays the behaviour of nominal and verbal morphology inevitably raises such questions.

8.2 Significance of patterns

The significance of frequency counts in the study of language contact was already considered by Hasselmo (1961) who used frequency of occurrence to distinguish between the codes. He measured the "impact of the loanwords on American Swedish by studying their frequencies in running text." His point was that

[t]hese frequencies are, of course, a matter of idiolect and subject matter, but by comparing a fairly large number of informants (10), we can at least get a picture of the range of variation and of the average and maximum amounts of borrowing. The loanword content of ASw is also a matter of style which was indicated by the fact that all counts at the beginning of an interview - before the informant had warmed up to the interviewer and relaxed - show markedly fewer loanwords than counts at the middle or the end of an interview. The highest frequencies were noted in recordings of uninhibited conversations between bilinguals, with the interviewer and the tape recorder [...] We could use the terms formal and informal style for the two kinds of ASw, the former being characterized by a relatively low percentage of loanwords, the latter by a higher percentage (1961:61-62).

However, frequency counts in language contact have been used to one end only: *integration* as the sole criterion to distinguish between code-switching and borrowing. This notion is particularly important within a theoretical framework where the core of the study lies in the distinction between borrowing and codeswitching. While the attempt to show the degree of assimilation of a loan in

the new language may be justified in various terms, the fact that it has not been explored further may not. Here, frequencies have been considered in terms of the patterns emerging in the Ozturk corpus rather than as supporting evidence for integration. In other words, borrowing-codeswitching distinction has not been a major concern in the present study.

There is no contradiction here. Naturally, a word list for the entire corpus has been generated and the frequencies of the word forms in use in the immigrant setting have been investigated, but to a different end. It is common practice in computational corpus analysis (e.g. Clear 1993) and also recently in language contact research (Myers-Scotton 1993a) that the cut-off point should be taken as three or more occurrences. It has been discussed throughout the thesis that frequency is a useful criterion but it does not give us the borrowing-code-switching distinction, it rather tells us whether our assumptions about the data are correct or not. In lexicographical practices, integration is an important criterion as the lexicographer has to decide which words will go into the dictionary and which ones will be left out. Yet, for the language contact researcher, there should be no such limitations although it is an issue re the centrality of the item in the structural system.

Baker (1992) points out that the process of translation often results in a specific type of distribution of certain features in translated texts. For example, common words such as *say* and *day* occur with a significantly higher frequency in English texts translated from Arabic than they do in original English texts. Similarly, texts produced in an immigration context can be expected to have different frequency and distribution of certain words than texts produced in monolingual contexts. In fact, words such as *eđitim* (education), *bilgi* (information), *yardıńm* (aid, grant, support), *toplum* (community), *ıþ* (work, job), *aile* (family), *sađýk* (health), *güvenlik* (security), *göç* (migration), *göçmen* (migrant), *hizmet* (service), *vergi* (tax), *okul* (school), *haklar* (rights), *baþvuru* (job application), *emeklilik* (pension) have the highest frequency in the corpus with the counts ranging between 400 and 50. It is to be expected, then, that these words are central to the analysis. I am however reluctant to see these words as forming a core vocabulary as they may simply be a reflection of the texts that form and inform the Ozturk corpus. A

larger corpus that spans 35 years of Turkish in Australia would be necessary to be able to make truly objective statements. Also, parallel migrant corpora would be useful to see whether the coreness of these words applies to other settings as well as other migrant communities within and outside Australia. However, the significance of the patterns investigated should be recognized with respect to the mixed code. These mixed patterns are the starting point of the C_{mx} considered in detail in Chapter 9.

8.3 Motivations

Numerous motivations for using foreign words have been identified in 5 decades of language contact. Some of these are summarised below:

1. The need of quoting English-speaking people and of using English terms which do not exist in the mother tongue (Haugen 1953:65).
2. To reduce decision-making strain and respiratory activity (Clyne 1967:80).
3. As a marker of high social prestige, religious or class identity or political conservatism (Kachru 1978).
4. To express emotion, close personal relationships and solidarity, and to exclude a third person from part of a conversation (Harding and Riley 1986:57-60).
5. To identify with EL culture or at least aspects of it or with a prestigious person who uses them (Myers-Scotton 1993a).

Here, I will deal with motivation by linking it to the notion of non-equivalence discussed in chapter 5. It is known that translators develop strategies in order to cope with specific types of translation problems arising particularly in relation to non-equivalence among languages. The problem of non-equivalence has been investigated at different levels (Baker 1992; Barnett *et al.* 1994 in relation to Machine Translation). Non-equivalence at word level and collocational level are of immediate interest to this study and they have been summarised in chapter 5.

In the course of this analysis, there has been frequent reference to the notion of 'unit of translation' (Tognini-Bonelli 1996a and b). The motivation for producing mixed units can be attributed to the migrant's attempt, somewhat like the professional translator, at the 'nearest natural rendition' in a context where translation plays an important role in everyday life. This may result in a number of units of translation with mixed elements that resemble neither L1 nor L2, as those of the 'third code' explored by Frawley (1984) (see chapter 5).

8.3.1 A word on speaker-innovation

Frequently, in the interpretation of language contact data, the researcher concludes that unfamiliar forms or patterns that cannot be accounted for on the basis of L1 and L2 grammars, are attributable to speaker-innovation. This is another way of looking at the question of motivation. Milroy (1992:169) differentiates between speaker-innovation and linguistic change on the basis of system-oriented and speaker-oriented approaches. He describes speaker-innovation "as an act of the speaker which is capable of influencing linguistic structure" whereas "a change is observed within the language system". Andersen (1989:10), on the other hand, explores the reasons for linguistic change by rejecting the dichotomy between the speaker and the society. He argues that the word 'change' has traditionally given rise to confusion and misunderstandings as a result of the preoccupation with diachrony.

He prefers the term 'innovation' and sees the emergence of a new linguistic entity as a result of "innumerable individual acts of innovation" (1989:14). According to Andersen, innovations can be pragmatically motivated, or they may "arise in the transmission of a language from generation to generation" or they may also arise "as it were, fortuitously out of nowhere" (1989:14). Undoubtedly, the first one is of relevance to language contact settings, *adaptive innovations*, as Andersen refers to them:

An adaptive innovation is a purposeful elaboration of an innovator's competence (a covert innovation), typically motivated by immediate

communicative needs and immediately realized in discourse (in an overt innovation). [...] Adaptive innovations may be premediated (as, for instance, terminological neologisms typically are), but most are not. Some are unquestionably intentional or may be rationalized ex post facto as intentional. But many appear to be made without conscious intent and may be produced in the here-and-now of discourse even without the innovator's being aware of their novelty.

It is clear that mixed units encountered in the present corpus have been produced without the migrant's being aware of their novelty in an ad hoc fashion. Yet they are coded, that is, they appear in the written text (Halliday 1985: Chap 3). So the question to be answered is: how coded are the units of the mixed code? *Codedness* has been defined as "part of the lexicalisation process, whereby strings of morphemes become institutionalised" although low frequency is not necessarily one of its determinants (Moon 1994:32). In an investigation of 'codedness' in fixed expressions, Moon (1994) has identified three criteria of different nature: 1. the quantitative criterion of institutionalisation; 2. the lexicogrammatical criterion of fixedness; 3. qualitative criterion of non-compositionality. In terms of institutionalisation, Andersen (1989:24-25) recognizes "two kinds of metadialogue which take place concurrently" while the innovations are under way. One of these metadialogues concerns the system and the other concerns the norms. That is, feedback on the innovations confirms or disconfirms the user's hypotheses about the language itself and at the same time all members of the community participate in the negotiation of the innovations in relation to the norms. While, this thesis has not approached language contact from this perspective, and nor is there an emphasis on the user and the community as such, the very appearance of mixed units in the community newspapers makes such considerations inevitable. An anecdotal account of this metadialogue is in place here. In a message released by the Turkish Embassy in 1996, on the occasion of a religious festivity, the former Turkish ambassador to Canberra listed some 150 English words he had identified in current Australian Turkish newspapers, such as *hot line*, *tune up*, *charcoal chicken* and *lease*; and invited the community to ban these words from their language (Kanberra'dan Haberler 1996). Whether this ban will be observed in the years to come is of course the subject of another study.

One way of investigating the operation of these metadialogues is to look at the legitimacy of the resources speakers draw on when they innovate. A rich source is the speaker's area of professional expertise. For example, those who work in the area of communication technology are familiar with the term *bandwidth* that frequently occurs in a phrase like 'I don't have the bandwidth to do it, today' meaning 'I don't have the time to do it'. Similarly, the expression 'they've got really poor gui' (=Graphic User Interface) which is used to comment on difficult people is quite common. These innovations, however, tend to remain within that specific community and do not reach the system, as such. Also brand names such as *Ajax* (Halliday 1966:11) and "catchphrases drawn from cinema or television, politics, and journalism and so on may become institutionalised as sayings and other kinds of formula" (Moon 1994:29-30). Similarly, a rich source of innovations for speakers seems to be other languages. In the Bank of English there are many words of Italian origin and some 216 of these occur more than ten times¹⁶. Interestingly, a common culinary word such as *pesto* is never used on its own in the corpus as it is in Italian but always in the company of the word *sauce*, hence the unit *pesto sauce*. This particular innovation may have been modelled on the basis of *white sauce*, *Worcestershire sauce*, *cheese sauce*, *tomato sauce* and so on yet the end result is a mixed unit. That *pesto sauce* should be considered a unit is supported by *apple sauce*, a similar unit which is now written as *applesauce* in American English without the space between the components.

When considered in terms of speaker innovation drawn on the available L2, the mixed units encountered in the Ozturk corpus may have been ad hoc creations yet the fact that they are written and repeated in community newspapers subject them to the community's metadialogue. Their codedness is difficult to capture in spoken language which brings us back to the question of the nature of the language contact data.

¹⁶ I am grateful to Liz Potter, Cobuild for this information, for many useful discussions on the topic and illustrative examples.

8.4 Summary

This chapter has examined a number of case studies identified as representative of the entire corpus and made specific reference to some problem areas in the analysis of language contact data, such as single switches, cultural borrowings, islands, delexical verbs, functors and nonce borrowings. It has considered the distribution of these examples across the corpus and explored the motivations for the emergence of mixed patterns with respect to translation and speaker innovation.

CHAPTER 9:

ON THE CHARACTERIZATION OF C_{mx}

In the light of the case studies analysed in the preceding chapter, this chapter aims at a characterization of the mixed code. Language contact research has often discussed the formation of a mixed code in bilingual and multilingual situations as an additional choice for communication, and the study of mixed codes has sought answers to a number of questions: 1) What produces a mixed code? 2) What are the linguistic rules which govern it? (Domingue 1990); 3) How should such a mixed language be classified? (Vildomec 1963:105). 4) Whether the two codes that produce a mixed one remain unaltered (Pfaff 1979a).

These questions have not found commonly accepted answers and "some of the misunderstanding among linguists on these questions" might have been "due to the fact that it is not clear in all cases whether a new language is the old one with some new elements borrowed, or, whether (and when) a change of language took place in which some elements of the old language remained preserved" (Vildomec 1963:105). For example, in the case of a language such as Albanian, mixture has reached such a level that the bulk of the vocabulary comes from Latin, Greek, the Slavonic languages and Turkish and the native words are limited to a few hundred.

Commonly, pidgins and creoles have been used to clarify the general principles governing the creation of mixed codes (e.g. Mühlhäusler 1980). Generally speaking, diachronic approaches to mixing consider the development of pidgins and creoles. According to Thomason (1995), approaches to language mixing, as far as pidgins and creoles are concerned, can be identified as follows: mixing without extensive bilingualism on the one hand, and mixing with extensive bilingualism on the other. Synchronic approaches to mixing, instead, consider mainly codeswitching in a variety of contact situations. This is a natural consequence of dealing exclusively with spoken data. Moreover, researchers have commonly worked in frameworks that did not allow the mixed code to be studied in its own right but in terms of L1 and L2.

For example, in her study of Swedish-Estonian bilinguals in Sweden and English-Estonian bilinguals in the US, Oksaar (1972) found that her informants violated the rules of both L1 and L2. This occurred particularly in translations and took the form of new collocations. She explained this with a third set of rules developed in addition to L1 and L2. However, Oksaar did not further analyse the inner dynamics of this new code. Instead, when she came to the conclusion that there was 'a new set of rules', she went on to analyse under which social circumstances this was done and by whom. She was following the trend of her time.

Later, a computational approach to the explanation of mixed code grammars in switching was taken by Joshi (1984). In his study of Marathi-English code-switching he attempted to formulate a computational system to characterize the third code. While he observed that the mixed sentences were not always attributable to either L1 or L2, his system was still a characterization of the mixed code in terms of Marathi and English grammars. He rejected Sankoff and Poplack's earlier claim of an independent third grammar and he was mainly concerned with the computational specification of a set of constraints on the switching rule. He was following the trend of early 80s.

More recently, Swigart (1992), with reference to codeswitching in Senegal, argued that Urban Wolof should be considered a third code¹⁷ in addition to Wolof and French. She analyses the situation from a socio-political perspective and concludes that this code cannot be explained in terms of the negotiation of social roles (Myers-Scotton 1993b) or in terms of the maintenance of social boundaries (Heller 1988).

In related fields such as *Translation Studies* there has also been reference to a 'third code' that results from the confrontation of the source and target codes and distinguishes a translation from both source and target texts (Frawley 1984). That is, the translation emerges as a new code, which dictates its own logic (Frawley

¹⁷ Stubbs et al (1985:132) make reference to a D.Phil thesis written by R K Agnihotri who studied the British-born Punjabi speakers in England and came to the conclusion that a mixed code was emerging and the Punjabi speakers selection of the mixed code depended on place, person, topic and certain internal factors. At the time of writing, this thesis was unavailable to me.

1984): "The translation itself, as a matter of fact, is essentially a third code which arises out of the bilateral consideration of the matrix and target codes; [...] that is, since the translation truly has a dual lineage, it emerges as a code in its own right, setting its own standards and structural presuppositions and entailments." Frawley also observes that the new code is not always reducible to target and matrix codes.

The matrix and Target codes provide only input into the third code. In a sense, they form part of the third code's redundancy, but insofar as the third code supersedes its matrix information and target parameters, it differentiates itself, emerges as new information derivative of the matrix and target redundancy, but further establishing its own predictability as an individual code.

A similar concept in the field of *Second Language Acquisition* is *Interlanguage* (IL) that is "the existence of a separate linguistic system" resulting from "a learner's attempted production of a TL norm" (Selinker 1972). The problem of *transfer* has been central to the study of *Interlanguage* as well as to the study of Language Contact. As defined by Selinker, it involves the transfer "of the structural patterns of one's native language to a foreign language" (Selinker 1969). Although the questions asked are the same for both *Interlanguage* and *mixed code* research (Jake 1998), *Interlanguage* studies are specifically concerned with learning and teaching perspectives while studies of the mixed code are relevant to immigrant settings. The questions "what language transfer consists of, what actually is transferred, how language transfer occurs, and what types of language transfer occur" (Selinker 1969) are also at the centre of language contact research.

The difference between these two lines of research is that the mixed code is used at the community level. It is used by migrants to communicate with the members of the same community. IL on the other hand, is used by language learners to communicate with native speakers, and because of the individual differences it does not become a shared communal property in the sense a mixed code does, no matter what similarities may exist between ILs of second language users.

The idea of mixed codes has been around since 1800s. In the early writings such as *Memoires de la SociØtØ Linguistique de Paris* or Whitney's *On Mixture in*

Language, the writers were mainly concerned with the question of a mixed grammar (see Hasselmo 1961 for a review). In his doctoral thesis, Hasselmo (1961) discussed the existence of 3 different codes used by the different generations of Swedish migrants in the US and called them 'layers'. His study treated grammar, phonology and lexicon separately.

The approach to the description of mixed-code grammars has so far been essentially contrastive. That is, the examples from the mixed codes under examination have been considered in contrast with what is and is not possible in either L1 or L2. However, forms that neither match the structures of L1 nor those of the L2 are common in such codes (Tognini-Bonelli 1996b).

An important issue that arises from these concerns is whether the mixed code will ever reach the status of *linguageness*. The reason why it usually does not may be better explained in relation to the absence of metaphor rather than theories of simplification, convergence and so on. This is often related to the process of nominalization in language. Yet, the frequency of nouns is an aspect of language contact data commonly taken for granted. The status of *linguageness* and that of *codeness* are discussed below.

9.1 Codeness vs linguageness

It is not clear within Language Contact whether the distinction between *language* and *code* is purely terminological. The debut of the term *code* probably dates back to the early decades of language contact.

Outside the field of language contact, Bernstein started his influential work, *Class, Codes and Control* in the late 1950s. His notion of codes originated from specific class-related questions raised by demographic studies in Britain and was mainly sociolinguistic. He argued that social relations regulated the meanings speakers created, which, in turn, acted selectively on lexical and syntactic choices. Bernstein drew a distinction between *restricted* and *elaborate* codes that controlled different ranges of syntactic alternatives and combinations within different

contexts (Bernstein 1987). This social perspective on *code* was not related to the codes examined in immigrant settings but it was appreciated within language contact circles (e.g. Gumperz and Hymes 1964).

Bernstein's work coincided with the equally powerful sociolinguistic notion of *domains* (see page 73) promoted by Fishman in the 1960s. He similarly assumed a restriction in language use in the context of migration. Also at this time Halliday, McIntosh and Stevens (1964) introduced their concept of restricted registers. Consequently, the term *code* in language contact came to mean that under certain conditions such as migration, the community language in question would inevitably be restricted. It follows from this line of thinking that when there is reference to *mixing* in such contexts, what is involved is *code* (e.g. code-alternation, code-switching, code-mixing, code-copying etc.) and when there is reference to mixing with respect to *pidgins and creoles* the preferred term is *language*.

The issue here then is when code is used it is commonly assumed that there is a restriction in the migrant's language. It seems so, at least in the words of Silva-Corvalán (1995b:10), who, in fact, explains how this restriction occurs:

In language contact situations bilinguals develop strategies aimed at making lighter the cognitive load [...]. In the use of the subordinate or functionally restricted language, beyond phonology, these strategies include (1) simplification of grammatical categories and lexical oppositions; (2) overgeneralization of forms, frequently following a regularizing pattern; (3) development of periphrastic constructions either to achieve paradigmatic regularity or to place less semantically transparent bound morphemes; (4) direct and indirect transfer of forms from the superordinate language; (5) code-switching, which involves the use of two or more languages by one speaker in the same turn of speech or at turntaking points.

This view inevitably leads to or facilitates the postulation of *attrition, shift and death continuum*. If, and only if restriction is not assumed, however, the

emergence of a third language can be considered, similar to the case of pidgins and creoles.

Another relevant argument that can be put forward with respect to *code - language* distinction comes from Sinclair's (1996b) work where he draws attention to two contrasting views in linguistic analysis. To date these two views remain irreconcilable. While the point of his argument is not directly related to language contact research (the contrast is between Natural Language Processing and Corpus-driven approach), there are striking similarities with the opposing but not clearly stated views within Language Contact:

In one view, language is primarily a carrier of messages. The propositional content of the sentences in a text - or most of the content of most of the sentences anyway - can be retrieved and symbolised in a knowledge base. The form of the sentences is only of value insofar as it does its job properly and allows the messages to be transmitted efficiently. The components of language text - words and phrases - have known meanings (such as are explained in dictionaries), and the process of construction of text is the selection and arrangement of these components according to the meaning that is to be delivered, and within the prescribed rules of construction - the grammar of language.

Sinclair points out that *code* is preferred by the proponents of this view (whom he refers to as *Academicians*) and examines their attitudes to such issues as *terminology*, *sublanguages*, *lexicons*, *selectiveness* and so on in contrast with the other view, the *Thespian* approach, according to which

language is a means of communication that deals in much more complex communications than messages, although it recognises that messages are important, even though very difficult to define. The form and the message cannot easily be separated, and the particular selections in a text interact with each other to such an extent that it is impossible to sustain the position that they deliver a stable unit of meaning on all occasions.

As I summarise Sinclair's argument in terms of the issues listed above, it will become clear why these views are relevant to Language Contact. The first

position's attitude to terminology is to protect individual terms from the normal effects of the usage of words (i.e. variation and change) and to keep their meaning constant, as is generally done.

