

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]



Université d'Ottawa • University of Ottawa

**ASPECTS OF SECOND LANGUAGE SPEECH: A VARIATIONIST
PERSPECTIVE ON SECOND LANGUAGE ACQUISITION**

by

Reza Ghafarsamar

Thesis submitted to
The Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for
the degree of
Doctor of Philosophy in Linguistics

*January 2000
University of Ottawa*



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-57041-X

Canada

ABSTRACT

The work reported in this study is based on empirical data collected from three groups of speakers: native speakers of English, native speakers of Persian and Persian speakers of English as a second language.

The study addresses, among others, two of the most important aspects of second language speech: 1) it provides an extensive sociolinguistic description of L2 speech by conducting multivariate analyses on features of English spoken by Persians. In this way it throws light on the issue of the systematicity of second language speech by providing evidence for systematic variation in L2, 2) it explores the relationship between native language of the speakers and their L2, on one hand, and the relation between the L2 and the language L2 learners try to acquire, on the other. In other words, it examines the sources of the patterns of variability found in second language speech.

The focus of our investigation is the variable contraction (deletion) of auxiliary verbs and variable use of relative and resumptive pronouns by L2 learners. We first use the principles of variation theory to make a detailed assessment of the behavior of the L2 learners with respect to the contraction of auxiliary verbs in their natural L2 speech at both low and high proficiency levels. The comparative method, then, is applied to systematically compare patterns of contraction in low and high proficient L2 speech with: a) patterns of contracting auxiliaries in our native English data and, b) patterns found in the samples of Persian speech produced mostly by the same L2 informants. Then, we extend the same procedures to examine variable use of relative and resumptive pronouns in restrictive relative clauses in the three contexts of English, Persian and L2 (high proficiency only).

Our results produced conclusive evidence that the variability found in the features of L2 speech studied in this thesis was indeed systematically conditioned by at least one linguistic factor group. There was found to be no difference between high and less proficient learners in this regard.

With respect to the sources of variability in L2 speech, we consider at least three possibilities: 1) that the variation can be explained by processes derived from English as the target language, 2) that the variation can be explained by the processes derived from Persian as the native language of the learners, and 3) that the variation and its conditioning system is unique to our learners' L2 speech, different from native or target languages. Parallel analysis and comparison of features of L2 and those of the informants' native and target languages demonstrated that variability in advanced second language speech was mostly conditioned by factor groups and factor weights that constrain native English speech. Informants' native language was found to play a trivial role in this respect. Less proficient learners, however, relied more on their native patterns of variability in cases where they have not acquired the target language patterns of variation yet. It was also demonstrated that patterns of variability change as learners advance in their L2 development. We did not observe any pattern unique to our informants' second language behavior, different from either English or Persian.

*Love rests on no foundation.
It is an endless ocean,
with no beginning or end.
Imagine,
a suspended ocean,
riding on a cushion of ancient secrets.
All souls have drowned in it,
and now dwell there.
One drop of that ocean is hope,
and the rest is fear.*
Rumi (Persian Poet)

I dedicate this to those who taught me the meaning of love, dedication and being
a 'human':

To Imam Khomeini, Iranian people and to all my teachers

ACKNOWLEDGEMENT

Among so many people whom I have become indebted to in the pursuit of my academic career, the late Dr. Mohammad Hadi Kamiabee was the one who made me understand the beauty and meaning of life and literature. I express my sincere gratitude to him and ask Allah bestow His blessings on his soul.

Professor Shana Poplack introduced me to sociolinguistics and broadened my knowledge of and interest in linguistics as a science. Her provocative teaching style, rigorous research training procedures, hard work, and ability to inspire excellence has provided her students with the most stimulating and challenging intellectual environment. I thank her for all and every single word she taught me. It was an exceptional privilege to have the opportunity to work with a scientist of such caliber.

I have also gained invaluable knowledge and experience through my participation in various courses offered by the faculty of the Department of Linguistics, University of Ottawa, among which were: courses in sociolinguistics, variation theory, bilingualism (S. Poplack), Psycholinguistics (P. G. Patel), Syntax (M-L. Rivero), L1 acquisition (H. Goodluck), Semantics (P. Hirschbuhler), L2 acquisition (G. Neufeld), and Phonology (H. Hung). I appreciate Helmut Zobl, Phil Hauptman, Jerald Neufeld, Eta Schneiderman, and Shana Poplack for their comments that led to a substantial improvement of my thesis. I myself, of course, am responsible for any remaining errors.

I am also indebted to all my informants, Persian and English. I am thankful for their gift of time and sharing which produced the interviews from which I have extracted my data and learned so much.

This dissertation could also not have been completed without the emotional and psychological support of friends and family. I would like to express my appreciation, among others, to Ahmad Moinszadeh, Dawn Harvie, James Walker, Jennifer Houghton and Gerard Van Herk for their support and comments on my data and thesis chapters. I have also enjoyed the friendship of several graduate students of the Department of Linguistics including: Olga Arnadova, Ejike Eje, Michelle Foley, Abdul-Hamid Godou, Stephen Goyett, Margarete Ling, Elizabeth Malvar, Laura Proctor, Vera Rolof, Danijela Stojanovic, and Mohammad Vahedi.

My family has been the underlying bedrock of stability and strength for the duration of my studies here and in Iran. Without their support and love none of it could have happened. My thanks to my wife, Maryam, for her love, support and sacrifices, to my daughter and son, Sara and Salman, for cheering up a usually busy and grumbling dad. I am also grateful to be blessed with supportive, loving and patient parents and in-laws. I thank them all from the bottom of my heart.

Several individuals and bodies provided financial support during my studies in Canada. Ministry of Culture and Higher Education of the Islamic Republic of Iran was the main source of funding for my studies. Department of linguistics also helped me with a couple of research assistantships for which I owe the most to P. Hirschbuhler. When all the sources of support were dried up, emotional and financial supports of my wife helped me to complete this study. My many thanks to them all.

Finally and most importantly, I thank God the Almighty for showering me with His blessings. This work has never been completed without those blessings.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION	5
1.1. BACKGROUND.....	5
1.2. SOURCES OF VARIATION IN SECOND LANGUAGE SPEECH.....	10
1.3. HYPOTHESES	12
1.4. GOALS	14
CHAPTER TWO: PREVIOUS STUDIES OF VARIATION IN L2	16
2.1. VARIATION IN SECOND LANGUAGE SPEECH	16
2.1.1. VARIATION: A FEATURE OF PERFORMANCE	16
2.1.2. VARIATION: AN INTEGRAL PART OF SLA.....	17
2.1.3. TYPES OF VARIABILITY: SYSTEMATIC/NON-SYSTEMATIC VARIABILITY	18
2.1.4. STUDIES OF VARIATION IN L2	19
2.1.4.1. Constraints on variation	20
<i>Linguistic context</i>	20
<i>Effects of topic</i>	21
<i>Effects of interlocutor</i>	21
<i>Effects of task</i>	22
2.1.4.2. Current research	24
<i>Adamson & Regan 1991, Adamson et al 1996, Regan 1996</i>	24
<i>Young 1991, 1996</i>	26
<i>Bayley 1994, 1996</i>	28
2.2. HYPOTHESES OF L2 ACQUISITION	29
2.2.1. THE TRANSFER HYPOTHESIS	29
2.2.2. THE INTERLANGUAGE HYPOTHESIS	31
2.3. SLA: DATA AND THE IMPORTANCE OF EMPIRICAL STUDY	33
CHAPTER THREE: METHODOLOGY AND DATA.....	38
3.1. THEORETICAL FRAMEWORK AND METHOD OF ANALYSIS.....	38
3.1.1. VARIABLE RULE ANALYSIS AND VARBRUL PROGRAM	41
3.1.2. COMPARATIVE METHOD	43
3.2. THE INFORMANTS.....	46
3.2.1. SELECTION OF INFORMANTS	46
<i>L2 background of the Informants</i>	47
3.2.2. PERSIAN INFORMANTS	50
<i>The Persian community in Ottawa</i>	50
<i>Persian Informants</i>	50
3.2.3. L2.....	50
<i>Assessment of English proficiency</i>	50
<i>L2: proficient informants</i>	51
<i>L2: less proficient informants</i>	51
3.2.4. ENGLISH	52
3.3. THE DATA	52

3.3.1. DATA COLLECTION AND CODING PROCEDURES	53
3.3.2. THE CORPORA	54
3.4. THE VARIABLES	55
<i>Auxiliaries</i>	55
<i>Relative clauses</i>	57
CHAPTER FOUR: AUXILIARY VERBS	60
4.0. INTRODUCTION	60
4.1. AUXILIARIES IN ENGLISH, PERSIAN, AND EL2	61
4.1.1. ENGLISH AUXILIARIES	61
4.1.2. PERSIAN AUXILIARIES	62
4.1.3. AUXILIARIES IN EL2	65
4.2. ANALYSES	66
4.2.1. VARIABLE CONTEXTS AND EXCLUSIONS	66
<i>English and L2</i>	66
<i>Persian</i>	67
4.2.2. CODING	68
4.2.2.1. Phonological environment	69
4.2.2.2. Type of subject	69
4.2.2.3. Following grammatical category	70
4.2.2.4. Underlying form of the auxiliary	71
4.2.3. DATA	72
4.2.4. RESULTS	74
<i>AM</i>	74
<i>IS</i>	77
<i>ARE</i>	82
<i>HAVE/HAS/HAD</i>	86
<i>WILL/WOULD</i>	89
4.3. DISCUSSION	91
4.4. SUMMARY OF CHAPTER FOUR	98
CHAPTER FIVE: RELATIVE CLAUSES: RELATIVE PRONOUNS 99	
5.0. INTRODUCTION	99
5.1. RELATIVE MARKERS IN ENGLISH, PERSIAN, AND EL2	100
5.1.1. RELATIVE MARKERS IN ENGLISH	100
5.1.2. RELATIVE MARKERS IN PERSIAN	102
5.1.3. RELATIVE MARKERS IN EL2	104
5.2. ANALYSES	106
5.2.1. VARIABLE CONTEXT, EXCLUSIONS AND CODING	106
5.2.2. FACTOR GROUPS	108
5.2.2.1. Semantic factor groups	109
<i>Definiteness of the head noun</i>	109
<i>Humanness of the head noun</i>	110
5.2.2.2. Syntactic factor groups	111
<i>Premodification of the head noun</i>	111
<i>Syntactic function of the NP head in the matrix clause</i>	112
<i>Syntactic function of the relative marker in the relative clause</i>	113