The other view instead questions the way single lexical items are perceived regardless of their status as terms or something else. It emphasizes the syntagmatic associations of words and the subsequent formation of new units of meaning, which require both (or more) words. This emphasis comes from the fact that the line between what is a term and what it is not cannot be drawn so easily. To quote some of Sinclair's 'quasi-term' examples, typically *mouse* and *window* in computing have not acquired the status of term.

The reason why this point interests language contact researchers is obvious. Many lexical items examined in language contact data collections are discarded as technical terms, specialised vocabulary or cultural borrowings. Flexibility as to their status on the part of the language contact researcher is low. This approach to loan material fits in well with the first view outlined above. The problems with this approach have been dealt with in chapter 3.

In terms of their attitudes to the notion of *sublanguages*, the first view is in agreement with the prevailing attitude in Language Contact. The assumption is that on certain occasions language users accept a set of restrictions on their expression. The restrictions normally occur in connection with a specialised topic that controls the vocabulary selection or with the function of the communication: "Hence the position arises that limitations in the messages of some varieties will lead to simplifications in the structure of the language used" (Sinclair 1996b).

This view on the limitations of use and the consequent simplification of structures brings in mind Fishman's concept of *domains* in the context of migration. The main point of this approach is the development of sublanguage under given conditions. However, as pointed out by Sinclair (1996b), "[t]he specification of a sublanguage is controversial, and so is empirical confirmation of their existence", although in Language Contact research there has been less controversy over the

argument that within migrant communities sublanguages would inevitably emerge. The 'empirical confirmation of their existence' is of course another matter:

To demonstrate the existence and importance of sublanguages, there are possible strategies. One is just to look for them. A user community that separated clearly its language about a subject matter area, and whose usage in that area differed markedly from its other usage and the usage of comparable communities, and the difference was that the language used in that area comprised a sub-set of the general language - such conditions would identify a sublanguage.

This picture is similar to the one described by Language Contact researchers who mainly concern themselves with the societal level of analysis, along the lines of Fishman's work. The following picture, on the other hand, is more familiar to the followers of Myers-Scotton (Sinclair 1996b):

Another way of identifying sublanguages is to imagine that they are **embedded** within less disciplined, more liberated text. Only certain sentences, in this approach, are examples of the sublanguages in action, and the rest are ordinary language sentences. [emphasis added].

As highlighted in bold type in the quote above, *embedded* is an adjective that figures prominently in Myers-Scotton's writings. Labels such as *insertion*, *convergence*, *encapsulated* and so on are also typical of this approach. Not surprisingly, "[t]he principles of limitation, restriction, selection, and simplification are central to the notion of sublanguages" (Sinclair 1996b). Equally central to this type of linguistic analysis is *sentence* as the common unit of analysis, and based on their structure some sentences are assigned higher value than others. In other words, only those sentences that perform a certain function are selected for examination. This is typical in the analysis of Language Contact data where only those sentences with L2 lexemes are valued.

In terms of the lexicon, this view favours "an entry for each word" approach and lists its morphological, syntactic and semantic characteristics with the assumption that attaching a set of features to a word as such is a possible procedure. It is also

assumed that "where the word is not in fact the appropriate unit of meaning, and perhaps a multi-word unit or a sub-word unit has to be identified, reliable criteria can be found to do this job" (Sinclair 1996b). However, this kind of analysis leaves out many of the commonest words in the language, namely the "grammatical" or "function" words as it also assumes that:

(1) there are two kinds of words in a language, one for which an explicit, permanent lexical profile is appropriate - the so-called "vocabulary words"; and (2) that objective, scientific criteria can be stated for deciding on the dividing line between these words and the rest [...].

This is again the position of most language contact researchers in relation to the matters of the lexicon. One common observation is that content words constitute the bulk of the borrowing process and few function words are found in language contact data. This distinction is discussed in more detail in Chapter 8.

From a thespian perspective however the lexicon looks quite different from the conventional one with the recognition that through collocation new meanings are created all the time:

There is no familiar set of properties permanently attached to each word, there is no discernible starting-point for a word, there are no preconceptions, no prerequisites; only a format for how the lexicon entry will be built **up through examining the usage of the word**. [emphasis added].

This is yet another point of corpus research highly related to Language Contact research as the conventional list of loan words would not work within such a lexicon. I shall return to this topic in chapter 10 where I propose a lexical approach to Language Contact.

In terms of *selectiveness*, the message-centered view will identify some sentences as more interesting and important than others with the notion of *sentence* "subtly redefined to support this prioritisation" (Sinclair 1996b). The preoccupation with

well-formedness also belongs to this view (discussed in 4.2.3.1). Within this approach, if sentences do not exactly fit the predefined unit of analysis, "with minor amendments the number of qualifying sentences can rise sharply. Often the amendments required are merely clarifications of sentence boundaries" (Sinclair 1996b). Such amendments are often necessary also in language contact data as clearcut units are difficult to find in natural texts:

The traditional grammar and many subsequent versions of grammar are only operative on certain classes of sentence, and those sentences that do not fit the analysis are put to one side, or adapted, or just ignored. This is part of the everyday routine of many linguists, and is remarkable [...] Instead of describing naturally-occurring text, the descriptions might apply only to texts which had been selected or adapted so that they fitted the description. This is the unavoidable risk that arises when a formal system meets raw data. A formal system defines what it can describe, and is restricted to that; since what it can describe is never exactly co-extensive with naturally occurring data, the fit, and the relevance, must be only approximate. The nature of the approximation, the way in which the rigorously defined categories are related to the data, is one of the central issues in linguistic theory.

The theoretical line of Language Contact research approaches data with a similar procedure. In fact, the ongoing debate on constraints is a good example of this within Language Contact.

To conclude this summary on the contrasting approaches to linguistic analysis, it should be emphasized that in the thespian perspective "everything does not fit together as the prevailing language models predict" (Sinclair 1996b). To the proponents of this approach "this is not just a reflection of the early state of data orientated language research, but is an inalienable feature of natural language" (Sinclair 1996b). It will be obvious by now that this study follows the thespian view and avoids making assumptions on restriction with respect to the data under examination.

9.1.1 Mixed Code and the absence of metaphor

The specification of mixed codes has posed a methodological problem to the language contact researcher and led to the impression that the mixed code can never be assigned the status of a proper language because its properties have never been fully described. And when they have, these descriptions have mostly concentrated on the syntactic properties of L1 and L2.

It goes without saying that, almost all of the available language contact studies have used spoken data, and this tradition has a bearing on the interpretation of the processes under investigation. However, it seems that, the problem of mixed code descriptions is not whether the emerging lexicogrammatical structures are well-formed or not. It is rather the randomness of *grammatical metaphors* in speech, used in the Hallidayian sense (e.g. 1985:93-96). This insight is provided by the writtenness of Ozturk corpus and needs to be explained. Let us consider the different nature of spoken and written texts. If we compare the way experience is reflected in speech and writing, we can see that few different verbs occur on the written side and a greater range of verbs on the spoken as Halliday's (1985:93-94) examples illustrate:

applause followed the announcement	'after the announcement people applauded'
this development could lead to a different outcome	'in view of this development, things might turn out differently'
he derived much satisfaction from this discovery	'because of this discovery he was very satisfied'
her speech covered five points	'she spoke about five points'

The verbs used in the spoken text are "turned by the written version into nouns [...] like *applause*, *speech*, *outcome*, and *satisfaction*," and in a "more spoken equivalent, *announcement*, *development*, and *discovery* might also appear as verbs" (1985:94). In other words, "the metaphor is in the grammar", where a verb

is represented as a noun. In written text, grammatical metaphor tends to occur more frequently than in speech and nominalization is the way to achieve this. In some cases grammatical metaphor is associated as much with speech as with writing. In speech, the principle at work is to represent a process by combining "a lexically 'empty' verb" (*have, do, make, take, go, give*) with a noun expressing the event or action: *have a bath, make a mistake, give a growl, and so on*" (1985:95). This point brings us back to the delexicalised verbs, investigated in chapters 7 and 8. These frequently appear in spoken language contact data in European research on Turkish. While we encounter combinations like *have a bath* with a Turkish delexical word and an L2 noun, a nominal clause such as *tradition dictated the wiping of food with hands* is unlikely. In the written texts of Ozturk corpus, translations do contain such nominal clauses (e.g. see the text about *Ombudsman*), however, mixed units as such are usually confined to simple nominalizations such as *lease yaptı*, and do not extend to more complex nominalising structures. However, if the 'code' is to be assigned the status of 'language' in contact settings, written text should also be considered. This should be done particularly in relation to the complexity of grammatical metaphor observed in L2 verbs exploited as nouns. While the frequent structure comprising of an L1 delexical verb and an L2 noun is a step towards the metaphorical use of the mixed code, it does not display higher level of sophistication. It should be mentioned here that grammatical metaphor is not usually acquired by children until the age of eight or nine (Halliday 1985:96). Grammatical metaphor considered with respect to mixed data in this sense is different from the *metaphorical switching* (Gumperz 1971) used by speakers to convey some of the social and cultural associations of the other code.

The other point related to metaphor emerges from the Bank of English, again. When we observe the loanwords of Italian origin in this collection, we can see that the metaphorical uses attached to these are limited¹⁸. A very strong example is the word *spaghetti* with frequent metaphorical uses as in *Spaghetti junction*. However, it is not common for loanwords to acquire metaphorical meanings and this aspect should be considered as another criterion for integration.

¹⁸ I am grateful to Liz Potter for this example.

9.2 The search for a unit of analysis

Traditionally, there have been two basic units of linguistic analysis: *the word* and *the sentence*. The status of the word was later challenged with the entry of the *morpheme* in the early decades, as the smallest and a more suitable unit of analysis. But the models available were concerned with the identification of morphemes combining into words, sentences and discourse, and although multi-word stretches had caught the attention of some researchers, they were never given priority (Sinclair 1996b:76). In such a treatment, *compounds*, which are of significant frequency in language contact data, were also "considered as marginal phenomena, almost aberrations, exceptions that prove the rules" (1996b:78). Similarly, phrasal verbs, "fixed phrases, variable phrases, clichés, proverbs, and many technical terms and much jargon" were recognised as patterns but they were "tucked-away, well-off centre" in grammatical description (1996b:78). They seemed "anarchic, individual, unstable, one-off items" which didn't "fit into a tidy description" (1996b:78). Sinclair (1996b:80) identifies the following types of relationships between individual words and the meaning in such patterns:

1. None of the words may appear to contribute directly to the meaning of the expression (*bear on = be relevant to*).
2. Some may, while others may not (*to beat someone up*).
3. Each still seems to mean what it normally *means* (*the rain beats down*).

With reference to the third type, *collocation*, Sinclair says this co-occurrence "does not have a profound effect on the individual meanings of the words, but there is usually at least a slight effect on the meaning, if only to select or confirm the most common meaning" (1996b:80). Thus, words are not totally independent in forming such relations as "other patterns cut across them and constrain them". So far the constraints postulated have been syntactic, and this has been the case in language contact, too (see Chap 3 for an overview). This thesis, following Sinclair (1996b), highlights *lexicogrammatical tendencies* that "operate often at the level of word choice" (1996b:80). Sinclair's explanation is based on his *open choice* and

idiom choice principles (see 4.2.3.1). The patterns that favour the open choice show a *terminological tendency*,

which is the tendency for a word to have a fixed meaning in reference to the world, so that anyone wanting to name its referent would have little option but to use it, especially if the relationship works in both directions.

This is related to the question of *cultural borrowings* in language contact where a word may be borrowed simply because its referent does not exist in the migrant's native language and therefore is treated as a technical term in traditional data analysis. However, cultural borrowings may gradually dispose of their cultural load and turn into common lexical items in language (see Chap 8). This point is further supported by the fact that words "cannot remain perpetually independent in their patterning unless they are either very rare or specially protected" like technical terms (1996b:82). Even technical terms go through a similar process, as for example illustrated by the case of *virtual*, a rapidly determinologized word (Meyer, Mackintosh and Varantola 1998). Consequently, they may start displaying interesting patterns of combinability, which then acquire features of idiomaticity (see Chap 8). Once

they begin to retain traces of repeated events in their usage, [...] expectations of events such as collocations arise. This leads to greater regularity of collocation and this in turn offers a platform for specialisation of meaning, for example compounds. Beyond compounds we can see lexical phrases form, phrases which have to be taken as wholes in their contexts for their distinctive meaning to emerge, but which are prone to variation. [...] this variation gives the phrase its essential flexibility, so that it can fit into the surrounding context (1996b:82-83).

Obviously, some patterns such as "*door* and *window*" with "*room* as a significant collocate" simply "correlate with the world" and do not contribute much to our knowledge of "collocational selectivity", "unlike *slammed* with *door* or *seat* with *window*" (1996b:82). In collocational selection, both lexical and grammatical

structure should be considered as "different components of the phrase carry out distinct functions; this 'division of labour' is a strong hint of a larger unit of meaning" (1996b:83). The co-occurrence of grammatical choices (Firth 1957) has been defined as colligation and in many cases the patterns observed can be explained as a blend of collocation and colligation (see Sinclair's analysis of *naked eye* in 1996b).

If the unit of analysis is taken to be the unit of meaning rather than 'single lexemes' the patterns in language contact data will require a different interpretation and Myers-Scotton's table of 'singly-occurring English lexemes' (1993a:15) will have to be re-examined:

Table 9

	<i>Types</i>	<i>Tokens</i>
Nouns	141	174
Verbs		
Finite	91	91
Infinitive	37	37
Past participle	13	15
Adjectives	27	36
Adverbs	11	11
Interjection	4	4
Conjunction	2	2
Pronoun	1	1
Possessive Pronoun	1	3
TOTAL	328	374

EL islands 121

EL sentences 44

In the light of the discussion presented in Chapter 8 in relation to the emergence of new units of meaning such as *delivery yapıyı*, *register belgesi* and *licenceli*, the validity of previous taxonomic approaches disappears. Extensive corpus research conducted in Birmingham shows that collocations develop a life of their own and

tend to acquire idiom status through repetition. So mixed units cannot be analysed on a slot-and-filler basis but should be seen in terms of co-selection tendencies. While Myers-Scotton's idea of *islands* recalls the idea of collocation, the idiom principle has a lot to tell us about these islands. For example, a unit such as *I guess* in Heller's data (1990) is a good example of how collocation and colligation blend in mixed data:

e.g. *je suis une Canadienne-française I guess* (Heller 1990:67)

The frequency of occurrence of *I guess* as a single unit is supported by the evidence from the Bank of English, as it appears mostly in the sub-corpora of US and Australian texts. As the examples below illustrate, when *I guess* occupies the final position in a sentence, it behaves differently from when it is the sentence initial unit and acquires the meaning of 'perhaps':

oznews/16 ck was coming, we all knew that I guess. <p> So why not just jump to the Flick?
usbooks/12 knew his dad was an alcoholic, I guess. Well, now his dad is dead." <p> My God!"
usbooks/12 honey. I could have said that, I guess. I intended to write, but I didn't. Not a
npr/05 pens up the door of vulnerability, I guess, for every museum in the country and
npr/05 I--but he can hide a select part, I guess, of his military capability, but he can't
npr/05 for H. Ross Perot is coming, I guess, just about three weeks after you said he

9.3 Empty C_{mx} lexicon first

If we look at the ways loanwords have been analysed and interpreted in Language Contact research so far, we will find that none of the models work because they assume that the C_{mx} lexicon is already full at the start. Before proposing a new model, let us first summarise the major approaches to loan-words in Language Contact. The broader picture is that language contact researchers dealing with the treatment of loanwords in bilingual data have been torn between what counts as borrowing and what counts as code-switching since the early days of the field of language contact (see chapter 3). While the code-switching-borrowing dilemma has never been solved, it has played an important role in the way loan-words have been handled in Language Contact. Traditionally borrowing has come to be seen

as a diachronic process whereas code-switching as a synchronic one. Although the implications of such a distinction are important for various areas of Language Contact research, the criteria for the classification of loans as either one or the other have never been truly reliable. The approaches to loan-words over the past few decades can be summarised under a number of headings (Kurtböke and Potter 1998).

Model - replica approach

This approach has been popular in the early days of Language Contact research (e.g. Johanson 1993) when bilinguals were not seen as 'skilled performers' but as confused individuals. Consequently, if the loan-word used by the bilingual did not match the original form, mostly in its phonemic shape, we would face "a case of partial learning" (Haugen 1950). Such an approach, of course, did not work when it came to unintegrated loan material such as *rendez-vous* (Myers-Scotton 1993a).

'Slot-and-filler' approach

Around the same time as the *model-replica approach*, there was another approach in fashion that saw the lexicon as an inventory of items. This was a natural consequence of the developments in the field of Linguistics at the time, and emphasized the centrality of syntax. Within this approach, bilinguals used loan-words to fill in the slots in their L1 structures. This approach has been popular until the new view on the Lexicon started taking over in the previous decade (e.g. Sinclair 1987, Sinclair 1991a).

'Integration' approach

A very common approach to the identification and analysis of loan-words in data has been the measurement of their frequency although initially it was not a criterion for integration but stood alongside it. The notion of frequency gained more importance as the debate on what counts as code-switching and borrowing has come to occupy a central position in Language Contact research. However, criteria based on frequency counts, as in Myers-Scotton's 3+ occurrence metric (1993a), have proved to be less reliable than previously thought for the distinction of borrowing from code-switching (Kurtböke and Potter 1998). Importantly, the emphasis in Language Contact has been on morpho-syntactic integration, and

lexical issues as such have been highlighted only when they served as further evidence towards syntactic constraints. The 'integration approach' is still very commonly used today (e.g. Backus 1996).

'Form' can be borrowed separately from 'meaning'

This view was a by-product of the approach that claimed different levels of analysis such as syntactic, morphological and semantic. In this view, the major premises of the lexicogrammatical approach, such as the co-dependence of form and meaning had no place. For a long time, it has been thought that the form had no impact on the meaning and vice versa. However, the way corpus examples has been analysed in Chapter 8 shows that out of a number of forms of a word, only one form is usually borrowed and a particular meaning is attached to that form. Also the collocational and colligational patterns have an impact on the loan as well as the loan on its environment.

'Cultural' borrowings vs 'core' borrowings

The latest of these approaches has been put forward in Myers-Scotton's MLF model (1993a:5) where "cultural borrowings represent objects or concepts new to the ML culture" as opposed to "core borrowings for which the ML always has viable equivalents." This thesis has argued that this distinction does not have any bearing on the status of mixed units.

As stated at the beginning of this section, all of these approaches consider the lexicon full at the start. Such views, however, do not allow the specification of the mixed code lexicon as an independent construct. Starting our analysis with a full lexicon is a problem as researchers are constantly preoccupied with the patterns of the L1 and L2 lexicons. Even if they subscribe to the view that a mixed code is emerging, it is difficult to make room for the lexicon of such a code. According to Sinclair (1996a), this problem can be tackled if:

The lexicon is considered empty at the start because nothing appears in it except what is gleaned from the study of the language in use - nowadays, through the study of corpora.

The implications of the empty-lexicon approach for Language Contact research are important. However, a major shift is required in the way the lexicon is perceived. Rather than going to our data with a set of assumptions, we should concentrate on usage. As Sinclair (1996a) puts it:

To build an adequate lexicon, we must start with usage. As speakers of the language, even as experts in its lexical structure, **we cannot reliably anticipate usage**, and so we have to study large samples of the language to uncover the regular patterns. [emphasis added].

The observation that 'we cannot reliably anticipate usage' raises the need to review a major question in language contact: the 'predictability of switches'. Within the view advocated here the popularity of the predictability debate disappears, as well. Sinclair (1996a) further explains why we should not assume our lexicon full at the start:

There is no assumption that meaning attaches only to the word; it is anticipated that meanings also arise from the loose and varying co-occurrences of several words, not necessarily next to each other. It is, thus, not possible to compile a list of entries in advance of analysing and interpreting the evidence, because the lexical items are not always words, and each word may enter into a variety of relationships with others to realise lexical items.

Thus, if we examine loans in relation to their environment, rather than as singly-occurring items, and observe their co-selection tendencies, we can see that the meaning of a word and its immediate new context have become inseparable and it may gradually acquire idiom status. So, rather than perceiving foreign words in data as "singly-occurring switches" we should start building the mixed code lexicon by paying attention to the patterns growing around the "foreign material".