<i>Category membership of the subject of the relative clause</i>	113
5.2.2.3. Formal factor groups	114
<i>Adjacency of the head noun and the relative marker</i>	114
<i>Length and position of the relative clause</i>	115
5.2.3. DATA	116
5.3. RESULTS	116
5.3.1. MARGINALS	119
5.3.1.1. Definiteness of the head noun	119
5.3.1.2. Adjacency of the head noun and the relative marker	121
5.3.1.3. Humanness of the head noun	122
5.3.2. VARIABLE RULE ANALYSES	123
5.3.2.1. Interactions	123
5.3.2.2. Variants and contexts	127
5.3.2.2.1. EL2 system of relative marking	127
5.3.2.3. Variants and contexts: EL2 compared to Persian and English	129
5.3.2.3.1. Zero	129
5.3.2.3.2. That	133
5.3.2.3.3. WH-forms	137
5.4. DISCUSSION	139
CHAPTER SIX: RELATIVE CLAUSES: RESUMPTIVE PRONOUNS	144
6.0. INTRODUCTION	144
6.1. RESUMPTIVE PRONOUNS IN ENGLISH, PERSIAN, AND EL2	145
6.1.1. RESUMPTIVE PRONOUNS IN ENGLISH	145
6.1.2. RESUMPTIVE PRONOUNS IN PERSIAN	145
6.1.3. RESUMPTIVE PRONOUNS IN THE ENGLISH OF PERSIAN SPEAKERS	147
6.2. ANALYSES	150
6.2.1. VARIABLE CONTEXT AND CODING	150
6.2.1.1. Syntactic factor groups	151
<i>Syntactic function of the relative marker in the relative clause</i>	151
<i>Syntactic function of the head noun in the matrix clause</i>	152
6.2.1.2. Formal factor groups	153
<i>Adjacency, length and position of the relative clause</i>	154
6.2.1.3. Semantic factor groups	155
<i>Definiteness and humanness of the antecedent</i>	155
6.2.2. DATA	156
6.2.3. RESULTS	157
6.2.3.1. English	157
6.2.3.2. Persian	158
6.2.3.3. EL2	160
6.3. DISCUSSION	161
6.4. SUMMARY OF CHAPTER SIX	167
CHAPTER SEVEN: DISCUSSION OF RESULTS AND CONCLUSIONS	168
7.1. FINDINGS OF THE STUDY	168

7.2. LIMITATIONS OF THE STUDY	174
--	------------

Appendix A

Appendix B

References

Chapter one: INTRODUCTION

1.1. Background

One issue that has recently been the subject of widespread debate in Second Language Acquisition (hereafter SLA) is the existence of variability in the speech of learners of second languages (Beebe 1988; Ellis 1987a, 1994a). Once this variation is recognized, the question arises as to whether such variation is random or if it is governed at least in part by rules which are discoverable by the researcher. Since the 1960s, when Selinker (1969) first put forward the proposition that the variation found in second language speech is in fact systematic, describing and explaining the highly variable speech of learners of second languages has been one of the abiding problems for SLA theory (Bayley 1996; Young 1991). Several researchers have found that L2 variation is indeed systematic, while others (e.g., Ellis 1985, 1991, 1994a, 1994b; Towell et al. 1993) insist on the existence of unsystematic variation in L2. Of many variable features of second language speech, only a few have been subject to multivariate (variationist) analysis, and almost all have been found to be systematically conditioned by at least some linguistic factors.

It is now believed that variation in L2 speech, systematic or not, is crucial to explaining the natural route and mechanisms of SLA, and change in learners' L2 knowledge over time. It is considered an important source of information about the way in which interaction in different social contexts, for instance, can influence both L2 use and its development. Explaining variation, then, not only helps researchers to describe the grammar of second language speech (facts of L2 speech), but also assists in explaining the nature of acquisition itself (Adamson 1988). Dealing with this variation is, therefore, not a marginal pursuit but an obligatory part of SLA research (Andersen 1989). Studies of L2 speech variability, in this respect, have contributed a great amount to the study of both systematicity of second language speech and its acquisition/development.

There are good reasons to believe that the speech of second language learners is more subject to change, development, and variation than the varieties of speech found in native (L1) speaking speech communities. In the 1950s and early 1960s it was believed that the main causes of L2 variability (then called 'errors') had their

roots in the influence of second language learners' native language on L2 learning. Later, after failure of contrastive analysis to predict errors which occurred in L2 speech based on L1 influence, SLA theorists posited that what had previously been attributed to influence from the learners' L1, is in fact, to some extent, independent of either first or target languages (Corder 1967, Nemser 1971; Selinker 1969, 1972). From that time on the notion of the systematicity of *interlanguage* (Selinker 1972) has become almost axiomatic in the contemporary field of SLA (Young 1991). Although modern multivariate analyses of second language speech (Bayley 1994, 1996; Regan 1996; Young 1991) have found both inherent systematicity in L2 and constraints on it, one issue that variationists in SLA have not dealt with yet is the exploration of the sources of variation in L2 speech. If variable use of second language features is indeed systematic, what are the patterns governing this variation? Do these patterns rely on the learner's L1 system of variation (transfer), are they created by L2 learners, independent of and different from either native or target languages, or are they acquired from the target language(s)? How can we demonstrate if a variable feature of a target language is acquired by its L2 learners? And if variation is acquired, are the constraints on variation in the target language acquired along with it? If it is assumed that patterns of variation are acquired from the target language, an immediate question to be answered will be if SLA is acquisition of categorical rules (Gregg 1989) or involves the acquisition of target language variable rules* (Ellis 1987a).

In this dissertation I make an attempt to tackle these issues. My quantitative examination of variation in second language speech adds to the bulk of evidence on the nature of L2 variability and contributes to the unresolved issue of the systematicity of second language speech. Further, exploring the acquisition of variability will shed more light on the processes of second language development.

The theoretical approach on which this study is based falls within the framework of variationist sociolinguistics. Variation theory (Labov et al. 1968; Poplack 1993; Sankoff 1982, 1988a, among many others) involves the scientific investigation of language use in natural context. This theory holds that in the course of linguistic

* Following Ellis (1987b) I assume that there are both categorical and variable rules in language.

performance, speakers are engaged in a multitude of choice processes among different alternatives which have the same referential value or grammatical function (Sankoff 1988a). Alternate use of these choices is conditioned by particular features in the “phonological environment, the syntactic context, the discursive function of the utterance, topic, style, situation and personal and/or sociodemographic characteristics of the speaker or other participants” (Sankoff 1988a; Tagliamonte 1991: 18). The alternate forms are referred to as *variants* and the larger unit which they are a part of is called the *linguistic variable*. The set of related factors that may condition a choice process constitute a *factor group*. The hierarchical order of the factors in a factor group is referred to as the *constraint hierarchy*. The systematic influence of different factors on the choice process and their hierarchy form a quantitative *pattern* of occurrence. This study is an attempt to find if there is any such pattern of variation in second language speech and where this very pattern of variation in L2 (if any) comes from. In other words, it is the origin of the *pattern* of variation that is the main focus of this thesis (not the presence or absence of linguistic structures only).

I start my inquiry from L2 speech itself by first examining the variation and its conditioning factors, and then I pay particular attention to the parallels and differences between patterns of variation in the informants’ L2 and their native and target languages. In addressing the problem of the source of variation in SLA data, therefore, I have chosen to look at the speech of the informants in terms of three possibilities: 1) variation is transferred, 2) variation is created by L2 learners, and 3) variation is acquired.

In order to achieve these objectives, I have chosen to perform a detailed quantitative sociolinguistic analysis on the use of two uninvestigated variable structures in L2, namely relative clauses and the variable contraction/deletion of auxiliary verbs, in audio-recorded corpora of L2 speech from Persian learners of English as a second language. Auxiliary verbs, in general, are an essential component of most well-formed interrogatives, negatives, indicatives, ellipsis, emphatics, etc. They are among the earliest (and the easiest) grammatical morphemes acquired by first and second language learners (Dulay & Burt 1973; Krashen 1977; Lightbown 1987, for English). Despite their early acquisition, scarcity of auxiliary contraction in

EL2 speech is one of the distinctive features of “foreigner talk”¹, easily detected by native speakers of English (Meechan 1996). However, because of the lack of studies which concentrate only on the acquisition and/or variable use of forms of auxiliary verbs in SLA, much of our knowledge comes from investigations into areas in which the auxiliary was not the primary focus of attention. These include studies of negation, question formation, past and perfect marking, etc. The focus of this thesis, in this part, will be to compare and contrast the frequency and constraints on the variable contraction (and deletion, in some cases) of the copula and auxiliary verbs in English spoken by Persian speakers with those found in their native (Persian) and target (English) languages.

One point that some researchers have acknowledged is the relation between proficiency and transfer on the one hand, and proficiency and variability on the other. It has been argued that less proficient learners rely more on transfer (Taylor 1975, cited in Odlin 1989) and as the learners’ target language knowledge increases, they draw less on their native language. Young (1991) has also studied variable plural marking patterns in the speech of second language learners of English with lower and higher proficiency levels and found that patterns of variation change as acquisition proceeds. A variationist study on the variable features of less proficient learners’ L2 speech, compared to that of advanced learners’, can both throw light on the effect of proficiency on language transfer and show if the level of L2 proficiency exerts any effect on the L2 system of variation.

Mention was made that auxiliary verbs (at least some forms of them) are acquired at the early stages of L2 learning (e.g., Krashen 1977); therefore, different forms of auxiliary verbs (forms of *BE*, for instance) are frequently used by *less* proficient learners of English. I will analyze the variable use of auxiliary verbs by a group of less proficient speakers, and compare its results with those of other analyses. A comparison of the patterns of variable contraction of auxiliaries by less proficient L2 learners with those of a group of advanced learners will demonstrate whether the patterns of variation in L2 speech change as acquisition proceeds. Moreover, comparison of patterns of variation in L2 speech with the patterns of variability in

¹ ‘Foreigner talk’ here means L2 spoken by foreigners.

native and target languages may indicate what the effects of these languages are on L2 variation.

Acquisition of relative clauses, on the other hand, has been one of the most widely investigated structures in adult L2 research (e.g., Gass 1979; Gass & Ard 1980, 1984; Tarallo & Myhill 1983; Flynn 1989). Relative clause constructions are a part of syntax that goes beyond phrase structure and comprises a complex area, the acquisition of which may occur in the later stages of L2 acquisition with a slower rate of progress. Moreover, it has been claimed that acquisition of different forms of relative clauses (use of relative markers, in particular) may indicate the relative proficiency of the language learners (Yorozo 1995). Variable use of elements of these constructions (relative markers, for instance) have also been the topic of several variationist studies (e.g., Guy & Bayley 1995) in native English. Because of these properties, relative clauses provide an important domain of investigation of L2 learning. In order to discover patterns of variation in relative clause usage in L2 and the effects of learners' native and target languages on this variation, I will perform a quantitative analysis on the variable use of relative and resumptive pronouns in the English spoken by Persians, and compare the results with those obtained from similar analyses on natural speech of native Persian and English speakers. Since relative clause constructions are rarely used by less proficient L2 speakers (only a few tokens per some informants), only the L2 speech of advanced learners will be incorporated in this part of the study.

Persian and English relative clause constructions are both post-nominal and finite, while the nature of relative marker usage and resumptive pronoun marking is different in these languages. Variable contraction of present forms of *BE* is common in both languages. English, however, contracts auxiliary forms of *HAVE* in present perfect, while Persians tends to delete them. A study of the behavior of Persian learners of English with respect to these linguistic variables can demonstrate what direction second language learners tend to take when the native and target language patterns parallel or contrast each other. Contrasting sites are of particular interest since this is the context where the effects of either native or target languages can be clearly discovered. For instance, if it is found that Persian uses more relative markers after nominal subjects and less after pronominals, but English favors them after pronominal subjects, then I will conclude that L2 learners

transfer their L1 patterns to L2 if they are found to use more relative markers after nominal subjects, or, on the contrary, they have acquired English patterns if they are found to favor more relative markers after pronominals.

The corpora I examine in this dissertation consist of the speech of 12 proficient and 10 less proficient Persian learners of English, 10 native speakers of English and 17 native speakers of Persian. Because of the uniform linguistic and extra-linguistic characteristics of the speakers in each corpus, the corpora will hopefully provide a reliable contribution to the issues investigated in the present study.

1.2. Sources of variation in second language speech

The first possible source of variability in L2 may be assumed to be variability in the target language, i.e., variation is acquired. Another option is that L2 is an independent interlanguage system which has its own patterns and mechanisms of variability, different from those of the target language, i.e., created by speakers as an *interlanguage*. Looking for the effects of the target language on L2 variation involves a variationist examination of certain variable features of L2, and comparison of the results to the behavior of the same features in the target language. Similarity or differences of the patterns of variation in two contexts will demonstrate not only the extent (or lack) of the effects of the target language variability on L2 speech but will also determine what is really acquired and what is not. As an example, variability between the presence or absence of a linguistic item in L2, e.g., a relative marker, may indicate not lack of acquisition, rather native-like acquisition, if this variability and the constraints on it mirror that of the target language.

A third option as a source of L2 variability is the native language of the L2 learners, i.e., variability is transferred. This phenomenon of first language influence on the learning (and production) of a second language has been recognized for centuries (e.g., Sweet 1899, cited in Odlin 1989; Gass 1996); and it has been at the center of every model proposed for SLA². It is now emphasized that accounting for the learners'

² Contrastive Analysis (Lado 1957), for instance, had crosslinguistic influence as its major concern; and the Interlanguage Hypothesis (Selinker 1972) dealt with it as one of the main processes of second language acquisition. Even for the models that had denied a major role for L1 influence on L2 development (e.g., Creative Construction theory (Dulay & Burt 1972)) it took a great amount of research to provide support for the idea that transfer plays little role in L2 development.

prior linguistic knowledge is an integral part of any theory of L2 acquisition, and that no theory of L2 acquisition that ignores the learners' prior linguistic knowledge can be considered complete (Ellis 1994a; also Gass & Selinker 1994). Although the evidence for L1 influence has grown over the last four decades, 'understanding the nature of the transfer process remains one of the most formidable challenges facing second language researchers' (Odlin 1994: 29).

To examine the effects of either native or target language on the patterns of L2 variation, comparison of at least two linguistic systems (L1-L2 and/or L2-target) seems unavoidable; and this necessitates some kind of contrastive analysis (Gass & Selinker 1983, 1992; Odlin 1989). How can one compare and contrast patterns of languages involved in a learning situation without being trapped in the "comparison of the category inventories of the languages in contact" (Labov 1996: 246), i.e., traditional contrastive analysis? Remember that this kind of contrastive analysis has been under criticism for decades because of the challenges faced by contrastivists concerning the predictive power of their method and the relation between first and second language acquisition (Odlin 1989).

One of the main problems faced by transfer (contrastive analysis) researchers has been to find an appropriate method of comparing languages involved in a SLA situation (the problem of comparison) (Ellis 1994a; Odlin 1989). Different procedures have been suggested (e.g., first describe the structures of the languages to be contrasted by means of one and the same theoretical models, then contrast these descriptions for the specification of similarities and dissimilarities (Sajavaara 1981: 37)) and used by several researchers (cf. Sajavaara & Lehtonen 1981; Dechert, Bruggemeir, and Futterer 1984), but there is little agreement about which constitutes the optimal method (Ellis 1994a; Odlin 1989). As Odlin asserts, "since native language influence interacts with nonstructural factors"³ (Odlin 1989: 28), for a thorough understanding of language transfer more than just structural comparisons (or according to Labov 'comparison of the category inventories (grammatical features, for instance)' (1996)), prevalent in contrastive analysis, are necessary.

³ Discourse or treatment of the topics in discourse, for instance (Odlin 1989: 31).

The work on which this study is based employs the methods of variationist sociolinguistics. Here, first, systematic quantitative analyses of rather large corpora of spoken language from native, target and L2 speakers provide accountable data bases from which the actual frequency, distribution and conditioning of variable forms is extracted. Second, patterns of variability in L2 are compared to those of native and target languages with respect to phonological, syntactic and semantic factors. Phonological factors can be subject to universal phonological processes (e.g., preference for CVC syllable structure) or certain performance factors such as difficulty of articulation of certain L2 sounds by L2 learners (e.g., difficulty in articulating English /r/ or /θ/ (as in *threw*) by Persians). Syntactic (and semantic) factors are, therefore, of more importance since they are more relevant to the underlying grammar of the learners. My comparisons, unlike traditional contrastive analysis, go beyond simple comparison of surface structures (Odlin 1989) or category inventories (presence or absence of linguistic features and their frequency), and take into consideration *patterns* of variability.

Mention was made that one of the major problems of contrastive analysis has been to find an optimal method of comparison that takes both structural and nonstructural factors into consideration. Comparison of patterns of variability where all structural and nonstructural factors (discursive, semantic, etc.) might quantitatively be dealt with, can be of great help in solving this problem of contrastive analysis that according to Labov (1996) 'remains the most promising route for those who want to apply a concrete knowledge of language structure to the problem of language learning' (P. 246).

1.3. Hypotheses

Sankoff (1988a, among others) believes that to arrive at a comprehensive understanding of the underlying structure of a given grammar, e.g., second language speech, one way is to examine the features of the linguistic context which influence the choice speakers make between alternative forms in their grammatical repertoire. Furthermore, parallel analysis and comparison of these features among different languages in a contact (learning) situation can help us infer the relationship between them. My basic assumption in this thesis concerning the grammar of L2 speech is that the underlying grammatical structure of second language speech (with respect to the features studied here) is discernable from an examination of the distribution and conditioning on

the variable use of those features in the samples of conversational speech from English spoken by Persians. The null hypothesis in variable rule analysis is that “none of the factors examined has any systematic effect on the choice process” (Sankoff 1988b: 987). If it is found that a single factor or group contradicts the null hypothesis by showing a real effect on the data, then we can prove that “random processes alone are unlikely to have resulted in the patterns of proportions observed” (Sankoff 1988b: 987). Hence, with respect to the systematicity of L2 speech I hypothesize that:

- ***Hypothesis one:*** if, in a given context, it is found that variable use of linguistic variables is not systematically conditioned by at least one conditioning factor, my interpretation will favor existence of non-systematic variation in L2 speech.

Thus for the purpose of this study, *system* refers to the patterns of relation between the language forms and the contexts in which they occur (see also Young 1991). Whenever this relation holds between L2 forms and their contexts, it will be argued that L2 variation is systematic.

With respect to the sources of variability in L2 speech, I consider at least three possibilities: 1) that the variation can be explained by processes derived from English as the target language, 2) that the variation can be explained by the processes derived from Persian as the native language of the learners, and 3) that the variation and its conditioning system is unique to the learners' L2 speech, different from native or target languages. Parallel analysis and comparison of features of L2 and those of the informants' native and target languages can help us demonstrate which of these options are operational in L2. According to these possibilities, the following hypotheses are made:

- ***Hypothesis two:*** if it is found that, in a given context, EL2 variation is conditioned by similar factors (factor weights and constraint hierarchies⁴) as those of English (and these simultaneously differ from Persian), I hypothesise that that feature of the target

⁴ In an ideal situation factor weights and constraint hierarchies will be similar, as well.

language has been acquired by the learners from their target language. Transfer effects hardly exist in these cases.