As for the format of such a lexicon, Sinclair (1996a) proposes three components: *a form of a lexical item, an environment, and a meaning*.

A word becomes associated with a meaning through its repeated occurrence in similar contexts. The distinction between the item and

its environment is not clear-cut, because the choice of a meaning has a profound effect on the surrounding text, one which is not suddenly cut off at a boundary, but which is correlated with adjacent meanings. Similarly, the domain of meaning does not consist of discrete entities, "meanings", to each of which can be linked a form; it is assumed to be an amorphous area that is ordered by the number and type of lexical items. Hence the construction of the lexicon requires us to vary all three components against each other [...].

An aspect less considered in language contact is the multiple meanings of words as the "one word-one meaning" attitude to the lexicon prevails. This approach gains importance when we consider the fact that nouns are the most easily borrowed category. An approach that insists on the centrality of verbs, regardless of the overwhelming number of nouns, provides little insight into the behaviour of singly-occurring nouns and nominal groups. Unless we start considering the grammar of nouns in terms of their syntagmatic environment, language contact research will continue along the paradigmatic axis. The insistence on paradigmatic axis and hierarchy has been dominant in the study of language contact (Muysken 1990). In the work of Poplack (1980) for example switching is allowed when the orders of both languages are respected; or as in the work of DiSciullo, Muysken and Singh (1986) switching is allowed when the demands of grammatical coherence of both languages are respected. Syntagmatic relations have received scattered attention, never been fully explained, and explored to a different end, as can be seen in the words of Muysken (1990).

In the case of borrowing syntagmatic relations need to be maintained in the system, because elements with valency (verbs and prepositions) can be integrated less easily than elements without valency (nouns and adjectives). The paradigmatic relations in the system must be maintained so that function words, that are defined in terms of a feature system [...] can be borrowed more difficultly than content words.

Such views, of course, favour the patterns of L1 and L2 and an example such as the following

...welche wir im Continent nich haben (Clyne 1967:127)

where *im* would not occupy this position in either language is considered *collocational failure*¹⁹ Similarly, Silva-Corvalán (1995b:6) reports *en los sábados* 'on the Saturdays' where *en* and *los* should not go together according to either Spanish or English rules, although she calls this *overgeneralization*.

This brings us to a major question facing those researchers who subscribe to the emergence of a mixed code in immigrant settings: *when does the mixed code establish itself?* From the discussion so far the answer is probably when mixed collocations develop a life of their own. In other words, when *the idiom principle* (Sinclair 1991) starts operating. This point is also closely related to the debate on grammaticalness vs. naturalness as each migrant setting is likely to set its own naturalness criteria. Accordingly, a noun that collocates with a certain set of adjectives out of a number of possibilities in L1, may not collocate with a similar set when it is used in the L2 setting. If it does, we can assume that the collocational patterns of either of the participating languages are not disturbed in the case of contact. But if it does not, it should be granted that the new pattern is the starting point of the mixed code. So far, the study of language contact has concerned itself with the borrowing of those lexical items which seemingly exist in both languages. Thus, the major question has been 'if both languages have the same lexeme why is it that the user still feels the need to borrow?' Such a view assumes equivalence between languages, of course. However, if we look at the experience of professional translators and recent corpus work on translation, we can see that equivalence is a problematic notion. In fact, non-equivalence is more common than equivalence (see Chap 5). Consequently, rather than 'equivalents' we should consider the comparable units of meaning. The end result will be a unit of translation that will not resemble either L1 or L2, similar to those mixed units that emerge in immigrant settings.

¹⁹ I have come across this term in Owen (1986).

9.4 Mixed Code and change: A Comparison of the 80s and 90s

In this thesis, the material from Ozturk corpus has not been presented within a 'language change' framework in spite of the diachronic nature of the corpus that covers a period of 15 years. This section will not attempt either, to link the points discussed to a change, or a gradual 'turnover' to use the popular terminology. The reason for this is my conviction that on the basis of the available data collections, which are both synchronically and diachronically of very modest size, such claims cannot be validated. We need larger corpora covering longer periods if such claims are to be made. Also, generationally speaking, it is difficult to say how Turkish will change in Australia, or elsewhere for that matter, across the next few generations, although it is a commonplace to do this (e.g. Backus 1996). That is because, as explained above, we cannot reliably predict usage. The present section rather goes back to the corpus, again with an open mind, to see whether there are differences in the patterns used, between texts from 1980s and 1990s. This is done in the following way: word lists have been computed for each year separately. However, in terms of text numbers there is a discrepancy as can be seen below (see also chapter 7):

1980 - 118 texts	1990 - 305 texts
1985 - 207 texts	1995 - 371 texts

In order to keep the size of the sub-corpora even, the frequency counts of year 1 and year 2 texts have been collated, which amount to 325. The balance has been achieved by comparing them with only those 371 texts from year 4.

An interesting aspect of the data is the change in the frequency of occurrence of some lexical items over time. For example, the word *emekli* (pensioner) and its various forms appear a total of 4 times in 1980 in comparison to 50 times in 1995:

<u>word form</u>	<u>year</u>	<u>frequency</u>
emekli	1980s	8
(pensioner)	1995	21

emekliler	1980s	1
(pensioner+pl)	1995	5
emeklilere	1980s	4
(pensioner+pl+dat)	1995	1
emeklileri	1980s	1
(pensioner+pl+acc)	1995	2
emeklilerin	1980s	1
(pensioner+pl+gen)	1995	0
emeklilerimizi	1980s	0
(pensioner+pl+poss+acc)	1995	2
emekliliđi	1980s	0
(retirement+acc)	1995	1
emekliliđine	1980s	0
(retirement+gen+dat)	1995	1
emeklilik	1980s	5
(retirement)	1995	13
emekliye	1980s	0
(pensioner+dat)	1995	2
emeklilikde	1980s	1
(retirement+loc)	1995	0

total	1980s	21
	1995	48

This indicates, understandably, a change of content in a period of 15 years. At the time of arrival in Australia, the members of the first generation were 30 years younger than they are now and for a while their priorities were housing, work and childcare. Naturally, the early members of the first generation are approaching the

pension age now and the frequent occurrence of *emekli* in 1995 texts shows that 'retirement' is moving up on their agenda.

A comparison of early and recent texts in terms of frequency is undoubtedly an attractive way of analysing language contact data. However, the corpus in hand, while in terms of the number of texts used can be considered large, in terms of word-count remains modest. Therefore, it does not allow the study of a wider range of words to justify change claims. This limitation is discussed in Chapter 10.

9.5 Summary

This chapter has dealt with the notions of *codeness* vs *linguageness* as well as metaphor that play an important role in the evolution of codes into languages. It has been argued that in the study of language contact there has been a problem in the characterization of mixed codes and that problem has been methodological. A new unit of analysis has been proposed as opposed to traditional units, word and sentence. A new construct, C_{mx} , has been proposed on the basis of the 'empty lexicon' approach. Finally, the data have been observed with a view on language change.

CHAPTER 10: Conclusions

The section below deals with the question: "What have I found in the course of this study?". The answer has been sought in relation to the set of questions listed in 1.1 as issues of general concern in language contact. There is also a summary of how the three research questions listed in 1.2 have been dealt with in a framework that has brought a number of research areas together; namely, Language Contact, Corpus Research and Translation Studies. The reason for bringing corpus methodology into this study was to offer an alternative to the traditional data gathering and analysis techniques in language contact research. Translation studies, on the other hand, provided support for the points made about non-equivalence and mixed code in this thesis. The four preceding parts, *Introduction; A meeting place for language contact, Translation and corpus research; Ozturk corpus; Internal operations* have been organised into nine chapters. The first part has set the background to the study and summarised the current state of research on Turkish in contact with various languages in and around Europe. Chapter 3 represents a novel type of literature review. Rather than taking the available concepts and theories as the centre of the overview, it has traced the development of the current debates in Language Contact across five decades, since 1950. That chapter has concluded that our thinking on certain issues and approaches has been shaped by the work of the early generations and needs to be revisited. Chapter 4 has surveyed the major areas of corpus research and highlighted the work of Sinclair and his colleagues who have played an important role in the development of the current project. Chapter 5 has emphasised the importance of translation in the migrant's everyday life and drawn parallels between language contact research and translation studies. Chapters six and seven in Part three have looked at the Turkish community in Australia as the provider of the newspaper texts used in the compilation of Ozturk corpus. The rationale behind text selection and processing has been explained. It has also examined some characteristics of written texts such as *nominalization* and *delexicalization*. Chapters eight and nine in Part four has dealt with the analysis of the corpus and the mixed code. These are detailed in the sections 10.3.1 to 10.3.10 below.

Before a summary of the thesis in relation to the broader areas of concern, the lexical tendencies observed in the data will be listed (10.1) and a lexicogrammatical approach to language contact will be proposed.

10.1 Lexicogrammatical tendencies in language contact

According to Thomason (1997) counter examples have been proposed for all the grammatical constraints postulated so far and no constraints should be imposed on language contact data. Grosjean (1990) also states that constraints on processing should be seen as general tendencies rather than absolutes. In fact, what follows is a list of some tendencies observed throughout the analysis of the *Ozturk Corpus*. The main premise of this thesis has been that "foreign material" in the host language will have an effect on its environment, and a distinctive pattern will develop around it regardless of the way it has been perceived so far, typically as one or a combination of the categories below:

- intersentential
- intrasentential
- individual item or unit
- content word
- function word
- cultural loan
- core borrowing
- assimilated
- unassimilated
- inflected
- uninflected
- marked
- unmarked

Consequently, the tendencies of the "foreign material" encountered in the present collection of texts can be summarised as follows:

1. Strong collocations will remain as a single lexeme in the process of contact.

E.g. *Mature Age Allowance*

2. Weak collocations can be split and the new collocate may not match the pattern in the original language.

E.g. *register belgesi*

3. L2 lexemes may develop new collocational as well as new colligational tendencies.

E.g. *delivery yapı*

4. Collocational failure in L1 or L2 does not result in collocational failure in C_{mx}.

E.g. *aile birleştirilmesi*

5. Verbs tend to be nominalized in the migrant setting. This results from the nature of the language rather than the contact process.

E.g. *lease yapı*

6. Delexical verbs will be employed in borrowing. They will frequently appear in nominal constructions.

E.g. *delivery yapı*

7. Co-selected L1 and L2 lexemes will acquire idiomacity and should not be analysed on the basis of slot-and-filler principle.

E.g. *tax idari*

8. At the early stages of Cmx a number of collocates will appear in the environment of "foreign material". However, frequency will determine the stronger co-selection tendencies that will acquire the idiom status.

E.g. *register kačlybelgesi*

9. Both the right and left collocates should be considered in the analysis of single lexemes.

E.g. *ɸkayetinizi Ombudsman'a.../ Ombudsman'a ɸkayet...*

10. A lemma has various word forms. Not all forms of a lemma will play an important role in language contact settings. Different forms will be associated with different grammatical patterns and with different meanings.

E.g. different forms of *yap-*

10.2 A lexicogrammatical approach to language contact

One of the main points of this thesis has been to challenge the conventional approach to language contact data and emphasise the suitability of a lexical model. This is in accordance with the recent developments regarding the nature and place of the lexicon in linguistic description (e.g. Sinclair 1991, 1987). While there are models available with a lexical claim such as Myers-Scotton's MLF model, the way these models are developed makes one wonder if they really have a base to justify this claim. The MLF model was reviewed in section 3.8. The present chapter concludes that some of the major problems of Language Contact will never be resolved if the *sentence* remains the central unit of analysis. However, if we approach language contact data from the perspective of lexical relations, the need for such a distinction disappears.

Previous approaches, while putting a great deal of emphasis on the lexical level of analysis, have not favoured a *lexical* approach. Lexical analysis has been

taxonomical and resulted in long lists that may be useful in showing what material is more likely to be borrowed than others however they have been of little theoretical use. The approach favoured here is based on the view that syntax is driven by lexis. The difference between traditional and lexical approaches to grammar has been outlined in Francis (1993):

Basically, most grammars are interested in lexis only insofar as it is useful, indeed indispensable, for the illustration of syntactic structures. The assumption is, or appears to be, that grammar and lexis are quite distinct, and that grammar is an activity which is concerned solely with the description of syntactic structure. Thus they see the clause consisting essentially of Subject, Verb, Object, Adjunct and so on, each of whose internal structure is similarly described in terms of its smaller constituents. All these constituents are seen as slots into which lexical items can be fitted, subject to certain obvious restrictions: for example, the modifier slots before the head of a nominal group are typically filled by members of the class of adjective or noun, and the qualifier slot by a prepositional phrase. The implication is that so long as you choose a word from the right class to fill the slot, then the result will be a grammatical utterance.

However, we do not communicate simply by selecting syntactic structures and independently choosing lexis to slot into them. "Instead, we have concepts to convey and communicative choices to make that require central lexical items, and these choices find themselves syntactic structures in which they can be said comfortably and grammatically" (Francis 1993).

The main point of the lexical approach then is that particular syntactic structures tend to co-occur with particular lexical items, and particular lexical items seem to occur in a limited range of structures. While specifically concerned with lexical access in the speech production of monolinguals, and does not consider the bilingual mind, Levelt's model (1989) also provides useful insights. According to Levelt, in speech production, except for the closed class items such as relative pronouns, it is the lexical selection that drives the syntactic constructions. Levelt's work has had a major impact on Language Contact researchers who have usually looked for emerging syntactic patterns in spoken data as the principal indication of

change. This was a natural consequence of seeing the lexicon and syntax as distinct systems. In the current study, the lexicon has been given a central place, and the mixed units have been analysed with the view that they reflect the usage in the immigrant setting and should therefore be considered as the starting point of an empty lexicon. It has been acknowledged over and over again that nouns occur more frequently in language contact data. The process of nominalization has been discussed in section 7.5.1 as well as in chapter 8 in relation to data analysis. In the section below, yet another aspect of nominalization is explored with respect to the mixed code. That a mixed code in an immigrant setting hardly reaches the status of 'languageness' may be explained in terms of the lacking grammatical metaphors realized by nominalization.

How do L1 and L2 items co-select each other out of a number of possibilities? This point has hardly been considered in language contact. The consensus has been that when L1 and L2 lexemes come together they usually fit into the lexicogrammatical patterns of either of the languages. Consequently their lexical environment will not be any different from that of a similar lexeme in L1 or L2. But this is not always the case and new patterns emerge in contact settings. At this point, it should be granted that a mixed code might be developing in its own right. Can we leave our preoccupation with grammaticalness aside and accept the naturalness of these patterns in immigrant settings? This thesis has argued that we should. However, Language Contact has had a methodological problem. As the importance of lexical relations is more and more emphasized in different fields, it seems possible that this methodological problem may be removed from language contact research by adopting the 'empty lexicon' approach.

Traditionally, grammars and dictionaries have taken word meanings for granted. However, Cobuild research has shown that words do not have inherent meanings, but depend on their environments to select and/or confirm their meaning. For example, if a word has two different meanings, a distinctive pattern will grow around each meaning. Consequently, the meaning will have an effect on the environment (Francis and Sinclair 1994). Lexicogrammar has major implications for linguistic theory as findings from corpora gradually challenge long-standing assumptions about lexis-syntax relations. The importance of collocations is

increasingly discussed by researchers and the criteria to define these lexical units (from completely frozen to highly variable) are being refined as it is recognized that such units are a central organizing principle in language as well as in the mental lexicon. Given the frequency of such units in language contact data, a lexicogrammatical approach such as the one proposed in this thesis, would help understand more easily those "internal operations" we are interested in.

10.3 "What have I found?"

This thesis has proposed a new way of looking at the mixed code lexicon that takes usage as the starting point. It has also discussed why the 'code' in immigrant settings has never been assigned the status of 'linguageness'.

Whether the features of the data discussed here are only characteristic of diaspora Turkish, or can be generalized to modern Turkish spoken in Turkey can be confirmed only after extensive analysis has been carried out on comparable corpora. In a similar case, that of Spanish spoken in different geographical areas of the world, Silva-Corvalán finds that "some of the changes cut across varieties; others clearly reflect" the impact of "the specific contact language" (1995b:11). It can be said that, within the limitations of this study, Turkish in Australia shows similar characteristics to Turkish in contact with European languages. This claim seems particularly valid when it comes to the use of delexical verbs (e.g. *yap-*) replacing verbal morphology. Whether such aspects can be described with reference to the participating languages or to a mixed code depends on where the researcher stands.

10.3.1 Mixed code or not?

The starting point for the mixed code preference here was the claim (e.g. Backus 1993) that in addition to the two available codes for communication, in this case English and Turkish; a mixed code has been developing in the immigrant setting. However, the methodology chosen has not followed the conversation analysis model employed for example in Auer (1998). While research conducted on the

development of mixed codes has used mainly spoken corpora, linguists have still relied on their intuitive knowledge of the language and their conviction of what is right and wrong to interpret the way the mixed code worked. In looking for the distinctive patterns in a corpus, however, "the human researcher needs to abandon his or her preconceptions about the languages concerned, and review the evidence in a detached way" (Sinclair *et al* 1996:177). Thus, in this approach, the corpus has provided the information, rather than intuition, for the description of the mixed code.

10.3.2 The place of the Lexicon in Language Contact

During the early decades, the lexicon was seen as an inventory of items in language contact and not explored any further. However, as a result of major corpus research, two revived Firthian notions, collocation and colligation, have been placed at the heart of lexicogrammar. Both notions have been used in this thesis to explain the inner dynamics of the mixed code that cannot be explained by counting the singly occurring items nicely fitting in a number of grammatical categories. Here, the new patterns emerging are placed in an empty lexicon rather than seen as an extension of L1 and L2 lexicons or grammars.

As a result of this narrow view of the lexicon three questions in relation to the loans in data have hardly been answered by the five generations of language contact researchers:

- How can "foreign material", call it switching or borrowing, have no effect on its environment?

- How can a loanword have the same meaning in the host language

or

- a different meaning in the host language without a distinctive lexical pattern growing up around the new meaning?

Within a lexicogrammatical framework, however, these questions gain central importance and the relationship between the loanword and the new language can be explained from a dynamic perspective.

10.3.3 Singly occurring switches

This thesis has argued that singly occurring switches do not really occur singly. The meeting point of the word forms from L1 and L2 is important as these forms enter into a relationship that previously was not there. The curious thing is how these items co-select each other out of a large number of probabilities? These considerations emerging from the open-minded observation of the corpus, link up with the notions of collocation and colligation, and show that what have so far seemed individual borrowings do have an impact on the lexicogrammatical patterns that develop in migrant settings. They have been treated differently from the previous discussions in relation to syntax (e.g. Poplack 1980).

10.3.4 Which elements facilitate mixing?

Some verbs such as yap- have traditionally been treated, in grammars and dictionaries, in a way that needs to be rethought. Such verbs have been delexicalized over time and are better considered grammatical rather than content words. While losing their semantic content, they have acquired a new role, at least in contact settings, that of inflecting the borrowed verb stem. Further study is needed to find other elements of this kind that facilitate mixing.

10.3.5 Borrowing vs. codeswitching

The distinction between borrowing and code-switching has been critically examined in this thesis, although not from Myers-Scotton's perspective, that is, not with reference to the status of foreign material in the mental lexicon. The notion of 3+ frequency metric as further support to the age-old view on integration has been questioned. Nominalization and metaphor have been introduced as new possible ways of looking at the question of integration. However, it has been

emphasized that integration as such is more of concern to the lexicographer than to the language contact researcher.

10.3.6 Morphological integration of words

The status of verbal vs. nominal morphology in language contact has been examined through various examples from Turkish-German data as well as Ozturk corpus. The observation that L2 verb stems usually appear in uninflected form in L1 structures as opposed to the flexibility of nominal morphology in L1+L2 patterns, leads to the tentative conclusion that verbal and nominal suffixes have a different status in the mental lexicon.

10.3.7 The question of Cultural borrowings

One of the main premises of this thesis has been a reconsideration of what is usually discarded from data analyses as cultural loans. That the role of culture-bound words should be rethought becomes evident when we come across metaphors like 'spaghetti junction'. The insight Ozturk Corpus has provided is that lexical patterns do grow around cultural words and that even simple dictionary words may acquire cultural significance in certain contexts. Also, their cultural load as such tends to disappear gradually and they become mainstream lexical items attached to patterns worth examining. Leaving out lexemes as cultural loans, is leaving out masses of valuable data.