- **Hypothesis three:** if the results of the analyses show that in a given context L2 variation is conditioned by factors similar to those of Persian, and different from English, I will infer that language transfer is the source of variation in that context.
- **Hypothesis four:** if it is found that L2 variation is conditioned by factors independent of and different from Persian or English, my interpretation will be in favour of an L2-specific pattern of variation. This will suggest that L2 speech is an independent linguistic system (interlanguage) with its own system of variation.

The findings of this study with respect to hypothesis two will contribute to our understanding of whether the acquisition of variable forms implies acquisition of the system of variation that operates in the target language. Crosslinguistic influence, according to the third hypothesis, will not only involve transfer of surface linguistic features of L1 but also patterns of variation that are operative in the native language of the L2 learners. And finally, if it is found that L2 has its own independent patterns of conditioning variable use of linguistic features, I will be able to confirm, on an empirical basis, the existence of an interlanguage system with its own patterns of variation in the contexts examined here.

1.4. Goals

The first major goal of this thesis, then, is to describe in detail the system of relative and resumptive pronoun marking and variable use of contracted (deleted) and full forms of auxiliaries/copula in second language speech of the Persian informants. I will also examine the influence of phonological, syntactic, discursive, and semantic factors on the variable use of these items in a quantitative manner. This work will expand our knowledge of the structure and the system of variation in L2 speech and will contribute to a more comprehensive understanding of L2 grammar. The second major goal is to compare the results of my quantitative analyses on variation in L2 to those of native English and native Persian, target and native languages of the informants respectively. I hope that my detailed characterization of the system of L2 speech and my findings with respect to the differences and similarities of L2 patterns of variation to those of Persian

and English will be a contribution to the field of SLA. This will also indicate both the importance of studies of variability in second language speech and the advantages of application of recent variationist methods to SLA data.

In subsequent chapters I will follow these goals in greater detail. In chapter two I review the literature on variation in second language speech, and transfer and interlanguage hypotheses in SLA research to situate the problems I will address. I will also provide a short argument for the utilization of naturalistic data in this study. Chapter three presents information about the theoretical framework adopted in this dissertation, methodology, and data. Chapter four is devoted to the analysis of the behavior of auxiliary verbs in EL2 and comparison of the results to the findings of similar analyses on native Persian and English data. In chapters five and six I present the findings of this study on relative pronoun and resumptive pronoun usage respectively. Chapter seven includes my discussion on and interpretation of the finding of chapters 4-6.

Chapter two: PREVIOUS STUDIES OF VARIATION IN L2

2.1. Variation in second language speech

All natural languages display variability in at least some of their structures, be it phonological, morphological, syntactic, etc. Variability in second language speech has also been generally acknowledged by studies of SLA. This variation may be transitory: it can disappear or change in its rate and frequency as the learner progresses from one stage to the next stage in his/her L2 development (vertical variability). Variation can also be a permanent feature of L2 speech when the learner comes to an end in his/her L2 development: either a target (variable) form is acquired or her/his language is fossilised (horizontal variability). At this stage of development L2 speech resembles other natural languages in which variability is integral (Ellis 1994a, Adjemian 1976). The question is how SLA researchers view this variability. Depending on the view of the theory of language/linguistics held, there seems to be two opposing approaches to the L2 variability in the literature. One is to simply ignore variability, and the other one is to try to account for it in a principled way.

2.1.1. Variation: a feature of performance

The general claim in the first approach is that in order to study language, in general and L2 in particular, one should abstract what learners “know” from what they “do”. The linguist can investigate the linguistic knowledge (competence) of the speaker if he gains access to invariable data by idealizing the actual utterances of an ideal speaker-hearer in a homogeneous speech community (Gregg 1989, 1990; White 1989a). Variability, then, is a feature of performance (what speaker does), and because the goal of the linguist is to describe the learner’s competence rather than his/her actual use of that knowledge, variability is ignored. This position is clearly evidenced in the following quotation from Gregg (1989): “... if we are careful to establish the domain of a theory of second language acquisition so that it is confined to the acquisition of linguistic competence, then we will not be compelled to account for those data on variability as far as that theory is concerned” (Gregg 1989: 22). According to this perspective, variable patterns “cannot, by their nature, be acquired” (Gregg 1989: 21) at all and SLA involves acquisition of categorical rules only.

The results of this thesis lend little (if any) support to this claim by demonstrating how variable patterns are learned by Persian informants with the same rates and constraint hierarchies as those of native speaker English, the informants' target language.

2.1.2. Variation: an integral part of SLA

For the proponents of the second position, variation in L2 speech is crucial to explaining the mechanisms of SLA, or change in learners' L2 knowledge over time (Regan 1996). Study of variation, therefore, provides important information about the use and development of second language in different contexts. A reliable account of L2 variation informs researchers of the facts of L2 speech (descriptive adequacy), and assists them to get closer to explaining the nature of acquisition itself (explanatory adequacy) (Adamson 1988). Another argument put forth in favor of studying variability in SLA is that if second language development involves even a brief period of systematic use of 'nonlanguage' rules* (as in 'What are you worrying?' (without preposition) (Preston 1996)), it is surely important to understand their nature, source, and contribution (Preston 1996). Eckman (1994) also argues that if variation in L2 could be accounted for without unduly complicating SLA theory, there would be no good reason to exclude this data from a theory of SLA.

The most important argument, in this regard, was put forth by Labov (1966 and elsewhere) for linguistic theory. He argues that the mechanism of language change can be explained only by a theory of language that recognizes *inherent* variability. Language change in linguistic theory correlates in SLA theory with change** in learners' second language speech, where permeability (Adjemian 1976) is assumed to be one of its major characteristics. Dealing with L2 variation is, therefore, an obligatory part of SLA research (Andersen 1989). In this respect, variationist studies of L2 speech, although few, have contributed a great amount to the study of second language acquisition, including:

- It has been demonstrated that the variation found in learners' speech (the majority of the features that have already been studied in individual or group L2) is indeed systematic (Tarone 1983, 1988, Young 1991).

* So-called by Preston (1996).

**Despite similarities between change in L1 and L2, there might be differences between the two, which are not of my concern here.

- It has been found that L2 variation is conditioned by several linguistic and extra-linguistic factors (Regan 1998; Young 1991).
- Classroom practitioners have been made to understand why their language learners fail to generalize newly learned patterns to contexts outside the classroom or vice versa, i.e., fail to generalize forms used in conversations with peers to more formal situations (Tarone 1995).
- Researchers have been helped to sort out the learners' individual grammars and assess their linguistic and sociocultural competences (Beebe 1988).

On several occasions (e.g., contraction of *ARE*, use of resumptive pronouns), the findings of this dissertation indicate that surface presence or absence of linguistic features and their similar/different rates of frequency to native or target language features are by no means an indication of the underlying structure of L2 speech. Nor is the presence/absence of features evidence for their acquisition from target language or their reliance on the informants' native language. Variationist methods and means are of great help in discovering the actual behavior of L2 learners (facts of L2) and its origin, as the results demonstrate empirically.

2.1.3. Types of Variability: Systematic/Non-systematic variability

The speech of second language learners is necessarily more subject to change, development, and variation than the speech of native language speakers (e.g., Adjemian 1976). Less progress has been made in reducing this variation in the speech of the L2 learners to rules, not because the variation is not rule governed but, because variationist studies have not been applied to L2 data for long. The majority of the researchers accounting for L2 variability from a variationist perspective have found that variation in L2 is subject to systematic effects of linguistic or extra-linguistic factors (Bayley 1994, 1996; Dickerson 1975a, Regan 1998; Tarone 1983, 1985; Young 1991, to name a few). Ellis (1985, 1991, 1994a, 1994b, and others; see also Towell et al. 1993 and Young 1996), however, emphasizes that there are two types of non-systematic variability in L1 and L2 languages. The first type is the result of performance lapses: false starts, deviation from rules, etc. The second type is the result of competing rules in the learner's competence that are "acted upon quite haphazardly" (Ellis 1991: 127).

Ellis's argument is based on Gatbonton's (1978) diffusion model, where it is claimed that two consecutive phases are important in the process of acquisition of target language forms: a) an acquisition phase in which the learner acquires the competing forms on the basis of positive evidence and holds them in free variation, using first one in all contexts and then introducing the other and, b) a replacement phase in which the learner progressively uses the forms in their appropriate contexts.

Towell et al. (1993) use the diffusion model as their theoretical framework and in a study on the variable use of 'de/à + participle' structure by advanced learners of French claim to have found evidence to show that advanced learners use forms in free variation over a long period of time. Following Ellis, they assume that if two forms occur in the same situational, linguistic and discourse context, if they perform the same illocutionary meaning, and if there is the same amount of attention paid to their production, these forms are used in a non-systematic way in the language (or interlanguage). Ellis further assumes that non-systematic variability is an "essential requirement of language because it serves as a resource for later language development" (Ellis 1994b: 141). Preston (1996), on the other hand, argues strongly against presence of non-systematic variability in both L1 and L2.

As mentioned before, one goal of the present study is to find out if the variable use of relative/resumptive pronouns and variable contraction of auxiliary verbs in L2 are systematically conditioned by certain linguistic factors or not. The results obtained in this study will throw light on this and indicate if there is any instance of non-systematic variation in my L2 data.

2.1.4. Studies of variation in L2⁵

Studies of L2 variation, a summary of which will be given below, have been carried out within different theoretical paradigms: Sociolinguistic (e.g., Labovian, dynamic), Psycholinguistic (Monitor, cognitive, planning, etc.), and so on. Except for a very few, all studies were able to locate systematic variability in the L2 of their informants. My focus is only on the most important factors (whose effects have been

⁵ I will not refer to all the studies of L2 variation here, and in some cases a very short summary of the studies will be presented. Young (1991) and, specially, Tarone (1988) are both good sources of information in this respect.

studied in the previous literature) and the way they are dealt with here. A selection of studies that have employed the recent versions of variable rule analysis will be reviewed in more detail in the final part of this section.