10.3.8 The question of "psycholinguistic stress"

This question has been investigated in relation to the experience of professional translators, as the presence of psychological and social stress in migrant settings results in the need for translation. However, translation equivalent is a concept that should be questioned in relation to the question of stress. Under conditions of stress specifically developed strategies are activated and lead to the emergence of units that do not resemble the patterns of the participating languages.

Consequently, the units of translation created under these conditions lead to new units of meaning.

10.3.9 Waiting for an integrated framework: will it ever come?

This thesis has not made a deliberate attempt to integrate models from psycholinguistics, sociolinguistics, and anthropological linguistics and so on into a single framework. The claim is not that it cannot be done. However, it has always been done at the expense of the data that should be the top priority of language contact research. Instead, there has been a strong emphasis on the speaker, the setting, the conditions of migration, the status of L1 and L2 within that context, the attitudes towards the migrants and their languages, their perception of their own languages and cultural values and the list can go on. In all these, the data sets used have seemed to be playing a central role when in the actual fact they weren't. This thesis has also made reference to the setting and the speakers but what has driven the study has been the corpus rather than the other elements.

10.3.10 The nature of language contact data

Until now, the spoken text has been the key player in language contact. The spokenness of the data, which is indispensable for psycholinguistic considerations, has naturally created a certain degree of bias in the characterization of the so-called language contact phenomena, as spoken texts contain certain linguistic features that are not common in written texts. Nominalization and the use of delexical verbs as well as metaphor are typical areas where written and spoken texts do not match up. This thesis has emphasised the importance of the written sources in language contact settings and the Ozturk corpus has provided fresh insights in this respect.

On a different note, we can expect texts produced in contact settings to be centered around a number of issues that are of importance to the migrant. Therefore, the frequency of certain words appears to be higher in language contact data. Such

words include *migrant, family, services, resources* and so on. However, this claim may be justified only after corpora of similar size and nature have been compared.

10.4 Limitations of this thesis

The generalizability of findings from one corpus of Turkish to another both in the diaspora and in the mainland is a necessary next step from here. For example, "[t]he list of the fifty most common words in the Brown Corpus is replicated almost exactly in a corresponding corpus of British English, the Lancaster-Oslo/Bergen Corpus (known as the LOB Corpus). In this very limited respect, therefore, the two corpora are virtually equivalent samples. As more corpora representing different language varieties are compared, it will become evident how far a sample may be regarded as representative of the language as a whole, or of some variety of it" (Leech 1987). At present, there are no corresponding corpora of diaspora Turkish that can be used to compare my findings. And the limited Turkish corpora have been compiled for reasons other than migrant research (e.g. Hankamer's *keġ texts*, Oxford Text Archive, Bilkent) and are not readily accessible. An interesting follow up would be to compare the *Ozturk corpus* with a mainland corpus as well as with another diaspora Turkish corpus of equal size. It has been explained in Chapter 7 that the corpus consists of a considerable number of 1000 texts. However, in terms of word count, the size is limited to 250,000 and there are approximately 50,000 unique words. A future development would be the completion of the spoken component of equal size. Currently, the corpus contains no transcribed recordings.

The limitations have not only been in terms of sample size but also in terms of the local computer capacity and the tools used. For example, the acquisition of the texts in electronic form was not possible and only some 200 texts were provided on disk. Scanning newspaper texts generally of poor paper quality, proved even more difficult as some Turkish characters are not recognized by the software and the texts had to be entered manually.

The processing tools used were basic that enabled word counts and concordances, with some search flexibility provided by the database. However, a powerful analytic tool such as the LOOK UP program specifically developed for the Cobuild lexicography project would have been very useful for the retrieval of the significant collocates. Unfortunately, such programs are not commercially available.

Finally, it should be remembered that actual language use is rich, subtle and dynamic; and any abstractions or generalizations made in the process of describing the patterns of language will be limited by the time period and the conditions of the setting under investigation. Then they will break down.

10.5 Implications of this study for other research areas

This study has implications for a number of areas that have been, directly or indirectly, touched upon in this thesis. The perspectives will naturally differ, however, collaboration between these areas and Language Contact would stimulate further research.

10.5.1 Lexicography

The difference between lexicographical and language contact perspectives in relation to "lexical material" does not seem clear either in the field of lexicography or in language contact. This thesis has discussed that while the question of 'integration of loanwords' is important for both parties, it should ultimately play a more significant role in compiling dictionaries than language contact research. The insights provided by the corpus, and how it will change the lexicographic description in mono- and bilingual dictionaries of Turkish is yet another question. The assumptions so far made on a delexical verb like *yap-* for example have been challenged here. Concordances have displayed interesting patterns such as - *lik+yap-* where a suffix gains more importance as the carrier of the semantic load, much more than the preceding noun or the following delexical verb. Such an observation has never been made before. This leads to the conclusion that these

patterns will have to be accommodated within future grammatical descriptions of Turkish.

10.5.2 Grammar

Lexicogrammatical patterns of Turkish have not frequently attracted attention, yet notions such as collocation and colligation offer us a new view on grammatical descriptions. Such patterns become visible only through corpus analysis that is yet to be incorporated into mainstream Turkish research.

For example, corpus evidence leaves no doubt that the distinction between grammatical and content words gets blurred when delexical verbs are used in certain patterns and it is unhelpful to try to fit these items into well-established western-grammar-categories such as *auxiliary*. It is also common to offer translations in traditional grammars for delexical constructions such as *olmak üzere* as in 'altýsý kýz, dördü erkek *olmak üzere* , on talebem var' (I have ten pupils, six-of-them girls, four-of-them boys' (Lewis 1967:167). Lewis explains this pattern as follows:

olmak üzere, lit. 'on-the-basis-of-to-be', may sometimes be translated by 'being' or 'as being' but can often be left untranslated (Lewis 1967:168).

However, these patterns can only be described as units of meaning rather than in terms of the meanings of their individual components. Other than delexicalization, nominalization as a process has also been taken into account in this study with particular reference to the multitude of the word forms such processes can generate in an agglutinative language as opposed to inflected languages. This is also an important aspect of the Turkish grammar that requires attention. Only in NLP applications have such processes have attracted attention but mainly to see what the machine could do.

10.5.3 Mental lexicon

The challenge in the fields of language contact and bilingualism is to be able to provide an insight into the workings of the bilingual's mental lexicon. This is particularly interesting when the languages involved are not typologically related. Turkish from the perspective of the bilingual's mental lexicon has not been studied, and there has been little work on the monolingual Turkish speaker's mental processes. Unfortunately, Hankamer's (1989) early work on Turkish has not been elaborated, although it has caught the attention of language contact researchers (e.g. Myers-Scotton 1993a:62). A set of mixed examples provided by the Ozturk Corpus indicate that verbal and nominal morphology may have different representations in the mental lexicons of Turkish and English speakers. Yet, standard lexical tests carried out in bilingual settings fail to answer such questions (Schaufeli 1992) and such test design needs to be carefully reconsidered.

10.5.4 Translation

Any language contact situation offers very rich translated material, or rather any language contact situation creates an ongoing search for translation equivalents. However, translation equivalents produced in such settings hardly match those listed in bilingual dictionaries. As summarised by Teubert (1997:153)

Comparing texts and their translations step by step with the pertinent information offered by dictionaries, we find, not very surprisingly, that the main difference is the size of the translation unit. The translation units that we can identify in parallel corpora, i.e., in collections of texts and their translations, are rarely words and commonly they are phrases, anything from multi-word to collocations, idioms, extended metaphors, set phrases, and even larger units.

What do parallel corpora enable us to do apart from obtaining comparable units of translation? That is, corpora of a certain size (as quantity is important), if we are to make generalizations. According to Teubert (1997:153):

if the same translation equivalent occurs time and again for a given translation unit, we can say it manifests fairly established translation knowledge. Therefore only large parallel corpora will reveal the translation knowledge that we can apply for producing new translations. Statistic software will make sure that only 're-usable' translation equivalents are extracted from the corpus.

The implications of this view for language contact settings such as Australia where government agencies regularly put out information to migrants in a considerable number of languages are significant.

Australia's constant need for translation will hopefully lead to the construction of a "computational translation platform that combines a lexicon featuring the information contained in existing bilingual dictionaries with dynamic translation memories [...] and with simple tools to align translation equivalents in parallel corpora" (Teubert 1997:153). This would be a step towards the improvement of frequent inconsistencies encountered in such translated information packages (Kurtböke 1997).

10.5.5 Natural Language Processing

Currently, corpus research is not popular among Turkish linguists who continue to work with invented examples. In this respect, *Ozturk corpus* fills an important gap. Large collections of machine-readable Turkish texts are still to be created and the existing collections are not accessible by the research community. However, electronic texts of Turkish are important if NLP applications are to be improved:

Natural Language Processing has become the driving force behind corpus development. With it many new related areas, such as lexicon building, knowledge extraction and automatic parsing, have come into the limelight: here a corpus is used to provide training and validating techniques. Indeed, part of the motivation behind the growing interest in corpus development is the perceived importance of many practical applications such as machine translation or speech recognition (Tognini-Bonelli 1996:2).

In these areas of NLP, there has been some research (e.g. Hankamer 1989, 1993; Oflazer and Bozrahin 1994) so far.

10.5.6 Teaching

Currently, there is a tension between corpus-versus task-based approaches to language teaching, a tension that should not be there. That is, a corpus in a language teaching situation is to do with the availability of naturally occurring examples for the instant use of the teacher and the student. Until now, language course books have used invented examples, as in traditional grammars and dictionaries. A large electronic corpus simply provides a more realistic picture of the language being learnt. Task-basedness, instead, is to do with the application of corpus examples and in this respect; these two lines are complementary rather than in opposition.

The use of a Turkish corpus for the teaching of Turkish as a foreign language (TFL) is extremely useful as the provider of authentic examples. Unfortunately, most course books of TFL are based on EFL books, as are grammars and dictionaries, and display some very peculiar examples, indeed. Similarly, the teaching of Turkish to the second and third generation Turks in Australia requires such a rich source of naturally occurring language.

10.6 Language Contact 2000

The question that comes to mind at the end of such a project is: if computational corpus research so strongly challenges the traditional approaches to linguistic analysis and provide such firm evidence, why don't we see more of it in Language Contact research? It seems that the problem lies in the amount of textual, financial and human resources required for corpus creation. In addition, when these problems are solved, the researcher should know what to look for or rather what not to look for in the corpus. Tools such as concordances have been used in language contact before but to check on what the researcher thought was there (e.g. Poplack 1990).

Theoretical developments within linguistics have been putting increasing emphasis on the lexicon as it has proved to be a central source of semantic as well as syntactic information. Communication among those with an interest in the lexicon such as lexicologists, lexicographers, linguists, computational linguists, publishers, and commercial software developers has increased as they share common objectives as well as skills and knowledge. It appears that all of these professional groups work with lexicons organized differently with different contents that make it difficult or impossible to share linguistically relevant information across systems. Initial experiments show that it may be possible to construct 'neutral lexicons' that can be shared, with different theories selecting relevant linguistic information through appropriate interfaces. Where does the Mixed Code lexicon stand in relation to all these developments?

Language Contact researchers have discussed issues such as the distinction between borrowing and code-switching, nonce-borrowings and so on, as outlined above. My aim in this thesis has been to divert attention to a more challenging platform, along the lines of computerised research. A common area of interest in Natural Language Processing is the notion of "reversible grammar". The claim of NLP researchers is that a single grammar can be used for both the analysis and the generation of natural language. The results so far have been provided by theoretical linguists and computational linguists. Strzalkowski (1994a) summarises these approaches respectively. According to the former,

a reversible grammar is a specification of a language that can be used for both analysis and synthesis of well-formed expressions. Since a grammar can be viewed as a transducer between linguistic surface strings and some internal representations of meaning, reversibility means the ability to perform such translations in both directions, using in each case the same grammar. The practical aspects of this bidirectionality [...] involve completeness and soundness of the linguistic coverage, redundancy and ambiguity of representation, sharability of information and various types of preferences.

and according to the latter it is an efficient program that can interpret a given grammar both

as a parser (to recognize well-formed strings and assign them appropriate representations), as a generator (to produce appropriate surface strings given an internal representation of meaning and speaker's intentions).

Ideally, Language Contact researchers would like to develop a grammar that can be used in both directions. That is, the analysis of structures produced as a result of language contact and the prediction of such structures beforehand should be explained by one grammar. But the main problem here has been the preoccupation with well-formedness as can be seen in the words of Strzalkowski. Whether a parser will ever be developed for the analysis of mixed code grammars is an open question; however in any formalization of this kind the lexicon will play a central role. Halliday (1998) predicts that the future of linguistics in general, and Language Contact we should add, lies in two directions: computational corpus research and the mental lexicon.

APPENDIX I

Turkish Arrivals 1992-95

1992-93	1993-94	1994-95
625 (0.8)	556 (0.8)	718 (0.8)

Source: BIMPR 1995 (Table 1.13 p 22)

Permanent Departures to Turkey

1992-93	1993-94	1994-95
113 (0.4)	70 (0.3)	56 (0.2)

Source: BIMPR 1995 (Table 1.14 p 24)

Turkey-born persons in Australia by ancestral origin

Ancestry	Number	Percentage
Armenian	1402	5.7
Assyrian	214	0.9
Australian	30	0.1
Greek	1492	6.1
Kurdish	118	0.5
Turkish	19914	81.2
Other	1350	5.5
Total	24529	

Source: ABS 1986 Census microfiche

Turkish-speaking persons by birthplace

Birthplace	Number	Percentage
Australia	6073	19.4
Bulgaria	214	0.7
Cyprus	3797	12.2
Greece	336	1.1
Turkey	19488	62.4
UK and Ireland	176	0.6
Yugoslavia	192	0.6
Other	328	1.1
Not stated	624	2.0

Total 31230

source: ABS 1986 Census microfiche

Turkish nationals granted Australian citizenship

1992-93	1993-94	1994-95
2,459	1,728	1,468

Source: BIMPR 1995 (Table 2.2 p 34)

APPENDIX II

LETTER TO THE EDITOR - TURKISH

Sayın Gazete Sahibi,

Monash Üniversitesi, Dilbilim Bölümünde doktora çalışması yapmakta olan bir Türk vatandaşıyım. İngilizce konuşulan ve yazılan bir ortamda anadilimizin hangi ölçüde etkilendiğini araştırmaktayım. Avustralya'da yayınlanan Türk gazeteleri bu çalışmada gerekli veritabanı için çok değerli bir kaynak oluşturuyorlar. Uzun süre Avustralya'da yaşamış gazetecilerin yazdığı yazılarda TAB, Medicare, Upper House vb gibi kültür bağımı sözcüklerin ve kavramların nasıl ortaya çıktığı, İngilizce'nin Türkçe tümceleri nasıl etkilediği vb. gibi sorulara cevap aramak amacındayım.

Çalışmamın ikinci aşamasında, gözlemlerimi daha geniş bir çerçeveye oturtabilmek için, makale yazarlarından dil kullanıcılarıyla ilgili 15 soru içeren bir anket yanıtlamalarını isteyeceğim.

Veritabanını oluşturacak makaleleri bilgisayara yükleyerek ileri bilgisayar programlarıyla inceleyecek ve araştırmayı 1998 yılı içerisinde sonuçlandıracağım.

Çalışmamda yararlanacağım, gazetenizde basılmış makaleleri, bu araştırmanın anadilimizin korunmasına önemli bir katkısı olacağı göz önünde bulundurarak, kullanmama izin vereceğinize ve yardımınız esirgemeyeceğinize inanıyorum. Ayrıca, gazetenizde yayınlanmış yazıları araştırmamın amacı dışında kullanılmıyacağına emin olursanız aşırıda adı geçen komisyona başvurma hakkınız olduğunu hatırlatmak isterim.

The Secretary
Standing Committee on Ethics and Research on Humans
Research Services Division
Monash University
Clayton 3168

tel: 905 2052 Fax: 905 5342

Saygılarımla,

Petek Kurtböke

LETTER TO THE EDITOR - ENGLISH

Dear Editor,

I am a native speaker of Turkish, conducting Ph.D research at Linguistics Department, Monash University. I am investigating to what extent our mother-tongue is affected in an English-speaking and English-writing environment. Turkish newspapers printed in Australia are an invaluable source for the database which is essential for this study. My aim is to answer questions such as how culture-dependent words and concepts as TAB, Medicare, Upper House etc. appear in text and how Turkish sentence structure changes under the influence of English, by looking at articles written by journalists who have lived in Australia for a long time. After collecting the illustrative articles I will enter these into the computer and use specific computer tools to analyse the content. I expect my research to be finalised in 1998.

In the second stage of my study, in order to obtain a broader view, I will also ask you and the journalists involved, to answer a questionnaire about their language use and writing process.

It is my hope that you will allow me to use articles from your newspaper bearing in mind that this study will make a significant contribution to the maintenance of our language. I would also like to remind you that you have the right to apply to the committee whose details are given below should you have any doubts regarding the appropriate use of the material from your newspaper.

The Secretary
Standing Committee on Ethics and Research on Humans
Research Services Division
Monash University
Clayton 3168

tel: 905 2052 Fax: 905 5342

Yours sincerely,

Petek Kurtböke

APPENDIX III

S O R U L A R

1. Gazeteniz kaç yıldır/aydır çıkıyor?
2. Gazete çıkarma kararınızda hangi etkenler rol oynadı?
3. Türkiye'de gazetecilik deneyiminiz var mıydı?
4. Gazetenin kadrosunda bu güne kadar değişiklik oldu mu?
5. Gazetenizin yayın yaşamını nasıl değerlendiriyorsunuz? (başarılı/başarısız/ne başarılı ne başarısız) Neden?
6. Yazılar baskıya girmeden önce nasıl bir işlem görüyor?
7. Gazetenizin geleceğini nasıl görüyorsunuz?
8. Gazete çıkarmadaki amacınız neydi? / nedir?
 - a. haber verme/bilgilendirme
 - b. ticari
 - c. Avustralya'da yaşayan Türklerin dillerini unutmaması
 - d. diğer (lütfen açıklayınız)
9. Avustralya'da 25 yıllık geçmişlerine karşın Türk göçmenlerin günlük hayatta hala İngilizce'den çok Türkçe kullanıyor olmasında gazetelerin payı nedir?
10. Gazetenizin okurları kimler?
11. Türk göçmen sayısının diğer gruplara oranla düşük olmasına karşın çıkarılan gazete sayısı yüksek. Bunun nedenlerini nasıl açıklıyorsunuz?
12. Kaç yıldır Avustralya'da oturuyorsunuz?
13. Günde kaç saat Türkçe kaç saat İngilizce konuşuyorsunuz?
14. İngilizce seviyenizi nasıl değerlendiriyorsunuz?
 - a. çok iyi
 - b. iyi
 - c. orta
 - d. zayıf
15. Eğitim seviyeniz
 - a. ilk
 - b. orta
 - c. lise
 - d. Üniversite
16. Gazeteniz kaç günde bir çıkıyordu ve kaç adet basılıyordu?

QUESTIONNAIRE

1. When did you start publishing?
2. Which factors determined your decision to publish a newspaper?
3. Did you work as a journalist in Turkey?
4. Have there been any staff changes over the years?
5. How would you rate your newspaper in terms of success?
6. What sort of a process do the articles go through before they are printed?
7. How do you see the future of your newspaper?
8. What is/was your aim in publishing a newspaper?
 - a. to inform
 - b. commercial
 - c. maintenance of Turkish
 - d. other (please explain)
9. In spite of their 25 years of presence in Australia, Turkish still plays an important role in the migrants' everyday life. To what extent do the newspapers contribute to the maintenance of Turkish?
10. Who are your newspaper's readers?
11. Turkish migrants, although fewer in numbers compared to some ethnic groups, enjoy a higher number of newspapers. How do you explain this?
12. How long have you been in Australia?
13. How many hours a day do you speak Turkish , and how many hours English?
14. How do you rate your level of English?
 - a. very good
 - b. good
 - c. average
 - d. poor
15. What is your educational background?
 - a. primary school
 - b. secondary school
 - c. high school
 - d. University
16. What is the frequency of publication and the number of copies distributed?

APPENDIX I

Turkish Arrivals 1992-95

1992-93	1993-94	1994-95
625 (0.8)	556 (0.8)	718 (0.8)

Source: BIMPR 1995 (Table 1.13 p 22)

Permanent Departures to Turkey

1992-93	1993-94	1994-95
113 (0.4)	70 (0.3)	56 (0.2)

Source: BIMPR 1995 (Table 1.14 p 24)

Turkey-born persons in Australia by ancestral origin

Ancestry	Number	Percentage
Armenian	1402	5.7
Assyrian	214	0.9
Australian	30	0.1
Greek	1492	6.1
Kurdish	118	0.5
Turkish	19914	81.2
Other	1350	5.5
Total	24529	

Source: ABS 1986 Census microfiche

Turkish-speaking persons by birthplace

Birthplace	Number	Percentage
Australia	6073	19.4
Bulgaria	214	0.7
Cyprus	3797	12.2
Greece	336	1.1
Turkey	19488	62.4
UK and Ireland	176	0.6
Yugoslavia	192	0.6
Other	328	1.1
Not stated	624	2.0

Total 31230

source: ABS 1986 Census microfiche

Turkish nationals granted Australian citizenship

1992-93	1993-94	1994-95
2,459	1,728	1,468

Source: BIMPR 1995 (Table 2.2 p 34)

APPENDIX II

LETTER TO THE EDITOR - TURKISH

Sayın Gazete Sahibi,

Monash Üniversitesi, Dilbilim Bölümünde doktora çalışması yapmakta olan bir Türk vatandaşıyım. İngilizce konuşulan ve yazılan bir ortamda anadilimizin hangi ölçüde etkilendiğini araştırmaktayım. Avustralya'da yayınlanan Türk gazeteleri bu çalışmada gerekli veritabanı için çok değerli bir kaynak oluşturuyorlar. Uzun süre Avustralya'da yaşamış gazetecilerin yazdığı yazılarda TAB, Medicare, Upper House vb gibi kültür bağımı sözcüklerin ve kavramların nasıl ortaya çıktığı, İngilizce'nin Türkçe tümceleri nasıl etkilediği vb. gibi sorulara cevap aramak amacındayım.

Çalışmamın ikinci aşamasında, gözlemlerimi daha geniş bir çerçeveye oturtabilmek için, makale yazarlarından dil kullanıcılarıyla ilgili 15 soru içeren bir anket yanıtlamalarını isteyeceğim.

Veritabanını oluşturacak makaleleri bilgisayara yükleyerek ileri bilgisayar programlarıyla inceleyecek ve araştırmayı 1998 yılı içerisinde sonuçlandıracağım.

Çalışmamda yararlanacağım, gazetenizde basılmış makaleleri, bu araştırmanın anadilimizin korunmasına önemli bir katkısı olacağı göz önünde bulundurarak, kullanmama izin vereceğinize ve yardımınız esirgemeyeceğinize inanıyorum. Ayrıca, gazetenizde yayınlanmış yazıları araştırmamın amacı dışında kullanılmıyacağına emin olursanız aşırıda adı geçen komisyona başvurma hakkınız olduğunu hatırlatmak isterim.

The Secretary
Standing Committee on Ethics and Research on Humans
Research Services Division
Monash University
Clayton 3168

tel: 905 2052 Fax: 905 5342

Saygılarımla,

Petek Kurtböke

LETTER TO THE EDITOR - ENGLISH

Dear Editor,

I am a native speaker of Turkish, conducting Ph.D research at Linguistics Department, Monash University. I am investigating to what extent our mother-tongue is affected in an English-speaking and English-writing environment. Turkish newspapers printed in Australia are an invaluable source for the database which is essential for this study. My aim is to answer questions such as how culture-dependent words and concepts as TAB, Medicare, Upper House etc. appear in text and how Turkish sentence structure changes under the influence of English, by looking at articles written by journalists who have lived in Australia for a long time. After collecting the illustrative articles I will enter these into the computer and use specific computer tools to analyse the content. I expect my research to be finalised in 1998.

In the second stage of my study, in order to obtain a broader view, I will also ask you and the journalists involved, to answer a questionnaire about their language use and writing process.

It is my hope that you will allow me to use articles from your newspaper bearing in mind that this study will make a significant contribution to the maintenance of our language. I would also like to remind you that you have the right to apply to the committee whose details are given below should you have any doubts regarding the appropriate use of the material from your newspaper.

The Secretary
Standing Committee on Ethics and Research on Humans
Research Services Division
Monash University
Clayton 3168

tel: 905 2052 Fax: 905 5342

Yours sincerely,

Petek Kurtböke

APPENDIX III

SORULAR

1. Gazeteniz kaç yıldır/aydır çıkıyor?
2. Gazete çıkarma kararınızda hangi etkenler rol oynadı?
3. Türkiye'de gazetecilik deneyiminiz var mıydı?
4. Gazetenin kadrosunda bu güne kadar değişiklik oldu mu?
5. Gazetenizin yayın yapımını nasıl değerlendiriyorsunuz? (başarılı/başarısız/ne başarılı ne başarısız) Neden?
6. Yazılar baskıya girmeden önce nasıl bir işlem görüyor?
7. Gazetenizin geleceğini nasıl görüyorsunuz?
8. Gazete çıkarmadaki amacınız neydi? / nedir?
 - a. haber verme/bilgilendirme
 - b. ticari
 - c. Avustralya'da yaşayan Türklerin dillerini unutmaması
 - d. diğer (lütfen açıklayınız)
9. Avustralya'da 25 yıllık geçmişlerine karşın Türk göçmenlerin günlük hayatta hala İngilizce'den çok Türkçe kullanıyor olmasında gazetelerin payı nedir?
10. Gazetenizin okurları kimler?
11. Türk göçmen sayısının diğer gruplara oranla düşük olmasına karşın çıkarılan gazete sayısı yüksek. Bunun nedenlerini nasıl açıklıyorsunuz?
12. Kaç yıldır Avustralya'da oturuyorsunuz?
13. Günde kaç saat Türkçe kaç saat İngilizce konuşuyorsunuz?
14. İngilizce seviyenizi nasıl değerlendiriyorsunuz?
 - a. çok iyi
 - b. iyi
 - c. orta
 - d. zayıf
15. Eğitim seviyeniz
 - a. ilk
 - b. orta
 - c. lise
 - d. Üniversite
16. Gazeteniz kaç günde bir çıkıyordu ve kaç adet basılıyordu?

QUESTIONNAIRE

1. When did you start publishing?
2. Which factors determined your decision to publish a newspaper?
3. Did you work as a journalist in Turkey?
4. Have there been any staff changes over the years?
5. How would you rate your newspaper in terms of success?
6. What sort of a process do the articles go through before they are printed?
7. How do you see the future of your newspaper?
8. What is/was your aim in publishing a newspaper?
 - a. to inform
 - b. commercial
 - c. maintenance of Turkish
 - d. other (please explain)
9. In spite of their 25 years of presence in Australia, Turkish still plays an important role in the migrants' everyday life. To what extent do the newspapers contribute to the maintenance of Turkish?
10. Who are your newspaper's readers?
11. Turkish migrants, although fewer in numbers compared to some ethnic groups, enjoy a higher number of newspapers. How do you explain this?
12. How long have you been in Australia?
13. How many hours a day do you speak Turkish , and how many hours English?
14. How do you rate your level of English?
 - a. very good
 - b. good
 - c. average
 - d. poor
15. What is your educational background?
 - a. primary school
 - b. secondary school
 - c. high school
 - d. University
16. What is the frequency of publication and the number of copies distributed?

APPENDIX IV

SAMPLE TEXT FROM OZTURK CORPUS

\ref dün024
\code 001Aa
\y 3

\tx Vatandaşlık yemini deđiyor.

Federal Hükümet Göç ve Etnik İşler Bakanı Senatör Nick Bolkus, 1994 yılı Ocak ayından itibaren vatandaşlıkta geçecek olan kimselerin yeni kabul edilen yemini ederek vatandaşlıkta geçeceklerini bildirdi. Senatör Bolkus Meclisten geçen yeni kanunla bu yeminin Avustralya'yı ve Avustralyalıların inancını daha iyi temsil edeceğini belirterek şöyle dedi: "Tamamen çokkültürlü bir toplum olan toplumumuzda bu yemin insanların daha anlamlı olacaktır."

Ayrıca bu yeminin vatandaşlık yemini dışında diğer resmi kuruluşlarda da kullanılacağını ve kendimizi daha iyi tanımamıza ve milli birliğimizin pekişmesine yardımcı olacağını belirtti.

Yeni kabul edilen yemin şöyle: "From this time forward, under God, I pledge my loyalty to Australia and its people whose democratic beliefs I share, whose rights and liberties I respect, and whose laws I will uphold and obey."

Yemin içinde geçen Allah adının isteyenler tarafından kullanılmayabileceğini belirten Senatör Bolkus önümüzdeki yıl vatandaşlık yemini için çağrılacak müracaatçıların en yakın Göçmen Bakanlığı ofisleriyle temas ederek daha fazla bilgi alabilirler.

APPENDIX V

SAMPLE WORDLIST FROM OZTURK CORPUS

\id WOMAN.WLS from source: WOMAN.TXT Wed Sep 30 18:08:17 1998

\w ve
\c 000290
\w bir
\c 000204
\w ÝĐ
\c 000188
\w veya
\c 000124
\w ÝÇÝN
\c 000113
\w bu
\c 000084
\w eđitim
\c 000079
\w Çalyđma
\c 000066
\w ile
\c 000065
\w Federal
\c 000051
\w ilgili
\c 000042
\w zaman
\c 000040
\w daha
\c 000039
\w size
\c 000038
\w bulma
\c 000036
\w iđe
\c 000035
\w iđer
\c 000034
\w telefon
\c 000034
\w yardđm
\c 000034
\w toplu
\c 000033
\w Çocuk
\c 000032

- Aarts, J and W Meijs 1984** *Corpus Linguistics*. Rodopi, Amsterdam.
- Aarts, J and T Van den Heuvel 1984** Linguistics and Computational Aspects of Corpus Research. In Aarts and Meijs, pp83-94.
- Ata A W and C Ryan (eds) 1989** *The Ethnic Press in Australia*. Academia Press and Footprint Publications, Melbourne.
- ABS 1992** *Access and Equity. Awareness, Perceptions and Developments In the City of Brunswick*. Statistical Consultancy, Melbourne.
- Afendras, E (ed) 1980** *Patterns of Bilingualism*. SEAMEO Regional Language Centre, Singapore University Press.
- Aijmer, K 1987** Oh and Ah in English Conversation. In Meijs, pp61-86.
- Akçelik, R and J Elley (ed) 1988** *Turkish Community in Australia*. Australian-Turkish Friendship Society Publications 3, Melbourne.
- Akçelik, R (ed) 1993** *Turkish Youth in Australia*. Australian-Turkish Friendship Society Publications 4, Melbourne.
- Akçelik, S and A Pappas 1988** Culture and Communication issues and the Turkish Community. In Akçelik and Elley, pp229-242.
- Alatis, J E (ed) 1970** *Bilingualism and Language Contact: Anthropological, Linguistic, Psychological and Sociological Aspects*. Monograph Series in Languages and Linguistics 23. Georgetown University Press, Washington DC.
- Alderson A D and F Ýz 1959** *The Concise Oxford Turkish-English/English-Turkish Dictionary*.
- Alderson A D and F Ýz 1978** *The Oxford English-Turkish Dictionary*. 2nd edition.
- Alderson A D and F Ýz 1984** *The Oxford Turkish-English Dictionary*. 3rd edition.
- Ali, A M 1991** The Turkish Speech Community. In Alladina and Edwards, pp202-213.
- Alladina, S and V Edwards (eds) 1991** *Multilingualism in the British Isles*. Longman, London.
- Allen, R F 1989** Survey of Natural Language Corpora in Computational Linguistics. In Batori et al., pp150-159.
- Allen, S 1992** Opening Address. In Svartvik, pp1-3.
- Altenberg, B 1987** Predicting Text Segmentation into Tone Units. In Meijs, pp49-60.
- Ammon, U., N Dittmar and K Mattheier (eds) 1987** *Sociolinguistics*. Walter de Gruyter, Berlin. 2 vols.
- Andersen, H 1989** Understanding linguistic innovations. In Breivik and Jahr, pp5-27.
- Apitzsch, G and N Dittmar 1987** Contact between German and Turkish adolescents: A case study. In Knapp et al., pp51-72.
- Appel, R and P Muysken 1987** *Language Contact and Bilingualism*. Edward Arnold, London.
- Armstrong, S (ed) 1994** *Using Large Corpora*. The MIT Press, Cambridge, Mass.
- Armstrong-Warwick, S 1994** Preface. In Armstrong, ppvii-viii.
- Atkins, S., J Clear and N Ostler 1992** Corpus Design Criteria. *Literary and Linguistic Computing* 7:1 pp1-16.
- Atkins, B.T.S. and A Zampolli (eds) 1994** *Computational Approaches to the Lexicon*. Oxford University Press, Oxford.
- Atkins, B.T.S., B Levin and A Zampolli 1994** Computational Approaches to the Lexicon: An Overview. In Atkins and Zampolli, pp17-45.
- Auer, P 1997** Roundtable Discussion. *First International Symposium on Bilingualism*. University of Vigo, Spain, 21-25 October.
- Auer, P (ed) 1998** *Code-Switching in Conversation*. Routledge, London.
- Backus, A 1992** *Patterns of Language Mixing: A Study in Turkish-Dutch Bilingualism*. Otto Harrassowitz, Wiesbaden.
- Backus, A 1993a** Turkish-Dutch code-switching and the Frame-Process Model. In Extra and Verhoeven (1993c), pp223-235.
- Backus, A 1993b** Bilingual norms and linguistic change. Paper presented at 22nd Annual Conference on New Ways of Analyzing Variation, Ottawa, Canada.

- Backus, A 1996** *Two in one. Bilingual Speech of Turkish Immigrants in the Netherlands*. Tilburg University Press.
- Backus, A and H Van der Heijden 1994** Life and Birth of A Bilingual: The Mixed Code of Bilingual Children and Adults in the Turkish Community in the Netherlands. Paper presented at 7th International Turkish Linguistics Conference, Mainz, Germany.
- Baetens Beardsmore, H 1982** *Bilingualism: basic principles*. (second edition 1986). Multilingual Matters Ltd, Clevedon.
- Baetens Beardsmore, H 1983** Interpreting the Scientific Literature on Bilingualism. In Nelde (1983b), pp19-34.
- Baetens Beardsmore, H 1990** Synthesis of the Workshop on Concepts, Methodology and Data. European Scientific Network on Code-Switching and Language Contact. Papers for the workshop on concepts, methodology and data. Basel meeting. pp1-2.
- Baker, M 1992** *In Other Words A Coursebook on translation*. Routledge, London.
- Baker, M 1993** Corpus Linguistics and Translation Studies: Implications and Applications. In Baker et al., pp233-250.
- Baker, M., G Francis and E Tognini-Bonelli (eds) 1993** *Text and Technology*. In Honour of John Sinclair. John Benjamins, Amsterdam.
- Bates M and R M Weischedel 1993** *Challenges in Natural Language Processing*. Cambridge University Press, Cambridge.
- Barnett, J., I Mani and E Rich 1994** Reversible Machine Translation: What to do when the languages don't match up. In Strzalkowski, pp321-364.
- Baş arýn, H H and V Baş arýn 1993** *The Turks in Australia*. Turquoise Publications, Melbourne.
- Batori, I S et al (eds) 1989** *Computational Linguistics. An International Handbook on Computer Oriented Language Research and Applications*. Walter de Gruyter, Berlin.
- Batrouney, T 1995** The Kurds in Australia (Interview). BIMPR Bulletin 13 p37.
- Beach, W A., Fox, S E and S Philosoph (eds) 1977** *Papers from the thirteenth regional meeting*. Chicago Linguistic Society, Chicago.
- Bednall, D 1992** *Media and Immigrant Settlement*. BIMPR. AGPS., Canberra.
- Bennett, P A et al 1986** *Multilingual Aspects of Information Technology*. Gower Publishing Company, England.
- Benson, M 1989** The Collocational Dictionary and the Advanced Learner. In Tikoo, pp84-92.
- Bernstein, B 1987** Social Class, Codes and Communication. In Ammon et al. pp563-579.
- Biber, D 1989** A Typology of English texts. *Linguistics* 27 pp3-43.
- Biber, D 1990** Methodological Issues Regarding corpus-based Analyses of Linguistic variation. *Literary and Linguistic Computing* 5:4 pp257-269.
- Biber, D 1992** On the Complexity of Discourse Complexity: A Multidimensional Analysis. *Discourse Processes* 15 pp133-163.
- Biber, D 1993** Representativeness in Corpus Design. *Literary and Linguistic Computing* 8:4 pp243-257.
- Biber, D 1994** *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge University Press, Cambridge.
- Biber, D and E Finegan 1994** Intra-textual variation within medical research articles. In Oostdijk and de Haan, pp
- Biber, D., S Conrad and R Reppen 1994** Corpus-based Approaches to Issues in Applied Linguistics. *Applied Linguistics* 15:2 pp169-189.
- BIMPR Immigration Update 1995** June Quarter. Australian Government Publishing Service, Canberra.
- BIMPR Community Profiles 1991 Census Turkey Born**. Australian Government Publishing Service, Canberra.
- Blom, J and J Gumperz 1972** Social Meaning in Linguistic Structure: Code-Switching in Norway. In Gumperz and Hymes, pp407-434.
- Bloomfield, L. 1933** *Language*. Holt Rinehart and Winston, New York.
- Boeschoten, H 1990** Asymmetrical code-switching in Immigrant Communities. In *European Scientific Network on Code-Switching and Language Contact. London meeting*. pp85-100.

- Boeschoten, H 1997** The Turkish Language Reform forced into stagnation. In Clyne, pp357-383.
- Boeschoten, H and L Verhoeven (eds) 1991** *Turkish Linguistics Today*. E J Brill, Leiden.
- Bolinger, D 1971** Introduction. In Di Pietro, ppvii-ix.
- Bosi, P 1980** Ethnic Press - A question of survival. *Media Information Australia*. **15** pp34-37.
- Borg, V 1998** Plenary Speech. Multicultural Aged Care Seminar. Richmond City Hall, Melbourne.
- Boyd, S 1985** *Language Survival A Study of Language Contact, Language Shift and Language Choice in Sweden*. Ph.D Thesis. University of Göteborg.
- Braunmüller, K 1997** Communication strategies in the area of the Hanseatic League: The approach by semi-communication. *Multilingua* **16:4** pp365-373.
- Breivik, L E and E H Jahr (eds) 1989** *Language Change Contributions to the Study of Its Causes*. Mouton de Gruyter, Berlin.
- Buseman, A et al 1996** *The Linguist's Shoebox for Windows and Macintosh*. Summer Institute of Linguistics, Waxhaw, North Carolina. Version 4.
- Caroli, F 1995** Types of Lexical Co-occurrences: Descriptive Parameters. In Steffens, pp203-218.
- Carter R and M McCarthy 1988** *Vocabulary and Language Teaching*. Longman, London.
- CETH Newsletter 1995** *Taking the Myth out of OCR*. Centre for Electronic Texts in the Humanities, New Brunswick, NJ. 3:1 pp2-3.
- Church, K and R Mercer 1994** Introduction to the Special Issue on Computational Linguistics Using large Corpora. In Armstrong, pp1-24.
- Christ, O 1996** *Corpus Exploration Tools for Lexicography. Tutorial script. EURALEX 96*, University of Göteborg, Sweden.
- Clear, J 1993** From Firth Principles - Computational Tools for the study. In Baker et al., pp271-292.
- Clear, J. et al 1996** COBUILD, The State of the Art. *International Journal of Corpus Linguistics* **1:2** pp303-314.
- Clyne, M 1967** *Transference and Triggering*. Martinus Nijhoff, The Hague.
- Clyne, M 1987** Constraints on code switching: how universal are they? *Linguistics* **25** pp739-764.
- Clyne, M 1991** *Community Languages The Australian Experience*. Cambridge University Press, Cambridge.
- Clyne, M (ed) 1997** *Undoing and redoing corpus planning*. Mouton de Gruyter, Berlin.
- Collins, P and P Peters 1988** The Australian Corpus Project. In Kytö et al. pp103-120.
- Cooper, R 1967** How can we measure the roles which a bilingual's languages play in his everyday behaviour? In Kelly, pp192-208.
- Costin, G and J Beadle 1995** *Text Analysis*. Summer Institute of Linguistics, Waxhaw, North Carolina.
- Crowdy, S 1993** Spoken Corpus Design. *Literary and Linguistic Computing* **8:4** pp259-265.
- Culler, J 1987** Towards a linguistics of writing. In Fabb et al., pp173-184.
- Cuocolo, R 1998** A Note from the Director. *Teatro* presented by IRAA Theatre and Canto Coro. Theatre program. Melbourne Festival.
- Cüceloğlu, D and D Slobin 1980** Effects of Turkish Language Reform on Person Perception. *Journal of Cross-Cultural Psychology* **11:3** pp297-326.
- Çevik, G 1996** Power or participation: Status of Turkish women. *BIMPR Bulletin* **17** p27.