2.1.4.1. Constraints on variation

Linguistic context

The Labovian approach to the study of variation in language, among other theoretical paradigms, is the one that emphasizes the role of linguistic context. In a study on the pronunciation patterns of Japanese learners of English, Dickerson (1975b) found that the learners' choice among a number of variants depended on the following and preceding phonological environments. Gatbonton (1978) found effects of phonological context on the production of English /ə/, /ɔ̃/, and /h/ by Canadian Francophones. Gatbonton suggests that the learners begin with a categorical point of origin, then they introduce variability (both *free* and systematic) into their speech in a *gradual diffusion* fashion. However, Dickerson claims that variation is systematically related to phonological environment from the beginning and it is never free.

Hyltenstam (1977, 1978, 1983) showed the effects of linguistic environment on the acquisition and use of grammatical forms in the L2 of the learners of Swedish as a second language. Adamson and Kovac (1981) for the first time utilized the VARBRUL program to reanalyze the data on negation collected by Schumann (1978) in his study of Alberto. They were able to give evidence that Alberto's variable use of negation was due to the effects of linguistic environment and the nature of task. Wolfram (1985, and also Wolfram and Hatfield 1984, Veronique 1987 for L2 French) argues for the importance of surface-level linguistic constraints in determining L2 variability in a study on the production of tense markers in the English spoken by Vietnamese speakers. Ellis (1988) analyzed the variable use of two forms: third person -s, and copula -s. His study supports the importance of linguistic factors in variation in L2 morphology and syntax.

In this dissertation I make an attempt to see if the effect of linguistic factors are specific to the L2 speech of the informants, or are affected by their native and target languages.

Effects of topic

Another factor that has been found to affect L2 variability is the topic of the speech (conversation, interview, etc.). Change in topic is one of the contextual ways to reduce the formality of the interview talk in Labovian methods of data collection (Preston 1989). Questions of personal experience, e.g., the danger of death question, and questions related to childhood games, have been effective in acquiring less carefully monitored speech. In order to explore the effects of topic, Selinker and Douglas (1985) collected data from one male informant, interviewed by two different interviewers on two different topics. They speculate that second language speech is developed within different *discourse domains* (topics). Smith (1989), however, did not find any statistically significant differences in the L2 production of his informants with respect to their speech on general and field-specific topics.

The protocols used in the interviews with the informants of this study were all similar. In this way, therefore, I was able to control the effect of the topic by discussing the same topics, mostly related to personal experiences, in the interviews with all the informants.

Effects of interlocutor

Interlocutor effect is one of the important tenets of Speech Accommodation Theory. To gain the social approval of listeners, to maintain their own positive social identities or to achieve efficient communication, L2 learners either adapt to the interlocutors' speech (convergence) or accentuate differences between themselves and interlocutors (divergence) (Thakerar, Giles, and Cheshire 1982). Beebe (1977) and Beebe and Zuengler (1983) used different native and non-native interlocutors to collect data from second language learners and found that L2 speakers shift their L2 depending on the interviewer's ethnicity. In his study of the production of nine morphemes in the speech of adult L2 learners of English, Young (1986) made an attempt to test the effects of interlocutor in a speech accommodation framework, but he achieved mixed results. When the accuracy scores (scores given to each morpheme based on the accuracy of its use) for the nine morphemes under study were totaled, a significant overall effect of interlocutor was observed. But when the scores for each morpheme were considered individually, the

accuracy of use for some morphemes, like plural -s, increased in conversation with native speakers; while the accuracy score for some other morphemes, like copula and articles, moved in the opposite direction, i.e. they scored lower in the conversations with the native speakers. Young 1991 (also Shohamy 1983) did not find a significant role for the interlocutor effect in his study of plural marking in L2, either. I do not examine the influence of interlocutor in my study, but in order to control its effects, Persian and L2 interviews were all conducted by the same interviewer.

Effects of task

For those researchers working within the Labovian framework, the amount of attention that a speaker pays to her/his speech when using an L1 (or L2) is a crucial factor, change of which may cause a shift in the style of the speech. For the proponents of psycholinguistic approaches, too, attention to speech is important: monitored or unmonitored speech has two different varieties of linguistic outputs (e.g. Krashen 1977), as does planned or unplanned discourse. For cognitive approaches to L2 variation controlled or automatic processing causes variation in the speech of L2 learners, as well. How have researchers been able to manipulate the linguistic context to gain data on attended/unattended, monitored/unmonitored, planned/unplanned, and controlled/automatic discourse? The most popular tactic in the literature is to utilize different tasks or conditions to make the informants change their attention from form to meaning or vice versa.

Hulstijn and Hulstijn (1984) used a storytelling task under four conditions of attention to form/content and with/without time pressure and an interview task to collect data and found that focus of attention on content resulted in more accurate production. Bialystok (1982) has however found that a gradual decrease in syntactic accuracy occurs for the tasks that require more focus on meaning and less on structure. Ellis (1987a) utilizes only one task but varies the planning time. His findings indicate that higher level accuracy occurred in planned discourse than in unplanned discourse.

Other researchers using a Labovian paradigm claim that shift in attention causes speakers to shift along a continuum of speech styles, ranging from casual (vernacular) to careful. Dickerson (1975b) (also Dickerson and Dickerson 1977) found that her L2

learners vary considerably across elicitation tasks in the production of /r/ and /z/ (from free speech to word list reading). Schmidt (1977) assigned three tasks to his Arabic learners of English in both English L2 and their native Arabic and found that the pattern of attention to speech holds regardless of the language the speakers used. In Fairbanks (1982) a Japanese learner of English never used the third person singular -s in his casual speech, but his careful style evidenced almost always target-like use of this morpheme. Tarone (1985) examines variation in the speech of Arabic and Japanese adult learners of English on different tasks (grammaticality judgment, oral interview, and oral narration). She measured the use of two bound morphemes-- third person singular -s and plural -s-- and two free morphemes --definite article and direct object pronoun *it*-- across the three tasks. Her results for the two bound morphemes are not that clear, since the tasks requiring more attention to speech elicit more target-like forms in some cases, but have no effects or show lesser accuracy in other cases. With respect to the two free morphemes, the accuracy decreases as the attention to speech increases, contrary to what has been found for phonological studies.

As will be elaborated later, the focus of this study is on L2 speech in its most natural situation. I therefore used sociolinguistic interviews to collect data that include samples of speech from friendly, informal conversations with the interviewees.

In summary, this brief review of the literature on the constraints of variation in L2 demonstrates that in many cases L2 variability is conditioned by linguistic and extra-linguistic factors in a systematic way. Non-systematic variability, however, is claimed to be found in L2 as well (e.g., Gatbonton 1978). Among the disadvantages of these studies is that many of them have explored the influence of only a few factors (in many cases only one factor) on L2 variation and that none has referred to their informants' native or target languages as the possible sources from which the constraints of variability may have originated (except for Schmidt (1987) who studies the effects of L1 phonological features on L2). In the present study I examine the contribution of several (linguistic) factor groups on L2 variability and pay particular attention to the sources of constraints on L2 variation.

2.1.4.2. Current research

The variationist approach to the study and description of language variation and language contact has proven to be one of the most promising research programs in linguistics (e.g., different studies of Labov on L1 (1966, 1969, 1975), Poplack on language contact (1989, 1993), and Young on SLA (1991, 1996), among others). Preston (1996) even claims that the variationist approach is a promising way of deciding among various theoretical approaches to SLA. This is of more importance in SLA since it is a field with an estimated fifteen to twenty theories, models, metaphors and perspectives (Long 1985), and many others brought from other fields (e.g., cognitive, sociolinguistic, ethnographic, linguistic, etc) (Selinker 1984).

On the one hand, researchers using the variationist approach can investigate the effects of different independent factors (linguistic or extra-linguistic), which were proposed by separate paradigms as the major causes of L2 variation, in a single study (multivariate analysis). On the other hand, application of statistical devices like, VARBRUL provides a way to go far beyond the surface linguistic productions of language learners. The probabilistic weightings of influences of different factors on varying forms enable the researcher to access the constraints on variability in an empirical, quantitative fashion.

In what follows, I will give a summary of the most recent studies of variation in L2 that have used a multivariate analysis on their data. Although the application of this approach in SLA research is still in its infancy, the findings of the following studies show a promising future for the empirical studies in the field.

Adamson⁶ & Regan 1991, Adamson et al 1996, Regan 1996

Adamson and Regan (1991) investigate the acquisition of the variable [*ing*] (with IN and I variants) by Vietnamese and Cambodian immigrants in Philadelphia. They apply the VARBRUL 2.0 program (Rand & Sankoff 1990) to analyze their data and their findings indicate that both linguistic (phonological and syntactic) and extra-linguistic

⁶ Adamson and Elliot (1997) is the most recent study of Adamson (and associates) on variation in L2. In this paper they explore the sources (causes) of variation from a 'prototype schemas' (psycholinguistic) perspective. Since they do not deal with quantitative data and the perspective utilized is totally different from ours, I do not review that study here.

(gender) factors affect L2 variation. One advantage of this study is that they use a sample of native English speech to compare to the behavior of their nonnative speakers. In this way, perhaps, they were able to find that ‘frozen forms’ (like *something*) are easier to learn than learning a productive rule. They also claim that they were able to give “a picture of how nonnative speakers acquire the speech norms of the community in which they live that makes sense from a sociolinguistic and a psycholinguistic perspective” (p. 21).

Adamson, Fonseca-Greber, Kataoka, Scardino, and Takano (1996) replicate Wolfram’s (1985) study on tense marking in English L2. They suggest a pedagogical implication for their findings, which maintains that narratives may be a good place for second language learners of English to practice the past tense since they contain a “safe environment: namely the foreground clause, where either the past or present form is acceptable” (p. 132).