- Dabéne, L and M Moore 1995** Bilingual Speech of Migrant People. In Milroy and Muysken, pp17-44.
- Danesi, M 1983** Contrastive Analysis Revisited. In Di Pietro et al., pp215-221.
- Deasey, M., R Mapolar and J Wheeler 1988** The Turkish elderly in the western suburbs of Melbourne: meeting their needs. In Akçelik and Elley, pp131-158.
- De Beaugrande, R 1991** *Linguistic Theory: the discourse of fundamental works*. Longman, London.
- De Beaugrande, R 1997** Theory and Practice in Applied Linguistics: Disconnection, Conflict, or Dialectic? *Applied Linguistics* 18:3 pp279-313.
- De Bot, K and M Clyne 1994** A 16-Year Longitudinal Study of Language Attrition in Dutch Immigrants in Australia. *Journal of Multilingual and Multicultural Development*. 15:1 pp17-28.
- De Bot, K and T Huebner 1992** Preface. In Fase et al., ppv-vii.
- De Haan, P 1987** Exploring the Linguistic Database: Noun Phrase Complexity and Language Variation. In Meijs, pp151-165.
- Demircan, Ö and A Erözden 1991** Özleştirme Kaynakçası. Dilbilim ve Türkçe. Dil Derneği Yayınları 3 pp138-176.
- De Ruiter, J 1991** *Talen in Nederland. Een beschrijving van de taalsituatie van negen etnische groepen*. Wolters-Noordhoff, Groningen.
- Delisle, J 1988** *Translation: An Interpretive Approach*. University of Ottawa Press, Ottawa.
- Di Pietro, R 1970** The Discovery of Universals in Multilingualism. In Alatis, pp13-23.
- Di Pietro, R 1971** *Language Structures in Contrast*. Newbury House Publishers, Rowley, Mass.
- Di Pietro, R., W Frawley, and A Wedel (eds) 1983** *The First Delaware Symposium on Language Studies (1979)*. University of Delaware Press, Newark.
- Di Sciullo, A M and E Williams 1987** *On the Definition of Word*. MIT Press, Mass.
- Di Sciullo, A M, P Muysken and J Singh 1986** Government and code-switching. *Journal of Linguistics* 22 pp1-24.
- Diebold, R 1961** Incipient Bilingualism. *Language* 37 pp96-112.
- Diebold, R 1962** A Laboratory for Language Contact. *Anthropological Linguistics* 4 pp41-51.
- Dil, A S 1971** *Language Structure and Language Use. Essays by Charles A Ferguson*. Stanford University Press, Stanford, California.
- Dittmar, N 1976** *Sociolinguistics. A Critical survey of theory and application*. Edward Arnold, London.
- Doğançay-Aktuna, S 1995** An Evaluation of Turkish Language Reform After 60 Years. *Language Problems and Language Planning* 19:3 pp221-249.
- Domingue, N 1990** Bi- and multilingualism: code-switching, interference and hybrids. In Polomé, pp527-534.
- Dorian, N 1981** *Language Death-The Life Cycle of a Scottish Gaelic Dialect*. University of Pennsylvania Press, Philadelphia.
- Dorian, N C (ed) 1989** *Investigating Obsolescence: Studies in Language Contraction and Death*. Cambridge University Press, New York.
- Doueïhi, J 1980** Ethnic Press - Its Role and Influence. *Media Information Australia* 15 pp38-40.
- Dowty, D., L Karttunen and A Zwicky (eds) 1984** *Natural Language Processing: Psycholinguistic, Computational and Theoretical Perspectives*. Cambridge University Press, Cambridge.
- Driessen, G 1992** First and Second Language Proficiency: Prospects for Turkish and Moroccan Children in the Netherlands. *Language, Culture and Curriculum* 5:1 pp23-40.
- Duranti, A 1997** *Roundtable Discussion. First International Symposium on Bilingualism*. University of Vigo, Spain, 21-25 October.

- Eastman, M C 1992a** Codeswitching as an Urban Language-Contact Phenomenon. In Eastman, pp1-17.
- Eastman M C (ed) 1992b** *Codeswitching*. Multilingual Matters, Clevedon.
- Elley, J 1985** *Guestworkers or Settlers? Turkish Migrants in Melbourne*. Ph.D Thesis. Monash University, Melbourne.
- Elley, J 1988** Çalışan hanýmlar: Turkish Migrant women at work. In Akçelik and Elley, pp79-99.
- Elley, J 1993** "I see myself as Australian-Turkish": the identity of second-generation Turkish migrants in Australia. In Akçelik, pp55-76.
- Engwall, G 1994** Not Chance But Choice: Criteria in Corpus Creation. In Atkins and Zampolli, pp49-82.
- Eppert, F (ed) 1982** *Transfer and Translation in Language Learning and Teaching*. SEAMEO Regional Language Centre, Singapore University Press.
- Ersöz, T 1995** Türk Dil Devriminin Günümüz Türkçesine Yansıması. *adapTürk Dili* 93 pp29-34.
- Ervin-Tripp, S 1969** Commentary. In Kelly, pp26-35.
- Ervin-Tripp, S 1970** Structure and Process in Language Acquisition. In Alatis, pp313-344.
- Ervin-Tripp, S 1972** On Sociolinguistic Rules: Alternation and Co-occurrence. In Gumperz and Hymes, pp213 -250.
- Ervin, S and C Osgood 1954** Second language learning and bilingualism. *Journal of Abnormal and Social Psychology Supplement* pp139-46.
- Extra, G 1993** Language acquisition, shift, and loss of immigrant minority groups in Europe. In Kettemann and Wieden, pp361-377.
- Extra, G and L Verhoeven (eds) 1993a** *Community Languages in the Netherlands*. Swets & Zeitlinger, Amsterdam.
- Extra, G and L Verhoeven 1993b** Community languages in cross-cultural perspective. In Extra and Verhoeven (1993a), pp1-28.
- Extra, G and L Verhoeven (eds) 1993c** *Immigrant Languages in Europe*. Multilingual Matters, Clevedon.
- Fabb, N et al. (eds) 1987** *The Linguistics of Writing*. Manchester University Press, Manchester.
- Fase W, K Jaspaert and S Kroon 1992** Maintenance and Loss of Minority Languages: Introductory Remarks. In Fase et al., pp3-13.
- Fase W, K Jaspaert and S Kroon (eds) 1992** *Maintenance and Loss of Minority Languages*. John Benjamins, Amsterdam.
- Ferguson, C 1959** Diglossia. *Word* 15 pp325-340.
- Ferguson, C 1966a** Sentence Deviance in Linguistics and Language Teaching. Reprinted in Dil, pp185-196.
- Ferguson, C 1966b** Applied Linguistics. Reprinted in Dil, pp135-147.
- Ferguson, C 1966c** On Sociolinguistically Oriented Language Surveys. Reprinted in Dil, pp149-156.
- Ferguson, C 1968** Language Development. Reprinted in Dil, pp219-232
- Fernandez, M 1993** *Diglossia. A Comprehensive Bibliography 1960-1990*. John Benjamins, Amsterdam.
- Ferrara, K., B Brown, K Walters and J Baugh 1988** *Linguistic Change and Contact. Proceedings of the XVI Annual Conference on New Ways of Analyzing Variation*. The University of Texas at Austin.
- Firchow, E., K Grimstad, N Hasselmo and W A O'Neill (eds) 1972** *Studies for Einar Haugen*. Mouton, The Hague.
- Firth, J R 1957** *Papers in Linguistics*. Oxford University Press, London.
- Fishman, J (ed) 1966a** *Language Loyalty in the US*. Mouton & Co. The Hague.
- Fishman, J 1966b** Language Maintenance and Language Shift as a Field of Inquiry. In Fishman, pp424-458.
- Fishman, J 1967** Bilingualism with and without Diglossia; Diglossia with and without bilingualism. *Journal of Social Issues*. 23:2 pp29-38.

- Fishman, J (ed) 1968a** *Readings in the Sociology of Language*. Mouton Publishers, The Hague.
- Fishman, J 1968b** Sociolinguistic Perspective on the Study of Bilingualism. *Linguistics* 39 pp21-49.
- Fishman, J 1972a** Problems and Prospects of the Sociology of Language. In Firchow et al., pp214-226.
- Fishman, J 1972b** Domains and the relationship between micro- and macro-sociolinguistics. In Gumperz and Hymes, pp435-453.
- Fishman, J 1972c** *Advances in the Study of Societal Multilingualism*. Mouton, The Hague.
- Fishman, J 1989** *Language and Ethnicity in Minority Sociolinguistic Perspective*. Multilingual Matters, Clevedon Philadelphia.
- Fishman, J et al (eds) 1986** *The Fergusonian Impact. In Honour of Charles A Ferguson*. 2 vols. Mouton de Gruyter. Berlin.
- Fishman, J., R Cooper, R Ma (eds) 1971** *Bilingualism in the Barrio*. Bloomington's-Gravenhage.
- Fisiak, J (ed) 1995** *Linguistic Change under Contact Conditions*. Mouton de Gruyter, Berlin.
- Fontenelle, T et al. (eds) 1998** *EURALEX '98 Proceedings*. University of Liège, Belgium. 2 Vols.
- Foott, D 1993** *Scanner technology and the addition of contents page to library records Monash University*. Department of Computer Technology. Report No. 92-06.
- Francis, G 1991** Nominal group heads and clause structure. *Word* 42:2 pp45-156.
- Francis, G 1993** A Corpus-Driven Approach to Grammar. In Baker et al, pp138-156.
- Francis, G and J Sinclair 1994** "I bet he drinks Carling Black Label": a risposte to Charles Owen on corpus grammar. *Applied Linguistics* 15:2 pp190-200.
- Francis, W N 1992a** Language corpora B.C. In Svartvik, pp17-32.
- Francis, W N 1992b** Review of Svartvik, J (ed) *The London-Lund Corpus of Spoken English: Description and Research*. *Language* 68:1 pp196-199.
- Francis, W N and H Kucera 1982** *Frequency Analysis of English Usage: Lexicon and Grammar*. Houghton, Mifflin, Boston.
- Frawley, W 1984a** Prolegomenon to a theory of Translation. In Frawley, pp159-175.
- Frawley, W (ed) 1984b** *Translation*. Associated University Presses, Ontario.
-
- Gal, S 1979** *Language Shift Social Determinants of Linguistic Change in Bilingual Austria*. Academic Press, NewYork.
- Galbally Report 1978** *Clearing House on Migration Issues*, Melbourne.
- Garfinkel, 1972** Remarks on ethnomethodology. In Gumperz and Hymes, pp301-324.
- Garner, M (ed) 1981** *Community Languages Their Role in Education*. River Seine Publications Melbourne.
- Garside, R et al (eds) 1987** *The Computational Analysis of English - a corpus-based approach*. Longman, London.
- Garvin, P (ed) 1970a** *Method and Theory in Linguistics*. Mouton, The Hague.
- Garvin, P 1970b** Introduction. In Garvin, pp9-12.
- Gass, S and L Selinker 1982** Language Transfer: A Conference Report. In Eppert, pp5-26.
- Gass, S and L Selinker 1983** *Language Transfer in Language Learning*. Newbury House Publishers, Rowley, Mass.
- Gellerstam, M 1992** Modern Swedish Corpora. In Svartvik, pp149-163
- Gellerstam, M et al. (eds) 1996** *Euralex '96 Proceedings*. Göteborg University Press, Dept. of Swedish.
- Gerver, D and H W Sinaiko (eds) 1978** *Language Interpretation and Communication*. Plenum Press, NewYork.
- Gilbert, G G (ed) 1970** *Texas Studies in Bilingualism*. De Gruyter, Berlin.
- Gilbert, G G 1972** Creole and Immigrant Languages: Basilect+Code-Switching or Restructuring? In Firchow, pp236-242.

- Giles, H 1973** Accent mobility: A model and some data. *Anthropological Linguistics*. 15 pp87-105.
- Giles, H (ed) 1977** *Language, Ethnicity and Intergroup Relations*. Academic Press, London.
- Giles, H., R Bourhis and D Taylor 1977** Towards a theory of language in ethnic group relations. In Giles, pp307-49.
- Giles, H and W P Robinson (eds) 1990** *Handbook of Language and Social Psychology*. John Wiley & Sons, Chichester.
- Gilson, M and J Zubrzycki 1967** *The Foreign Language Press in Australia 1848-1964*. Australian National University, Canberra.
- Gimbel, J 1988** Immigrant Danish - A Multitude of Languages. Migracijske terme, godina 4 br. 4 Zagreb. Centre for Migration and Nationalities Studies, pp79-89.
- Gleason, H A 1962** The Relation of Lexicon and Grammar. In Householder and Saporta, pp85-110.
- Gödüp , B 1995** Demokrasi ve Kitle Dili. *‘ađlapTürk Dili* 88 pp2-4.
- Grahs, L., G Korlen and B Malmberg (eds) 1976** Theory and Practice of Translation. Peter Lang, Bern. **Greenbaum, S 1984** Corpus analysis and Elicitation Tests. In Aarts and Meijs, pp193-201.
- Grimshaw, A 1987a** Micro-/Macrolevels. In Ammon et al., pp66-77.
- Grimshaw, A 1987b** Sociolinguistics versus Sociology of Language: tempest in a teapot or Profound Academic Conundrum? In Ammon et al., pp9-15.
- Grishman, R 1986** Computational Linguistics. Cambridge University Press, Cambridge.
- Grosjean, F 1982** Life with Two Languages. Harvard University Press, Cambridge, Mass.
- Grosjean, F 1990** *Synthesis of the Workshop on constraints, conditions and models. European Scientific Network on Code-Switching and Language Contact. Papers for the workshop on concepts, methodology and data. London meeting.* pp1-3.
- Grosjean, F 1995** A Psycholinguistic Approach to Code-Switching. In Milroy and Muysken pp259-275.
- Grosz, B., K Sparck Jones and B L Webber (eds) 1986** *Readings in Natural Language Processing*. Morgan Kaufmann Publishers, Inc., California.
- Gumperz, J 1962** Types of Linguistic Communities. *Anthropological Linguistics* 4:1 pp28-40.
- Gumperz, J 1964** Linguistic and Social Interaction in Two Communities. In Gumperz and Hymes, pp137-153.
- Gumperz, J 1967** On the Linguistic Markers of Bilingual Communication. *Journal of Social Issues* 23:2 pp48-57.
- Gumperz, J 1969** How can we describe and measure the behaviour of bilingual groups? In Kelly, pp242-249.
- Gumperz, J 1968** The Speech Community. *International Encyclopedia of the Social Sciences*. MacMillan, NewYork, pp381-386.
- Gumperz, J 1970** Verbal Strategies in Multilingual Communication. In Alatis, pp129-143.
- Gumperz, J 1971** *Language in Social Groups*. Stanford University Press, Stanford, California.
- Gumperz, J 1972** Introduction. In Gumperz and Hymes, pp1-25.
- Gumperz, J and D Hymes (eds) 1964** The ethnography of communication. *American Anthropologist* 66:6 part 2 pp137-154.
- Gumperz, J and D Hymes 1972a** Preface. In Gumperz and Hymes. ppv-viii.
- Gumperz, J and D Hymes (eds) 1972b** *Directions in Sociolinguistics The Ethnography of Communication*. Holt, Rinehart and Winston, Inc. NewYork.
- Gumperz J and R Wilson 1977** Convergence and creolization. First published in Hymes (1971b), pp151-167.
- Hakuta, K 1986** *Mirror of Language: The Debate on Bilingualism*. Basic Books, NewYork.

- Halliday, M A K 1966** *Grammar, Society and The Noun*. H K Lewis & Co Ltd., London.
- Halliday, M A K 1985** *Spoken and Written Language*. Deakin University Press, Victoria.
- Halliday, M A K 1987** Language and the order of nature. In Fabb et al., pp135-154.
- Halliday, M A K 1994** *An Introduction to Functional Grammar*. 2nd edition. Arnold, London.
- Halliday, M A K 1998** Representing the child as a semiotic being (one who means). Plenary Speech. *Representing The Child Conference*. Monash University, Melbourne. 2-3 October.
- Halliday, M A K and J R Martin 1993** *Writing Science*. The Falmer Press, London.
- Halliday, M A K., A McIntosh and P Stevens 1964** *The Linguistic Sciences and Language Teaching*. Longmans, London.
- Hamers, J and M Blanc 1989** *Bilinguality and bilingualism*. Cambridge University Press.
- Hankamer, J 1986** Turkish generative morphology and morphological parsing. Paper presented at Second International Conference on Turkish Linguistics, Istanbul.
- Hankamer, J 1988** Parsing Nominal Compounds in Turkish. In Wallace, pp123-143.
- Hankamer, J 1989** Morphological parsing and the lexicon. In Marslen-Wilson, pp392-408.
- Hankamer, J 1993** keçitexts: Text-Based Analysis of Morphology and Syntax in an Agglutinating Language. ACH-ALLC Conference, Georgetown, Washington DC. [abstract]
- Harding, E and P Riley 1986** *The Bilingual Family: a Handbook for Parents*. Cambridge University Press, Cambridge.
- Harris, Z 1954** Transfer Grammar. *International Journal of American Linguistics*. **20** pp259-270.
- Hasselmo, N 1961** *American Swedish*. Ph.D Thesis. Harvard University.
- Hasselmo, N 1969** How can we measure the effects which one language may have on the other in the speech of bilinguals? In Kelly, pp122-141.
- Hasselmo, N 1970** Code-Switching and Modes of Speaking. In Gilbert, pp179-210.
- Hasselmo, N 1972** Code-Switching as Ordered Selection. In Firchow et al., pp261-280.
- Haugen, E 1938** Language and Immigration. *Norwegian-American Studies and Records* **10** pp1-43.
- Haugen, E 1950** The Analysis of Linguistic Borrowing. *Language* **26:2** pp210-231.
- Haugen, E 1953** (2nd Edition 1969) *The Norwegian Language in America. A Study in Bilingual Behaviour*. Indiana University Press, Bloomington.
- Haugen, E 1956** *Bilingualism in the Americas. A Bibliography and Research Guide*. American Dialect Society.
- Haugen, E 1970** Linguistics and Dialinguistics. In Alatis, pp1-7 and 40-41.
- Haugen 1972** Bilingualism, Language Contact, and Immigrant Languages in the United States: A Research Report 1956-1970. In Fishman, pp1-111.
- Haugen, E 1973** Bilingualism, Language Contact and Immigrant Languages in the US. *Current Trends in Linguistics* **10** pp505-591.
- Hays, D G 1976** The Past, Present, and Future of Computational Linguistics - Closing Address. In Papp and Szepe, pp583-585.
- Hays, P 1997** *Collocational Similarity: Emergent Patterns in Lexical Environments*. Ph.D Thesis. Birmingham University.
- Hazai, G 1970** Linguistics and Language Issues in Turkey. *Current Trends in Linguistics* **6** pp746-758.
- Heller, M (ed) 1988** *Codeswitching: anthropological and Sociological Perspectives*. Mouton de Gruyter, Berlin.
- Heller, M 1990** The politics of code-switching: processes and consequences of ethnic mobilisation. In *European Scientific Network on Code-Switching and Language Contact*. Brussels meeting. pp53-76.
- Heller, M 1992** The Politics of Codeswitching and Language Choice. In Eastman, pp123-142.
- Hetmar, T and J N Jørgensen 1993** Multilingual Concepts in the Schools of Europe: Denmark. *Sociolinguistica* **7** pp79-89.