Regan (1996) reports on a quantitative, longitudinal study of the progress of learners learning the variable use of the French negative marker ‘ne’ and gives a detailed picture of the acquisition of the native speech community grammar. Data were gathered from advanced Hiberno-English learners of French at two different times: Time 1 (before time in France) and Time 2 (after time in France). The results for each time period is compared to the results obtained from another study on ‘ne’ deletion in native French. The findings of this study are that the overall rate of ‘ne’ deletion increased dramatically between Time 1 and Time 2 and the rule strengthens for nearly all the different factor groups. Interestingly enough, not only did the constraint hierarchies remain the same (except for one factor), but the ordering was generally the same as for the native speakers of French. Regan assumes that the advanced learners of her study “do not change in so major a way their mastery of the structures of the second language but rather their sociolinguistic knowledge” (p. 197). In other words, acquisition had already taken place before the informants’ arrival in France. Exposure to native French affected only the overall rate of ‘ne’ deletion. This clearly shows that overall rate of presence and/or absence of a (variable) linguistic feature cannot always indicate mastery (or lack of it) of that feature in L2. What is of more importance is the pattern of the variable use of that item by L2 learners.

In this dissertation, in addition to data from the target language of the speakers (as in Adamson & Regan 1991 and Regan 1996) and data from less proficient speakers of English L2 (similar to Time 1 in Regan 1996), I will analyze variability in the native language of the informants as well. My three-way comparisons of patterns of variation will not only indicate any change in the speech of L2 learners at two levels of proficiency, but will also hopefully demonstrate how L1 features affect that variation and how nonnative speakers acquire the speech norms of the target language community.

Young 1991, 1996

Young (1991) investigates the plural marking patterns in the L2 English of native Chinese speakers. Overall, it was shown that the English L2 is highly systematic, the pattern of variation in the learners' system changes as acquisition proceeds, and that the relative contribution of all factors that were hypothesized to have some effects on L2 variation was quantitatively determined. The factors whose effects shows considerable change as acquisition proceeds are phonological factors (selected as significant for only low proficiency informants) and the semantic factor groups of animacy and definiteness. Syntactic factors contributed similarly to plural marking in low and high proficiency groups. Unfortunately, except for the phonological factors that behave more like the target language at the higher proficiency levels i.e., the constraint hierarchies look similar) we do not know if changes in semantic factors are towards the target language norms (target patterns are not provided).

The nature of the change in L2 variability at different levels of proficiency is one of the issues explored in the present study. In chapter 4, I will investigate whether change occurs only in rates and/or patterns of variability, and whether the direction of the change is towards the target language patterns or not.

Young's 1996 study explores variation in article usage in the spoken L2 English of native speakers of Czech and Slovak, languages without overt articles. His results are of relevance to my study in some respects. First, concerning the proficiency level of the informants, he finds that accuracy in article usage increases as acquisition proceeds (measured by TOEFL scores); patterns of indefinite article usage, however, do not significantly differ at the two levels of proficiency. However, proficient learners' use of

the definite article is different from that of the lower proficient ones. For the lower proficiency group definite articles are “overgeneralized but there appears to be no clear form-function relation” (P. 170), evidence for ‘non-systematic flooding’ of definite articles. Chapter four of this thesis will show whether this holds for my study as well, i.e., whether we witness any nonsystematic variability at lower proficiency levels.

Second, with respect to the influence of the learners’ native language on their L2, Young finds two places where learners map L1 forms/meaning onto L2 forms in their initial hypotheses about the L2 article system. First, Czech and Slovak use of demonstratives as definite markers (anaphoric *this*) causes the L2 English of these learners to consistently use demonstratives for anaphoric reference and prefer them to definite articles. The second site where transfer from L1 to L2 is most apparent is the influence of L1 (pragmatic) word order on L2 definiteness marking. Czech and Slovak signal thematicity by the use of pragmatic word order (theme (old/given information) stated sentence-initially; rheme (new information) stated sentence-finally). Therefore L2 learners mark final rhemes by the *ZERO* article most because definiteness is marked by L1 word order in these cases; while final themes (not preferred in L1) are marked more with the definite article.

Young (1991) is important and inspiring since it is one of the first studies of L2 variation that attempts to account for variation by taking into consideration several different factors by utilizing the latest versions of the sophisticated tools for the analysis of the L2 data which VARBRUL (Pintzuk 1987) has provided. Young 1996 is also the first study that has utilized variable rule analysis and claims to have found nonsystematic variability and L1 influence in the L2 speech of second language learners.

Young, however, bases his accounts of the learners’ native languages on the prescriptive rules of grammar and does not show what the target language speakers actually do in these respects (i.e., variability in article or plural marker usage in English). I will tackle these issues by using the same variationist method as Young’s, but with further comparison of the results obtained for L2 with those of the same analyses on the informant’s native and target languages. In this way I will be able to empirically demonstrate the effects of L1 on L2 (chapters 4-6) and note any changes as acquisition proceeds (chapter 4).

Bayley 1994, 1996

Bayley (1994) studies the past tense marking patterns of L2 learners. The results of the variable rule analysis show that past tense marking is indeed highly systematic and subject to multiple linguistic, developmental, and social conditioning factors. Bayley (1996) uses variable rule analysis to examine the patterning of linguistic and social constraints on final /t/, /d/ deletion and affixation. He pays special attention to the effects of /t/, /d/ deletion on past tense marking of regular verbs. The results of the analysis show both convergence with and divergence from patterns found in studies of the same phenomenon in native dialects of English.

Summarizing, the tour of current variationist studies in SLA shows that variability in second languages used by L2 learners in their oral and written productions is mostly conditioned by linguistic and extra-linguistic factors. Presence of possible nonsystematic variation still continues to be claimed in these studies (Young 1996). Young (1991, 1996) and Regan (1996) also found that variability in L2 speech changes as learners progress from early stages of L2 acquisition toward the mastery of the target language features, but this change is mostly in overall rate of rule application, not in the patterns of conditioning factors. Influence of L1 features on L2 variability was also explored by Young (1996), but he did not analyze L1 in a quantitative, variationist manner, nor did he take into consideration variability in the target language of his informants.

Although these findings are all valuable to SLA researchers in the sense that they give a reliable description of the facts of second language speech and development, several issues are still in need of further investigation and reliable responses from variationist researchers in SLA. We are still facing claims that certain features of L2 speech may haphazardly be used by L2 learners (e.g. Young 1996). Therefore, we still need to know how far variation in L2 is systematic and if there is any evidence for unsystematic variation in it. Moreover, except for Young 1996, which uses only the prescriptive rules of L1 as a guideline, we are not provided with any information about the effect of learners' native language (variable features, in particular) on variation in L2. Another important issue is the acquisition of the variable features of the target language and their influence on L2 variation. If a linguistic feature is variably utilized by the native speakers of a certain language, e.g., relative marking in English, how do second language

learners acquire this feature? What effect does variation in the target language exert on variability in the L2?

In this dissertation I make an attempt to tackle these issues, in particular acquisition of variability and its effects on L2. I start my inquiry from L2 speech itself by first examining the nature of variation in it and the factors that condition that variation. This will help us to see if there is any unsystematic variation in the L2 of the informants. Then I will look for the sources of the constraints that condition L2 variation.

In addressing the issue of the sources of variation, I will look at the speech of the informants in terms of three possibilities: 1) variation is transferred, 2) variation is created, and 3) variation is acquired. The first two of these assumptions are in complete correlation with two theoretical orientations (hypotheses) in SLA, namely the *interference* hypothesis (transfer) and the *interlanguage* hypothesis (creation). The third possible source of variation (variation is acquired) focuses on the acquisition of variation from the target language.

The following section provides a short review of those two theoretical orientations in SLA and their role in the development of the field.

2.2. Hypotheses of L2 acquisition

2.2.1. The transfer hypothesis

The adult second language learner comes to the task of language acquisition with the knowledge of at least one other language that provides a major source to possibly affect the construction of L2 grammar. On the one hand, prior linguistic knowledge can inform the learner of a great deal about language in general, which provides a rich inventory of hypotheses for learners to draw on in the construction of an L2 grammar (Yip 1995). On the other hand, it causes incorrect hypotheses that result in non-target-like L2 grammar, which is usually called interference or negative transfer.

‘Transfer’ is a highly ambiguous term. In spite of more than four decades of research following Weinreich (1953) and Lado (1957), there is still little consensus on what transfer in language actually is (Dechert & Raupach 1989; Odlin 1989, 1994; among others). Sharwood-Smith and Kellerman (1986) have suggested replacing ‘transfer’ with a superordinate term, *crosslinguistic influence*, a term which is claimed to

be both comprehensive and theory-neutral. However, the term 'transfer' has persisted and its definition has been broadened to include most of the crosslinguistic phenomena. The definition of transfer that will be used in this study is adapted from Odlin (1989) who offers the following as a working definition of crosslinguistic influence:

Transfer is the influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired (1989: 27)

The interference hypothesis (negative transfer) maintains that when a learner attempts to acquire a second language, s/he will make errors in the L2 which are predictable from her/his native language. That is, the native language and its structure will interfere with learning the L2. In cases where L1 influence assists in learning the target language features, positive transfer is said to have occurred.

Historically speaking, the era of Contrastive Analysis (especially the 1960's) was the golden age of transfer studies. Contrastive Analysis, as practiced in the 1950's and 1960's, assumes that language learning is habit formation; thus, where L2 and L1 differ, the old habit hinders the formation of the new habit, and when they are identical, it assists the acquisition of the similar habit in L2 (see Lado 1957). This framework led the majority of the studies to the conclusion that the surface similarities or differences of the involved languages were the main and the only constraint on language transfer; and that the difficulties of second language acquisition could only be determined through contrastive analyses (Odlin 1989).

On the one hand, the comparison of surface features proved problematic and many of the predictions (hierarchy of difficulties, for instance) were found to be false, simply because L2 learners were found to behave differently from what had been predicted by contrastive analysts. On the other hand, the theoretical foundations of the contrastive hypothesis (structuralism in linguistics and behaviorism in psychology) were unable to account for much of the language learning processes (Huebner 1991). The collapse of behaviorist psychology and the emergence of the generative grammar followed by Creative Construction theory in SLA (Dulay & Burt 1972) led to the fall of

the transfer hypothesis in language acquisition studies. The 1970s were, then, the age of the ideas of ‘virtual nonexistence of L1 transfer’ (as mentioned by Sajavaara & Lehtonen 1989; Odlin 1989).

What at first appeared to have been a fatal blow to transfer analysis in general eventually led the way toward a fruitful reinterpretation of the influence of learners’ L1 in their second language speech (Kohn 1986). The conceptualization of transfer within the new trends in SLA and generative linguistics has initiated a great number of investigations and discussions on different types and hypotheses of transfer since the late 1970s (Kellerman 1977, 1978, 1979; Jorden 1980; Meisel 1982; Wode 1981, 1982, Zobl 1980, 1982, 1988, 1992, and several collections of papers on the topic (e.g., Gass & Selinker 1983)). The present state of the art of language transfer research is characterized by an enormous number and diversity of studies dealing with various aspects of crosslinguistic influence. It is now acknowledged that accounting for the learners’ prior linguistic knowledge is an integral part of any theory of L2 acquisition.

2.2.2. The Interlanguage hypothesis

The era after contrastive analysis witnessed a strong reaction against the transfer hypothesis in language learning/teaching. The learner was now seen as an autonomous creator of the language system. The Interlanguage hypothesis, initiated by the work of three prominent figures of SLA in 1970’s, was referred to this autonomous nature of second language grammar.

Corder (1967) argued that second language research should follow the example of first language research and view the learners’ development as a development of underlying linguistic competence (Chomsky 1965). He further suggested that L2 learning is an operation whereby the learners’ mental learning device builds a system using the language to which s/he is exposed and creates a particular transitional system. This transitional system is then modified by the new input, and the process goes on. Nemser (1971) proposed that second language development should be seen as a succession of evolving systems that took the learner nearer and nearer the target language, and farther from his L1 (approximative systems). Selinker (1972) proposed an approach similar to that of Nemser. His term “interlanguage” was eventually adopted by most people when talking about approximative or transitional systems.

The interlanguage hypothesis suggests that when one tries to learn L2 after having already acquired 'meanings' in a native language, the utterances which s/he will produce will not be identical with those produced by native speakers of that language, nor will they be an exact 'translations' from his/her native language. Rather, a new, separate language system will develop, a system of interlingual forms (Cancino et al 1974). Nemser's (1971) version of this hypothesis proposes that 'learner language' will not only be systematic, but will also evolve in successive acquisitional stages from initial through advanced learning.

Interlanguage is now a well-established notion in second language studies. Even variationists who have performed quantitative analyses on data from second language learners have tried to provide evidence for the independent nature of L2 speech, its systematicity, and its difference from the (descriptive) grammar of native and target languages⁷.

In the present study, I intend to test the transfer and interlanguage hypotheses in an empirical manner. Unlike previous studies of language transfer (studies in a contrastive analysis framework), I deal with variable features of L1 and L2 (not so-called categorical rules of grammar), and, moreover, incorporate data from the target language of the informants. The comparison of the results of the variationist analyses on L1, L2, and the target language data will demonstrate in what contexts, if any, L2 learners rely on their native language (transfer), in what contexts they follow patterns of variation in the target language (acquisition), and where in their L2 productions they create their own independent system of variability (interlanguage).

⁷ Another, a rather stronger, reaction against contrastive analysis in 1970s was the Creative Construction trend that is best exemplified by the works of Dulay and Burt (1972) and Baily, Madden & Krashen (1974 for adult L2 learners). The error analyses of the 1960s and early 1970s showed that some types of error are common in the emerging second languages of speakers of various L1 backgrounds. Brown's (1973) study, also, indicated the occurrence of the same errors (known as developmental errors) in the speech of children learning English as their native language. Dulay & Burt claimed that the effects in L2 of the second language learners previously accounted as L1 effects can be attributed to developmental stages of language acquisition. Language transfer, then, was claimed not to be a significant factor in L2 acquisition. The thrust of their argument was to support the idea that L2 acquisition is not different from L1 acquisition (L2 = L1), and tried to reject or downgrade the effects of native language on L2 acquisition. I will not deal with the L1 = L2 hypothesis in this study.

2.3. SLA: data and the importance of empirical study

One of the main challenges faced by research involving the interaction of two or more linguistic systems is the problem of developing a comprehensive description and comparison of the languages in the contact situation (Ellis 1994a; Odlin 1989). The dominant comparison method during the heyday of Contrastive Analysis (and after) was limited to a surface comparison/contrast of the linguistic properties and frequency counts of the tokens in natural or elicited data. Contrastive Analysis, however, proved inadequate empirically in accounting for second language acquisition processes and prediction of (and accounting for) errors made by these learners (Newmeyer & Weinberger 1988), one possible reason of which might have been the method used in describing (e.g., reliance on surface data) and comparing languages. This led Odlin (1989) to assert that for a thorough understanding of language transfer, more than just structural comparisons are necessary. An important problem with respect to the method of comparison is, thus, the nature of the appropriate data to base theoretical/empirical studies on, i.e., what to compare?

Linguists have in general tended to rely on three sources for data: 1) *grammaticality judgment data* (speaker intuition), 2) *data from introspection* (the linguist's intuition), and 3) *empirical data* (i.e., data produced by the speakers in natural situations). Overall, most of the controversy with respect to the appropriate data for SLA has been focused on the use of either empirical or judgment data.

The use of judgment data has both a practical and a theoretical rationale in linguistics. Grammaticality judgment tests are generally easier to administer, and it is easier to extract data from them. They can be administered to a large number of informants in a single session. Another practical reason advanced in support of grammaticality judgment tests is that some linguistic features are not accessible to investigation in production data, for the simple reason that the speakers either do not use them or their occurrence is not that frequent, while use of grammaticality judgment tasks helps researchers elicit whatever kind of data they need (Gass 1994).

The strongest claim made on behalf of the grammaticality judgment tasks in (theoretical) linguistics is that 'judgments of grammaticality provide better data for the study of a person's knowledge of his language than do actual utterances of the speaker

(Baker 1978:12). The motivation for this claim stems from Chomsky's early works (e.g., 1957) where it is argued that the proper object of study for linguistics is an idealization of speech behavior which is best accessed by means of introspection. The data obtained in this way is claimed to be removed from any non-linguistic context which might affect its interpretation (Lightbown 1984). Moreover, it is believed that when judging the grammaticality of the structures, the speakers need not access the processing systems responsible for using the underlying grammar in actual performance (Ellis 1991). They, then, provide a relatively direct window into linguistic competence of the speakers (although many believe that they are still performance data).

With respect to SLA research and use of grammaticality judgment tasks in this field, it has been argued that if we assume that learner languages are similar to natural languages, then it might also be reasonable to presume that they can be investigated by means of the same methods (Gass 1983; see also Schachter, Tyson, & Diffley 1976). However, grammaticality judgment data are fraught with difficulties, among which are:

- Grammaticality tests do not elicit linguistic behaviour per se but rather a response indicating the learner's belief about the L2 grammar⁸ (Munnich, Flynn, & Martohardjono 1994).
- Learners of second languages appear to deploy a variety of strategies when asked to render grammatical judgements (Goss, Zhang & Lantolf 1994). Grammaticality judgement tasks⁹ are, then, not devoid of extralinguistic factors (Birdsong 1989, 1991; Cowan & Hatasa 1994; Ellis 1987b, 1991; Flynn & Manuel 1991; Gass 1994; Jordens 1991; Schachter 1991). Parsing strategies (e.g., rejection of a structure because of difficulty of parsing), context and mode of presentation, pragmatic factors,

⁸ The researchers compare grammaticality judgments with elicited imitation tasks, here.

⁹ There is a distinction between tests and tasks. Tests are devised to measure what the learner knows and does not know of the target language. A subject's performance is measured against that of target language speakers. A task, on the other hand, is devised to reveal what a learner knows: the rules s/he uses and the system s/he has. We may sometimes infer something about the learner systems from the test results but it is not what the tests are devised to reveal. In this section I have used tests and tasks interchangeably because many researchers do not distinguish between them in their studies and, also, the arguments posited here are extendable for both of them.

speakers' mental/physical state, linguistic training, translation strategies, etc. are all factors that may affect the intuition of the subjects (Sorace 1996).

- If the effect of the extralinguistic factors on judgement data is approved, it is thus concluded that they do not measure what they are supposed to measure, i.e. linguistic competence of L2 learners. A sentence may be judged grammatical or ungrammatical for reasons that have nothing to do with its status in the internal grammar of the speaker. Grammaticality judgement tests, therefore, are not valid enough to tap the linguistic knowledge of the speakers (Cowan and Hatasa 1994; Sorace 1996; Swan 1987).
- Grammaticality judgement data are not reliable: it is possible for the judgement of two speakers to vary with respect to a particular utterance, and it is also possible for a single speaker to have different judgements in different replications of a judgmental test (Chaudron 1983; Cowan & Hatasa 1994; Kellerman 1985; Soarce 1996). Choice of vocabulary, selection of syntactic structures, L1 of the informants, experimental design, sentence length, psychological status of the informants, etc. have been claimed to contribute to the variability in judgement tasks (see also Christie and Lantolf 1991, Ellis 1990, and Gass 1994 for more on reliability of judgement data).
- It is not always possible to tease apart issues of grammaticality judgement and comprehension (Lantolf 1993 cited in Gass 1994). We cannot clearly determine if informants are making judgements of the grammaticality of the test sentences or their comprehensibility.
- Since a variety of performance strategies are involved in judgement tasks, it might be argued that these tasks are just as much performance tasks as any other kind of task¹⁰ (Birdsong 1991; Ellis 1991). Then, why should research bother to use a very controversial task, the result of which may not be a reliable measure of what was supposed to be tested?

¹⁰ Many theoretical linguists working in UG framework themselves believe that grammaticality judgment data are not a direct reflection of the competence of the speakers. Lasnik (1981, cited in Birdsong 1991, also Gass (1994) for SLA) observes that responding to a grammaticality query is an instance of linguistic performance. Some others, however, think that judgment data are performance data that are linked to linguistic competence of the speakers and can tap their underlying competence (Cook 1988, 1996).