- Heyd, U 1954** *The Language Reform in Turkey*. The Israel Oriental Society publications, Jerusalem.
- Hockey, S and D Walker 1993** Developing Effective Resources for Research on Texts: Collecting Texts, Tagging Texts, Cataloguing Texts, Using Texts, and Putting Texts in Context. *Literary and Linguistic computing*. 8:4 pp235-242.
- Hoey, M 1998** Some Text properties of Certain Nouns. *Colloquium on Discourse Anaphora and Reference Resolution*. University of Lancaster.
- House J and S Blum Kulka (eds) 1986** *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition*. G Narr, Tübingen.
- Householder, F W and S Supporta (eds) 1962** *Problems in Lexicography*. Indiana University, Bloomington.
- Hugo, G and C Maher 1995** *Atlas of the Australian People - 1991 Census - National Overview*. BIMPR.
- Hyltenstam, K and L Obler 1989a** Introduction. In Hyltenstam and Obler, pp1-12.
- Hyltenstam, K and L Obler (eds) 1989b** *Bilingualism Across The Lifespan: An Introduction*. Cambridge University Press, Cambridge.
- Hymes, D 1968** Why Linguistics Needs The Sociologist. *Social Research* 34 pp632-647.
- Hymes, D 1970** Linguistic Method in Ethnography: Its Development in the United States. In Garvin. pp249-311.
- Hymes, D 1971a** Preface. In Hymes, pp3-11.
- Hymes, D (ed) 1971b** *Pidginization and Creolization of Languages*. Proceedings of Conference held at the University of the West Indies, Mona, Jamaica, April 1968. Cambridge University Press, London.
- Hymes, D 1972** Morris Swadesh: From the First Yale School to World Prehistory. In Sherzer, pp228-270
- Hymes, D 1980** Foreword. In Sankoff, ppix-xv.
- Hymes, D and J Fought 1975** (reprinted 1981) *American Structuralism*. Mouton Publishers, The Hague.
- Ýçduygu, A 1991** *The Unintended settlement of Turks in Australia*. Ph.D Thesis. Australian National University, Canberra.
- Ýçduygu, A 1994** Facing Changes and Making Choices: Unintended Turkish Settlement in Australia. *International Migration* 32:1 pp71-90.
- Ýmer, K 1993** *Toplumbilim Kavramları İþýnda Türk Dil Devriminin Deðerlendirilmesi. Uygulamalı Dilbilim Açısından Türkçenin Görünü* . Dil Derneði Yayınları 5. pp75-87.
- Ýmer, K and E Uzun (eds) 1997** *Proceedings of the VIIIth International Conference on Turkish Linguistics August 7-6, 1996*. Ankara Üniversitesi Basımevi.
- Inglis, C and L Manderson 1988** Turkish families, their children and education in Australia. In Akçelik and Elley, pp47-78.
- Inglis, C., J Elley and L Manderson 1992** *Making Something of Myself*. Office of Multicultural Affairs, Canberra.
- Isaac, D 1988** A Transcultural educational and support group for Turkish women in poor health. In Akçelik and Elley, pp181-201.
- Jahr, E H (ed) 1992** *Language Contact. Theoretical and Empirical Studies*. Mouton de Gruyter, The Hague.
- Jake, J 1998** Constructing interlanguage: building a composite matrix language. *Linguistics* 36:2 pp333-382.
- Jakobson, R., C Fant, M Gunnar and M Halle 1952** *Preliminaries to Speech Analysis*. MIT Press, Cambridge, Mass.

- Johanson, L 1993** Code-copying in Immigrant Turkish. In Extra and Verhoeven (1993c), pp197-221.
- Johansson, S and A Stenström 1991 (eds)** *English Computer Corpora Selected Papers and Research Guide*. Mouton de Gruyter, The Hague.
- Johansson, S and K Hofland 1994** Towards an English-Norwegian parallel corpus. In Fries, pp25-37.
- Jørgensen, J N 1993** Children's Code Switching in Group Conversations. In *European Scientific Network on Code-Switching and Language Contact. Paris meeting*. pp165-181.
- Joshi, A K 1984** Processing of Sentences with Intra-Sentential Code-Switching. In Dowty, pp190-20.
- Jupp, J (ed) 1988** *The Australian People: An Encyclopedia of the Nation, Its people and Their Origins*. Angus & Robertson, Sydney.
- Jupp, J 1966** *Arrivals and Departures*. Chesire-Lansdowne, Melbourne.
- Jupp, J 1991** *Immigration*. Sydney University Press, Sydney.
- Kachru, B 1978** Towards Structuring Code-Mixing: An Indian Perspective. *International Journal of the Sociology of Language*. **16** pp27-46.
- Kachru, B 1980** Bilingualism. *Annual Review of Applied Linguistics* pp2-18.
- Kachru, B 1981** 'Socially realistic linguistics': the Firthian Tradition. **International Journal of the Sociology of Language** **31** pp65-89.
- Kardam, F and C Pfaff 1993** Issues in Educational Policy and Language Development of Bilingual Children in Berlin. In Kroon et al., pp 51-68.
- Karlsson, F 1992** Comments. In Svartvik, pp398-400.
- Keçeli, B 1992** Script of a talk at Ecumenical Migration Centre.
- Keçeli, B and Cahill, D 1995** Education and Inequality: A Case study of second generation Turkish Australians. Paper presented at *Aspects of Australian Multiculturalism Conference*. Monash University, Melbourne, 18-19 Sept.
- Kelly, L G (ed) 1969** *Description and Measurement of Bilingualism: an international seminar*. University of Moncton June 6-4 1967. University of Toronto Press, Canada.
- Kettemann, B and W Wieden (eds)** *Current issues in European Second Language Acquisition Research*. G Narr, Tübingen.
- Kjellmer, G 1987** Aspects of English Collocations. In Meijs, pp133-140.
- Kjellmer, G 1991** A mint of phrases. In Aijmer and Altenberg, pp111-127.
- Kloss, H 1966** German-American Language Maintenance Efforts. In Fishman, pp206-252.
- Knapp, K., W Enninger and A Knapp-Potthoff (eds) 1987** *Analyzing Intercultural Communication*. Mouton de Gruyter.
- Kolers, P A and E Gonzales 1980** Memory for words, synonyms and translations. *Journal of Experimental Psychology, Human Learning and Memory* **6** pp53-65.
- Kolers, P A and M Paradis 1980a** Introduction. In Kolers and Paradis, pp287-303.
- Kolers, P A and M Paradis (eds) 1980b** Psychological and Linguistic Studies of Bilingualism. *Canadian Journal of Psychology*. Special Issue. **34**:4.
- Koskeniemmi, K 1983** *Two-level morphology: A general computational module for word-form recognition and production*. Department of Linguistics Publications 11. University of Helsinki, Helsinki.
- Kroon, S and T Vallen 1995** Multilingualism and Education: An Overview of Language and Education policies For Ethnic Minorities in The Netherlands. In Wright, pp5-31.
- Kroon, S., D Pagel and T Vallen (eds) 1995** *Multiethnische Gesellschaft und Schule in Berlin*. Waxmann, Münster.
- Kurtböke, P 1996a** The Impact of Corpus Planning on Bilingual Dictionaries. In Gellerstam et al., pp591-596.
- Kurtböke, P 1996b** Türk Sözlükçülüğünün Sorunları II. *Kuram* **11** pp70-74.
- Kurtböke, P 1996c** A Corpus-based study of the Turkish Community Newspapers in Australia. A Progress Report. In Ýmer and Uzun, pp221-230.

- Kurtböke, P 1996d** Towards a Corpus of Australian Turkish. *University of Melbourne Working Papers in Linguistics and Applied Linguistics* **16** pp151-158.
- Kurtböke, P 1997** Migrants Mediation Mixed Code. *BAAL 30th Annual Meeting*. The University of Birmingham, 9-11 September 1997.
- Kurtböke, P 1998a** Delexicalised verbs in Turkish From a Corpus Perspective. *9th International Conference on Turkish Linguistics* University of Oxford, 12-14 August, 1998.
- Kurtböke, P 1998b** Non-equivalence of delexicalised verbs in bilingual dictionaries. In Fontenelle et al. Vol 2 pp397-404.
- Kurtböke, P 1998c** Ozturk Corpus. *TASG News* **46** pp34-40.
- Kurtböke, P in press** Delivery yapýlýr: single occurrence or co-selection? *SIB 97 Proceedings*, University of Vigo, Spain.
- Kurtböke, P and L Potter 1998** Neither a borrowing nor a Switch be: co-occurrence tendencies of foreign words in corpora. *BAAL 31st Annual Meeting* The University of Manchester 11-13 September, 1998.
- Kurtböke, P, M Verde and P King (in progress)** On doing and making: Delexicalization in Turkish and Greek.
- Kytö, M., O Ihalainen and M Rissanen (eds) 1988** *Corpus Linguistics, Hard and Soft*. Rodopi, Amsterdam
-
- Labov, W 1968** The Reflection of Social Processes in Linguistic Structures. In Fishman (1968a), pp240-251.
- Labov, W 1972** On the Mechanism of Linguistic Change. In Gumperz and Hymes, pp512-537.
- Labov, W 1982** Building on Empirical Foundations. In Lehmann and Malkiel, pp17-92.
- Lambert, W 1955** Measurement of the Linguistic Dominance of Bilinguals. *Journal of Abnormal and Social Psychology* **50** pp197-200.
- Lambert, W E 1972** *Language, Psychology and Culture: Essays by W E Lambert*. Edited by A S Dil. Stanford University Press, Stanford, California.
- Lambert, W E 1989** Language Attrition. *Review of Applied Linguistics* **83** pp1-18.
- Lambert, W and S Fillenbaum 1959** A Pilot Study of Aphasia Among Bilinguals. *Canadian Journal of Psychology* **13:1** pp28-34.
- Lambert, W, J Havelka and R C Gardner 1959** Linguistic Manifestations of Bilingualism. *The American Journal of Psychology* **72** pp77-82.
- Lambert, W, Havelka J and C Crosby 1958** The Influence of Language Acquisition Contexts on Bilingualism. *The Journal of Abnormal and Social Psychology* **56** pp239-244.
- Lambert, W and B Freed 1982** *The Loss of Language Skills*. Newbury House, Rowley, Mass.
- Larsen-Freeman, D 1996** Impressions of AILA 1996. *AILA Review* **12** pp87-92.
- Leech, G 1987** General Introduction. In Garside, pp1-15.
- Leech, G 1993** Corpus Annotation Schemes. *Literary and Linguistic Computing* **8:4** pp276-281.
- Lehmann, W P and Y Malkiel (eds) 1968** *Perspectives on Historical Linguistics*. (Reprinted 1982) John Benjamins Publishing Company, Amsterdam.
- Leopold, W F 1939-1949** *Speech development of a bilingual child*. 4 vols. Northwestern University Press, Evanston, Illinois.
- Levelt, W J M 1989** *Speaking: From Intention to Articulation*. MIT Press, Cambridge, Mass.
- Lewis, G L 1967 (reprinted 1985)** *Turkish Grammar*. Oxford University Press, Oxford.
- Lieberson, S 1966** Introduction. *Sociological Inquiry*. (Special Issue on Sociolinguistics) **36:2** pp131-134.
- LOTE Report 1994** Report to the Minister of Education, Ministerial Advisory Council on LOTEs 1994 Directorate of School Education, Melbourne.

- LOTE Strategy Plan 1993** Directorate of School Education and Ministerial Advisory Council on LOTES.
- Lux, P and W Grabe 1991** Multivariate approaches to contrastive rhetoric. *Linguas Modernas* **18** pp133-60.
- Lüdi, G 1987** *Devenir bilingue, Parler bilingue*. Actes du 2e colloque sur le bilinguisme. Université de Neuchâtel, 20-22 Sept. 1984. Max Niemeyer Verlag, Tübingen.
- Lüdi, G 1990** Synthesis of the Workshop on "Impact and Consequences: Broader considerations". European Scientific Network on Code-Switching and Language Contact. *Papers for the workshop on concepts, methodology and data. Brussels meeting*, pp1-3.
- Lüdi, G 1992** Preface: A plural approach to the dynamics of Languages in contact. *Multilingua* **11:1** pp7-10.
- Mackey, W 1953** Bilingualism and Linguistic Structure. *Culture* **14** pp143-149.
- Mackey, W 1956** Toward A Redefinition of Bilingualism. *Canadian Journal of Applied Linguistics* **2** pp4-11.
- Mackey, W F 1965** Bilingual Interference; its analysis and measurement. *The Journal of Communication* **15** pp239-249.
- Mackey, W F 1969** Introduction - How can bilingualism be described and measured? In Kelly, pp2-9.
- Mackey, W 1970** Interference, Integration and The Synchronic Fallacy. In Alatis pp195-223.
- Mackey, W 1993** Introduction. In Fernandez, ppxiii-xx.
- Macnamara, J 1967a** The Bilingual's Linguistic Performance - A Psychological Overview. *Journal of Social Issues* **23:2** pp58-77.
- Macnamara, J 1967b** Bilingualism in the modern world. *Journal of Social Issues* **23:2** pp1-7.
- Macnamara, J 1970** Bilingualism and Thought. In Alatis, pp25-45.
- Mallery, M 1995** Quality Issues and Electronic Texts. *CETH Newsletter*. **3:1**.
- Manderson, L 1988** Turks. In Jupp, pp818-825.
- Marcinkeviciene, R and N Volz (eds) 1997** *TELRI Proceedings of the Second European Seminar. Kaunas, Lithuania April 17-20*. Vytautas Magnus University Publishers, Kaunas, Lithuania.
- Martinet, A 1953** Preface. In Weinreich, ppvii-ix.
- Marslen-Wilson, W 1989** *Lexical Representation and Process*. A Bradford Book. The MIT Press. Cambridge, Mass.
- Meijs, W 1987** Preface. In Meijs, ppii-v.
- Meijs, W (ed) 1987** *Corpus Linguistics and Beyond - Proceedings of the Seventh International Conference on English Language Research on Computerised Corpora*. Rodopi, Amsterdam.
- Meyer, L, K Mackintosh and K Varantola 1998** From Virtual Sex to Virtual Dictionaries: On the Analysis and Description of a De-terminologized Word. In Fontenelle et al., pp643-654.
- McNaught, J 1993** User needs for textual corpora in Natural Language Processing. *Literary and Linguistic Computing* **8** pp227-234.
- Milroy, L 1980** *Language and Social Networks*. 2nd edition 1987. Blackwell, Oxford.
- Milroy, J 1992** *Linguistic Variation and Change*. Blackwell, Oxford.
- Milroy, J and P Muysken 1995** Introduction. In Milroy and Muysken, pp1-14.
- Milroy, J and P Muysken 1995** *One Speaker, two languages. Cross Disciplinary perspectives on code-switching*. Cambridge University Press, Cambridge.
- Mitchell, R., J Hooper and C Martin 1997** Linking language teaching and language learning: the case of prefabricated chunks. Paper presented at *BAAL 30th Meeting*, The University of Birmingham, 9-11 September 1997.

Moon, R E 1994 *Fixed Expressions and Text. A Study of the Distribution and Textual Behaviour of Fixed Expressions in English*. Ph.D Thesis. The University of Birmingham.

Muysken, P 1990 Concepts, Methodology and Data in Language contact Research: Ten Remarks from the Perspective of Grammatical Theory. *European Scientific Network on Code-Switching and Language Contact, Basel meeting*. pp15-29.

Muysken, P 1995 Code-switching and grammatical theory. In Milroy and Muysken, pp177-198.

Mühlhäusler, P 1980 Structural Expansion and the Process of Creolization. In Valdman and Highfield, pp19-56.

Myers-Scotton, C 1988 Differentiating Borrowing and Codeswitching. In Ferrara et al., pp318-325.

Myers-Scotton, C 1992 Comparing Codeswitching and Borrowing. In Eastman, pp19-39.

Myers-Scotton, C 1993a *Duelling Languages*. Clarendon Press, Oxford.

Myers-Scotton, C 1993b *Social Motivations for Codeswitching*. Clarendon Press, Oxford.

Nattinger, J R and J S DeCarrico 1992 *Lexical Phrases and Language Teaching*. Oxford University Press, Oxford.

Nelde, P (ed) 1980 *Languages in Contact and Conflict*. Franz Steiner Verlag, Wiesbaden.

Nelde, P H 1983a Plaidoyer Pour Une Linguistique des Langues en Contact. In Nelde (1983d), pp3-12.

Nelde, P 1983b Prefatory Notes. In Nelde, ppix-xi.

Nelde, P (ed) 1983c *Multilingualism*. Dümmler, Bonn.

Nelde, P H (ed) 1983d *Current Trends in Contact Linguistics*. Dümmler, Bonn.

Neumann, J H 1938 Notes on American Yiddish. *Journal of English and Germanic Philology* 37 pp403-421.

Nishimura, M 1986 Intrasentential Code-Switching: The case of Language Assignment. In Vaid, pp123-143.

Oblor K and M Albert 1978 *Language and Communication in the elderly: clinical, therapeutic and experimental issues*. Lexington Books, Lexington, Mass.

Oflazer, K and C Bozþ ahin 1994 *Progress Report on Turkish Natural Language Processing Initiative: Developing Computational Foundations and Applications for NLP in Turkish (TU-LANGUAGE) for NATO Science for Stability Program*. Bilkent University, Ankara.

Öim, H 1989 Language Understanding and Problem Solving: On the Relation Between Computational Linguistics and Artificial Intelligence. In Batori, pp277-283.

Oksaar, E 1967 Commentary. in Kelly, pp147-152.

Oksaar, E 1972a Bilingualism. *Current Trends in Linguistics* 9 pp476-511.

Oksaar, E 1972b On Code Switching. An Analysis of Bilingual Norms. *Proceedings of Third International Applied Linguistics Conference*. Vol 3 pp491-500.

Oksaar, E 1976 Interference, Bilingualism and Interactional Competence. In Grahs et al, pp295-305.

Oksaar, E 1980 Linguistic Aspects in the Development of Bilingualism. In Afendras pp305-321.

Oostdijk, N and P de Haan (eds) 1994 Corpus-based research into language: in honour of Jan Aarts. Rodopi, Amsterdam.

Osgood, C E., G Suci and P Tannenbaum 1957 *The Measurement of Meaning*. Urbana

Owen, C 1988 Naturalness and the language learner. *ELR Journal* 2 pp21-46.

Ozolins, U 1993 *The Politics of Language in Australia*. Cambridge University Press, Cambridge.

Özdemir, A Z 1995 Dilimizi Bozanlar. Çağdaş Türk Dili 94 pp7-9.

Özdemir, E and O N Poyrazoğlu 1995 Yazınbilimci Emin Özdemir'le Almanya'da Türkçe Öğretimi Üzerine. *Çağdaş Türk Dili* . 84. pp23-27.

Palmer, F (ed) 1968 *Selected Papers by J R Firth*. Longman, London.

Papp, F and G Szepe (eds) 1976 *Papers in Computational Linguistics - Proceedings of the 3rd International Meeting on Computational Linguistics*. Mouton, The Hague.

Paradis, M 1978 *Aspects of Bilingualism 4th Lacus Forum 1977*. Hornbeam Press, Inc.

Paradis, M 1980 The language switch in bilinguals. In Nelde, pp501-506.

Paradis, M 1983 Sociological Context and Cerebral Representation of Languages in Contact. In Nelde 1983, pp215-226.

Paradis, M 1987 *The Assessment of Bilingual Aphasia*. Lawrence Erlbaum Assoc. New Jersey.

Partington, A 1993 Corpus Evidence of Language Change - The Case of the Intensifier. In Baker et al., pp177-192.

Pawley, A and F Syder 1983 'Two Puzzles for Linguistic Theory: native-like selection and nativelike fluency'. In Richards and Schmidt, pp191-226.

Peal, E and W Lambert 1962 The relation of bilingualism to intelligence. *Psychological Monographs* 76:27 pp1-23.

Peköz, B 1993 Second-generation Turkish Youth's bilingual experience and its effects on individuals. In Akçelik, pp121-137.

Penfield, W and L Roberts 1959 *Speech and Brain Mechanism*. Princeton University Press, Princeton.

Perecman, E 1989 Language processing in the bilingual: evidence from language mixing. In Hyltenstam and Obler pp227-244.

Pérez Hernández, C 1996 A Pilot Study on Translation Equivalence between Spanish and English. In Sinclair *et al.*, pp218-237.

Petre, B 1995 *Perspectives on Greek-Australian Media and Language Maintenance in Melbourne*. M.A Thesis, Monash University, Melbourne.

Pfaff, C W 1979a Constraints on language mixing: intrasentential code-switching and borrowing in Spanish/English. *Language* 55 pp291-318.

Pfaff, C W 1979b Incipient Creolization in "Gastarbeiterdeutsch"?: An Experimental Sociolinguistic Study. (Paper presented at the Conference on Theoretical Orientations in Creole Studies). *Studies in Second Language Acquisition* 3 pp165-178.

Pfaff, C W 1988 Linguistic and Social Determinants of Turkish/German Bilingualism of Migrant Children. In L Durovic (ed) *Diaspora Languages in Western Europe*. Lund. *Slavica Lundensia* 12 pp49-74.

Pfaff, C W 1990 Mixing and Linguistic Convergence in Migrant Speech Communities: linguistic constraints, social conditions and models of acquisition. *European Scientific Network on Code-Switching and Language Contact, London meeting*. pp119-153.

Pfaff, C W 1991 Turkish in contact with German: Language maintenance and loss among immigrant children in Berlin (West). *International Journal of the Sociology of Language* 90 pp97-129.

Pfaff, C W 1993 Turkish language development in Germany. In Extra and Verhoeven (1993c), pp119-146.

Pfaff, C W 1998 The development of yapmak as auxiliary and main verb in the speech of Turkish/German bilingual children. Panel paper at *9th International Conference on Turkish Linguistics* University of Oxford, 12-14 August, 1998.

- Pierce, J 1961** A Frequency count of Turkish Affixes. *Anthropological Linguistics* 3:9 31-42.
- Pierce, J 1962a** Frequencies of occurrence for affixes in written Turkish. *Anthropological Linguistics* 4:6 pp30-41.
- Pierce, J 1962b** Possible electronic computation of typological indices for linguistic structures. *International Journal of American Linguistics* 28 pp215-226.
- Pierce, J 1963a** A Statistical Study of Grammar and Lexicon in Turkish and Sahaptin (Klikitat). *International Journal of American Linguistics* 29 pp96-103.
- Pierce, J 1963b** Typological Indices for written vs. Spoken Turkish. *International Journal of American Linguistics* 29 pp71-72.
- Pierce, J 1964** *Life in a Turkish Village*. Holt, Rinehart and Winston, New York.
- Pierce, J 1965** On the Validity of Genetic Classifications. *Linguistics* 13 pp25-33.
- Pike, K 1967** *Language in relation to a unified theory of the structure of human behaviour*. Mouton, The Hague.
- Pike, K 1961** Stimulating and Resisting Change. *Practical Anthropology* 8 pp267-274.
- Pike, K and C Fries 1949** Coexistent Phonemic Systems. *Language* 25:1 pp29-50.
- Polomé, E C (ed) 1990** *Research Guide on Language Change*. Mouton de Gruyter, Berlin.
- Poplack, S 1978** Quantitative Analysis of Constraints on Code-Switching. Centro de Estudios Puertorriquenos Working Papers.
- Poplack, S 1980** Sometimes I'll start a sentence in Spanish y termino en Español: toward a typology of code-switching. *Linguistics* 18 pp581-618.
- Poplack, S 1985** Contrasting Patterns of code-switching in two communities. In Warkentyne, H J (ed) *Proceedings of the Fifth International Conference on Methods in Dialectology*. University of Victoria Press. pp363-385.
- Poplack, S 1987** Contrasting Patterns of Code-Switching in Two Communities. In Wande et al., pp51-76.
- Poplack, S 1990** Variation theory and language contact: concepts, methods and data. In *European Scientific Foundation Network on Codeswitching and Language Contact. Basel meeting*. pp34-68.
- Poplack, S and D Sankoff 1984** Borrowing: The synchrony of integration. *Linguistics* 22 pp99-135.
- Poplack, S and M Meecham 1995** Patterns of language mixture: nominal structure in Wolof-French and Fongbe-French bilingual discourse. In Milroy and Muysken, pp199-232.
- Poplack, S, D Sankoff and C Miller 1988** The Social Correlates and Linguistic Processes of Lexical Borrowing and Assimilation. *Linguistics* 26 pp47-104
- Poplack, S, S Wheeler and A Westwood 1989** Distinguishing language contact phenomena: evidence from Finnish-English bilingualism. In Hyltenstam and Obler, pp132-154.
- Price, C A 1989** *Ethnic Groups In Australia*. Australian Immigration Research Centre, Canberra.
- Price, C A and J I Martin (eds) 1976** *Australian Immigration - A Bibliography and digest*. Number 3, 1975, Part 2. Department of Demography, Institute of Advanced Studies, Australian National University.

Quirk, R 1992 On corpus principles and design. In Svartvik, pp457-469.

- Rayfield, J R 1970** *The Languages of a Bilingual Community*. Mouton, The Hague.
- Rehbein, J 1987** Multiple formulae Aspects of Turkish migrant workers' German in intercultural communication. In Knapp et al, pp215-248.
- Renouf, A and J McH Sinclair 1991** Collocational frameworks in English. In Aijmer and Altenberg, pp128-143

- Richards, J C and R W Schmidt (eds) 1983** *Language and Communication*. Longman, London.
- Rissanen, M., O Ihalainen and M Kytö 1987** The Helsinki Corpus of English Texts. In Meijs, pp21-32.
- Ritchie, G D et al 1992** *Computational Morphology*. The MIT Press, Cambridge, Mass.
- Robinson, P W and H Giles 1990** Prologue. In Giles and Robinson, pp1-8.
- Romaine, S 1989** *Bilingualism*. Basil Blackwell, Oxford.
- Romaine, S 1991a** Introduction. In Romaine, pp1-24.
- Romaine, S (ed) 1991b** *Language in Australia*. Cambridge University Press, Cambridge.
- Ronjat, J 1913** *Le développement du langage observé chez un enfant bilingue* Champion, Paris.
- Röhr-Sendlmeier U M 1990** Social Context and The Acquisition of German by Turkish Migrant Children. *Journal of Multilingual and Multicultural Development*. 11:5 pp377-391.
- Salmons, J 1990** The context of language change. In Polomé, pp71-96.
- Sankoff, G 1974** A Quantitative Paradigm for the Study of Communicative Competence. (Reprinted in 1980 *The Social Life of Language*) pp47-79.
- Sankoff, G 1980** *The Social Life of Language*. University of Pennsylvania Press.
- Sankoff, D and S Poplack 1981** A formal grammar for code-switching. *Papers in Linguistics* 14 pp3-45.
- Sankoff, D., S Poplack and S Vanniarajan 1990** The case of the nonce loan in Tamil. *Language Variation and Change* 2 pp71-101.
- Sapir, E 1921** *Language* Harcourt Brace, New York.
- Sapir, E 1933** *Language*. Reprinted in Mandelbaum, D G 1963 *Selected Writings of Edward Sapir in Language, Culture and Personality*. University of California Press. Berkeley.
- Sayar, H 1988** Turkish Australians and Education in Australia. In Akçelik and Elley, pp101-110.
- Schaufeli, A 1991** *Turkish in an Immigrant Setting - A comparative study of the first language of monolingual and bilingual Turkish children*. Ph.D Thesis, University of Amsterdam.
- Schaufeli, A 1992** Attrition and Shift of Turkish in the Netherlands. Draft paper. Dept. of Applied Linguistics, University of Nijmegen.
- Schulze, B M et al., 1994** *DECIDE Designing and Evaluating Extraction Tools for Collocations in Dictionaries and Corpora*. MLAP Project 93-19.
- Scotton, C and J Okeju 1973** Neighbors and Lexical Borrowings. *Language* 49:4 pp871-889.
- Seliger, H and R Vago (eds) 1991** *First language attrition*. Cambridge University Press.
- Seliger, H and R Vago 1991** The study of first language attrition: an overview. In Seliger and Vago, pp3- 15
- Selinker, L 1969** Language Transfer. *General Linguistics* 9 pp67-92.
- Selinker, L 1972** Interlanguage. *IRAL* 10:3 pp209-231.
- Sgall, P 1989** Theoretical Aspects of Computer-Aided Description of Language Systems. In Batori et al., pp127-140.
- Sharwood Smith, M 1983** On First Language Loss in the Second Language Acquirer: Problems of Transfer. In Gass and Selinker, pp222-231.
- Sherzer, J (ed) 1972** *Morris Swadesh. The Origin and Diversification of Language*. Routledge & Kegan Paul, London.
- Silva-Corvalán, C 1986** Bilingualism and Language Change: The extension of ESTAR in Los Angeles Spanish. *Language* 62 pp587-608.
- Silva-Corvalán, C 1994** *Language Contact and Language Change. Spanish in Los Angeles*. Clarendon Press, Oxford.

- Silva-Corvalán, C 1995a** Lexico-Syntactic modeling across the bilingual continuum. In Fisiak, pp253-270.
- Silva-Corvalán, C 1995b** The Study of Language Contact: An Overview of the Issues. In Silva-Corvalán, pp3-14.
- Silva-Corvalán, C 1995c** *Spanish in Four Continents Studies in Language Contact and Bilingualism*. Georgetown University Press Washington DC.
- Sinclair, J 1984** Naturalness in Language. In Aarts and Meijs, pp203-210.
- Sinclair, J (ed) 1987a** *Looking Up: An Account of the COBUILD Project in Lexical Computing*. Collins, London.
- Sinclair, J 1987b** Introduction. Collins Cobuild English Language Dictionary. pp xv-xxi.
- Sinclair, J 1987c** Collocation: a progress report. In Steele and Threadgold, pp319-331.
- Sinclair, J 1988** Naturalness in Language. *ELR Journal* 2 pp11-20.
- Sinclair, J 1989** Uncommonly Common Words. In Tikoo, pp135-152.
- Sinclair, J et al. 1990** *Collins Cobuild English Grammar*. Collins, Glasgow.
- Sinclair, J 1991a** *Corpus Concordance Collocation*. Oxford University Press, Oxford.
- Sinclair, J 1991b** Shared knowledge. In Alatis, pp489-500.
- Sinclair, J 1992** The Automatic Analysis of Corpora. In Svartvik, pp379-397.
- Sinclair, J 1996a** The Empty Lexicon. *International Journal of Corpus Linguistics*. 1 pp99-120.
- Sinclair, J 1996b** The Search for Units of Meaning. *Textus* 9 pp75-105.
- Sinclair, J., J Payne and Pérez Hernandez (eds) 1996** Corpus to Corpus A Study of Translation Equivalence. *International Journal of Lexicography* 9:3.
- Sinclair, J and A Renouf 1988** A lexical syllabus for language learning. In Carter and McCarthy, pp140-160.
- Skutnabb-Kangas, T 1981** *Bilingualism or Not: The Education of Minorities*. Multilingual Matters Ltd, Clevedon.
- Snell-Hornby, M 1988** *Translation Studies. An Integrated Approach*. John Benjamins, Amsterdam.
- Sonyel, S R 1988** *The Silent Minority: Turkish Muslim Children in British Schools*. The Islamic Academy, Cambridge.
- Sperberg-McQueen, C M and L Burnard 1990** *Guidelines for the Encoding and Interchange of Machine-Readable texts*. Text Encoding Initiative.
- Sridhar S N and K Sridhar 1980** The syntax and psycholinguistics of bilingual mixing. *Canadian Journal of Psychology* 34 pp407-16.
- Steele, R and T Threadgold (eds) 1987** *Language Topics - Essays in honour of Michael Halliday*. John Benjamins, Amsterdam.
- Steffens, P (ed) 1995** *Machine Translation and the Lexicon*. Springer, Berlin.
- Stenström, A B 1987** Carry-on signals in English conversation. In Meijs, pp87-119.
- Stillings, N A et al., 1995** *Cognitive Science - An Introduction*. 2nd ed. The MIT Press, Cambridge, Mass.
- Storrer, A and U Schwall 1995** Description and Acquisition of Multiword Lexemes. In Steffens, pp35-50.
- Strzalkowski, T 1994a** Preface. In Strzalkowski, pp xiii-xxi.
- Strzalkowski, T (ed) 1994b** *Reversible Grammar in Natural Language Processing*. Kluwer Academic Publishers, Boston.
- Stubbs, M (ed) 1985** *The Other Languages of England - Linguistic Minorities Project*. Routledge & Kegan Paul, London.
- Stubbs, M 1995a** Collocations and cultural connotations of common words. *Linguistics in Education* 7:4 pp379-90.
- Stubbs, M W 1996** *Corpus Linguistics and Text*. Blackwell, London.
- Svartvik, J 1992a** Corpus linguistics comes of age. In Svartvik, pp7-13.
- Svartvik, J (ed) 1992b** Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82. Stockholm, 4-8 August 1991. Mouton de Gruyter, Berlin.
- Swadesh, M** (see Sherzer)
- Swigart, L 1992** Two codes or One? The Insiders' view and the description of codeswitching in Dakar. In Eastman, pp83-102.

Tamis, A 1986 *The State of Modern Greek Language as Spoken in Victoria*. Ph.D Thesis. The University of Melbourne.

Taylor, J and Young, C 1989 The Turkish Press in Australia. In Ata and Ryan, pp5-23.

Teubert, W 1997 Translation and the Corpus. In Marcinkeviciene and Volz, pp147-164.

Thomason, S 1995 Language Mixture: Ordinary Processes Extraordinary results. In Silva-Corvalán, pp15-33.

Thomason, S 1997 Plenary Speech. *First International Symposium on Bilingualism*. University of Vigo, Spain, 21-25 October.

Thomason, S and T Kaufman 1988 *Language Contact, Creolization, and Genetic Linguistics*. University of California Press, Berkeley .

Tietze, A 1962 Problems of Turkish Lexicography. In Housholder and Saporta, pp263-272.

Tikoo, M L (ed) 1989 *Learners Dictionaries: The State of The Art*. SEAMEO Regional Language Centre, Singapore.

Tognini-Bonelli, E 1996a *The Role of Corpus Evidence in Linguistic Theory and Description*. Ph.D Thesis. The University of Birmingham.

Tognini-Bonelli, E 1996b Towards Translation Equivalence from a Corpus Linguistics Perspective. In Sinclair et al., pp197-217.

Toury, G 1995 *Descriptive Translation Studies And Beyond*. John Benjamins, Amsterdam.

Treffers-Daller, J 1994 *Mixing Two Languages. French-Dutch Contact in a Comparative Perspective*. Mouton de Gruyter, Berlin.

Türker, E 1993 *Fremmedord bilmiyom. A Study of Second Generation Immigrant Turkish in Norway*. M.A. Thesis, University of Oslo, Norway.

Türker, E 1998 The use of yap- in Turkish-Norwegian codeswitching. Panel paper at 9th International Conference on Turkish Linguistics University of Oxford, 12-14 August, 1998.

Underhill, R 1976 *Turkish Grammar*. The MIT Press, Cambridge, Mass.

Vaid, J (ed) 1986 *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives*. Lawrence Erlbaum, New Jersey.

Valdman, A and A Highfield 1980 *Theoretical Orientations in Creole Studies*. Academic Press, New York.

Vallduvi, E 1988 On Lexical and Grammatical Language Mixing. In Ferrara et al., pp368-377.

Van der Linden, E 1994 Review of Bazergui, N, Connors, K, Lenoble, M and Majkrak, B (eds) 1990 *Acquisition du français(L2) chez des adultes a Montreal*. Office de la langue française. *Applied Linguistics* 15:2 pp233-236.

Van der Meer, G 1998 Collocations as one particular type of conventional word combinations. Their definition and character. In Fontenelle et al. pp313-322.

Van Els, T and B Weltens 1989 Foreign Language Loss Research from a European Point of View. *Review of Applied Linguistics* 83 pp19-35.

Van Hout, R and P Muysken 1994 Modeling lexical borrowability. *Language Variation and Change* 6 pp39-62.

Verhoeven, L 1991 Acquisition of Turkish in a Mono- and Bilingual Setting. In Boeschoten and Verhoeven, pp113-149.

Vermes, G (ed) 1988 *Vingt-cinq communautés linguistiques de la France* L'Harmattan, Paris.
Vildomec, V 1963 *Multilingualism*. A W Nijhoff, Leiden.

Wallace, K 1988 *Morphology as a Computational Problem*. UCLA Occasional Papers no7 Working Papers in Morphology. Dept. of Linguistics, UCLA.
Wande, E et al. (eds) 1987 *Aspects of Multilingualism. Proceedings from the 4th Nordic Symposium on Bilingualism, 1984*. Borgströms, Motala, Sweden.
Watson, J L 1977 *Between Two Cultures: Migrants and Minorities in Britain*. Blackwell, Oxford.
Weertz, J and E McClure 1977 Monolingual 'codes': Some remarks on the similarities between bilingual and monolingual code-switching. In Beach, *et al.*, pp706-713.
Weinreich, U 1953 *Languages in Contact*. Mouton & Co., The Hague.
Weinreich, U., W Labov and M I Herzog 1968 Empirical foundations for a theory of language change. In Lehmann and Malkiel, pp95-188.
Weltens, B, K de Bot and T Van Els (eds) 1986 *Language Attrition in Progress*. Foris, Dordrecht.
Winter, E 1977 'A clause relational approach to English texts: a study of some predictive lexical items in written discourse' *Instructional Science* 6:1 pp1-91.
Woolford, E 1983 Bilingual code-switching and syntactic theory. *Linguistic Inquiry* 14 pp520-535.
Wright, S (ed) 1995 *Languages in Contact and Conflict. Contrasting Experiences in the Netherlands and Belgium*. Multilingual Matters, Clevedon, UK.

Yaǎmur, K 1993 Ethnolinguistic vitality perceptions: intergenerational differences among Turkish-Australians. In Akçelik, pp97-120.
Yaǎmur, K 1997 *First Language Attrition Among Turkish Speakers in Sydney*. Ph.D Thesis. Tilburg University Press.
Yellow Pages Australia 1997 Pacific Access Pty. Ltd.
Young C and J Taylor 1985 *The Turkish and Yugoslav Press: a survey of content and readership of ethnic Newspapers in Melbourne*. Australian Government Publishing Service, Canberra.

Zubrzycki, J 1958 The Role of the Foreign-Language Press in Migrant Integration. *Population Studies* 12:1: 73-82.

GLOSSARY

colligation

This notion was first coined by Firth (1957:13):

The statement of meaning at the grammatical level is in terms of word and sentence classes or of similar categories and of the inter-relation of those categories is colligation. Grammatical relations should not be regarded as relations between words as such - between 'watched' and 'him' in 'I watched him' - but between a personal pronoun, first person singular nominative, the past tense of a transitive verb and the third person singular in the oblique or objective form.

and has been explored further only recently (Hoey 1998):

I suggest that colligation can be defined as the grammatical company a word keeps. Just as a lexical item may have a strong tendency to co-occur with another lexical item, so also that lexical item may have an equally strong tendency to occur in a particular position or (a separate point) to co-occur with a particular grammatical category of items.

collocation

In its broadest sense collocation is defined as the lexical co-occurrence of words. In this study, this notion is used to explain the way L1 and L2 items co-select each other out of a number of possibilities.

context (span)

Words that appear on either side of a word under examination form the linguistic environment or the context of that word.

concordance

A computer-generated index of words in a text which enables the linguist to observe significant language patterns in texts.

corpus

A machine-readable collection of spoken and written natural language texts representing a state, or variety of a language.

lemma

A lemma is a "composite set of word-forms" (Sinclair 1991:173). That is, a word usually has different forms and a lemma comprises of all the different word-forms of a word.

lemmatization

Lemmatization is the process of determining all the forms of a word and arranging them into lemmas.

node

The word whose lexical behaviour is under examination in a collocation, usually highlighted in bold type in a concordance line, is called the node.

running words

Refers to the length of a text where each successive word-form is counted once.

text

A text is a piece of spoken or written language which is complete and continuous.

word-form

Words in a language have several actual forms. For example, the verb give in English has the separate word-forms *give*, *giving*, *gave*, *given* and so on.