- A testing context has its own specific situational and psychological properties. It is a mostly *formal* situation in which one or more informants are under a certain psychological and situational (e.g., time, presence of a tester, formal setting, etc.) pressure of taking a test. The linguistic output of such a situation can by no means be an actual reflection of what language is naturally used for (see Larsen-Freeman & Long 1991). It resembles only one variant of the variable grammar (competence) of the informants (Tarone 1988). A reliable product of the linguistic knowledge of the informants involves a natural situation to elicit vernacular speech that is the speech used to communicate meaning in its most natural setting.

Given the problems of judgment data, only an empirical study that examines the L2 as actually used by the second language learners can offer insight into the L2 grammar. Selinker (1972), more than two decades ago, suggested that such data would be those “observable data to which we can relate theoretical predictions” (pp. 213-214). The only observable data, i.e., the utterances which are produced when a learner attempts to make meaningful performance, are: 1) utterances in the learner’s native language produced by the learner; 2) L2 utterances produced by the learner in meaningful performance situations; and 3) L2 utterances produced by native speakers of that L2. These three sets of data, for Selinker, are the psychologically-relevant data of second language learning¹¹. Tarone (1988) also asserts that if any L2 style is to be taken as basic, it is the vernacular style, because this style is the most unique to the L2, and contains the smallest number of borrowings from other language systems.

Naturalistic data, as the only directly observable behavior of the learners, do not have many of the deficiencies enumerated above for metalinguistic judgments. They are the actual linguistic behavior of the speakers (not their attitude); they are the outcome of whatever linguistic knowledge that is available at the time of speech (not the artifact of testing situation or extra-linguistic factors); and although sparseness of some linguistic features is in itself meaningful for an analyst, there are strategies to collect data about those constructions in a natural manner (such as using different interview protocols).

¹¹ But see Selinker (1984) for his argument on the usefulness of judgment data in SLA research.

Given the problems with the judgment data, given the principle of ‘the use of appropriate data’ incorporated in variation theory (Chapter 3), a framework that this study is based on, and given the fact that the use of quantitative techniques “has successfully overcome many of the difficulties associated with analyses which evaluate anecdotal data by intuitive judgment” (Tagliamonte 1991: 20), only an empirical study which examines language as actually used by the speakers can offer insight into the issues tackled in this study. The data that the present study is based on include (tape) recorded speech samples from learners of English as a second language, native Persian, and native English speakers.

Chapter three: METHODOLOGY AND DATA

3.1. Theoretical framework and Method of analysis

It is evident that variation is pervasive in human languages. In Persian, for instance, depending on the linguistic, social or historical contexts, one of the following variants may be used to express a single function of asking somebody to 'sit down'.

- | | |
|--|---|
| 1. a1. <i>beshin</i> | a2. <i>benshin</i> (both used in informal situations) |
| b1. <i>beshinid</i> | b2. <i>benshinid</i> (both used in less formal situations) |
| c1. <i>befarma:</i> | c2. <i>befarma: 'id</i> (both used in formal situations) |
| d1. <i>befarma: beshin</i> | d2. <i>befarma: 'id benshinind</i> (both used in formal situations) |
| e. <i>betamarg</i> (very informal, rude) | |
| f. <i>jolus konid</i> (archaic) | |

The branch of empirical linguistics that tries to account for variation and change, exemplified in 1, and scientifically investigate language use and structure as manifested in natural(istic) contexts is known as variation theory (e.g. Fasold 1984; Labov 1971, 1984, Poplack 1993; Sankoff 1982, 1988a).

Since one of the main concerns of this dissertation is to account for variability in second language speech, on the one hand, and compare patterns of variability in three contexts of Persian, L2, and native English, on the other, the theoretical approach adopted here is the quantitative paradigm incorporated in variation theory and the comparative method applied to language contact studies in the recent developments of this theory (Poplack & Meechan 1995; Meechan & Poplack 1996, Walker 1999, for instance).

Variation theory focuses on the examining of variables defined as *alternate ways of saying the same thing* (Labov 1971), as in 1. Researchers working in this field have come to recognize that in a corpus of natural speech, "it is not only the contrast between what is said versus what does not seem to be said which must be accounted for, but also that certain elements occur more frequently than others and that these frequencies, not just presence and absence, are systematically conditioned by co-occurring quantities"(Sankoff 1982). When scientifically accounting for the variability, therefore,

“striking and widespread regularities may emerge which pertain to the relative frequency of occurrence or co-occurrence of various structures, rather than to their existence or grammaticality” (Sankoff 1988a: 141). For variationists, then, elements undergoing variation are systematically conditioned by co-occurring linguistic/extra-linguistic factors that might include: the phonological environment, the syntactic context, the discursive function of the utterance, topic, style, situation, and personal and/or sociodemographic characteristics of the speaker or other participants (Tagliamonte 1991). Variationists seek to discover the variability, its frequency, and its conditioning factors. In other words, variation theory is centered on discovering the *patterns* of usage in languages, i.e., *the relative frequency of occurrence or co-occurrence of various structures, rather than their existence or grammaticality* (Poplack 1993; Sankoff 1988a).

The most systematic patterns are found in the *vernacular*, “the relatively homogeneous, spontaneous speech reserved for intimate or casual situations” (Poplack 1993: 252). Thus, the notion of appropriate data has gained importance in variation studies. ‘Appropriate data’ implies that data should be collected in environments where the language is used as naturally as possible, they should be representative of the speech patterns of the community, and there should be enough tokens of the variants being investigated to allow for patterns to emerge. Other principles of the variationist framework are the use of appropriate informants, the principle of accountable reporting, and the circumscription of the variable context (Labov 1966; Poplack 1993).

The use of appropriate informants implies that in order to find the speech patterns representative of a community, speech samples should be collected from individuals with shared community norms. The principle of accountable reporting requires that not only each instance of a particular variant, but also the number of instances in which it could have occurred, but did not, should be incorporated into the analysis. The sum total of occurrences and non-occurrences of variant realizations in a given context together constitute *the linguistic variable* (Poplack 1993: 279). And finally, circumscribing the variable context suggests that before anything else, a variationist must locate the environments in which choice between the alternate realizations of a form is an option. In other words, the object of the study should be clearly defined (Poplack 1993). In a variationist study, therefore, every single instance of each surface form that is within the

circumscribed area of variation is included, along with a detailed and systematic notation of its environmental characteristics (Tagliamonte 1991). The statistical procedures (and tools), known as *variable rule analysis*, are designed by variationists to enable the analysts to extract the patterns from this information.

Research on variability and change in SLA indicates that the same processes may be at work when second language learners alternately use different linguistic forms to express a single function (Fasold 1984). This means that in second language speech, too, certain elements occur more frequently and that these frequencies, not just their presence or absence, are systematically conditioned by co-occurring quantities. Application of the variationist means (e.g., computer programs) and methods, which have been able to discover patterns of variability in many first languages, to the SLA data can solve one of the abiding problems for second language acquisition theory, namely, how to describe and explain the highly variable speech of learners of second languages (Bayley 1996; Young 1991). Some of the previous studies that have already made use of the variationist techniques have contributed a great amount to the study of second language acquisition. Application of the most recent methods of multivariate analysis of linguistic variation developed by sociolinguists will enable us to examine the effects of conditioning factors on variation in the areas of SLA that have not been empirically investigated before, and determine the relative weight of constraints and their hierarchy.

The statistical method I employ in this study is Goldvarb 2.0, a variable rule application for the Macintosh (Rand and Sankoff 1990). This program uses a multiple regression procedure which is able to determine the relative contribution of different environmental factors to the occurrence of the variants under investigation (multivariate analysis). In the following section I present a sketch¹² of the procedures used by this program to analyze variable data.

¹² For a detailed analysis of variable rules see Sankoff 1988b. Rand and Sankoff (1990) is the most detailed explanation of the application of GoldVarb, a variable rule program for the Macintosh. Young & Bayley (1996) detail the procedure of applying variable rule programs to SLA data. The sketch of the variable rule analysis presented in this section has been taken from Sankoff 1988b.

3.1.1. Variable rule analysis and VARBRUL program (Sankoff 1988b)

A variable rule analysis is a statistical treatment of variability and its conditioning factors in language. Prerequisites for a variable rule analysis are:

- Existence of some sort of choice between two or more alternatives (variability).
- Unpredictability of the outcome of the choice process.
- Recurrence of the choice process.

The essence of the variable rule analysis is to examine how the choice process is influenced by different contextual (linguistic, extra-linguistic) factors. The set of related factors that condition a choice process constitute a *factor group*. For instance, different phonological segments preceding the English auxiliary/copula *IS* comprise a factor group that might include vowels, stops, glides, etc. The set of possible choices (e.g., contraction/non-contraction of *IS*) constitutes a *variable*. The variable context is the environment in which the variation (zero, contracted, or non-contracted) can occur.

Application of variable rule analysis to a set of variable data gives the *overall percentage of occurrences (rates)* of the variable under study, both in general and in each individual context (factor). Comparison of the results at this stage, i.e., comparison of the overall distribution of data in each factor, is called *comparison of marginals*. Marginal results, although helpful, are not always clearly interpretable. They indicate only the individual effect of each factor in the process of variable selection of each option. This shows only the *rates* (overall percentage of occurrences) of the variable use of the linguistic features, not their *patterns* of conditioning. For instance, if in the study of variable contraction of *IS* both preceding phonological and grammatical environments have been incorporated into the analysis, marginal results may indicate that preceding consonants and pronouns strongly favor contraction. But, if the majority of the pronominal subjects are tokens of *IT* and *THAT*, which are both consonant-final, how can one determine if *IS*-contraction is favored by preceding consonant or preceding pronoun? It is evident that the two factor groups of preceding phonological and grammatical environments have a particular relationship in the data. When this type of dependence among factor groups obtains, the examination of marginals may give highly misleading results. Here it is said that the factors have non-independent effects or that they *interact*.

A primary goal of the analysis is “to allocate contextual effects on the choice process among the different cross-cutting factors which make up the contexts, in a way which is able to correct both for dependent factors and for interacting factors” (Sankoff 1988b: 987). Here is where the actual process of variable rule analysis begins. In order to find the configuration of factor groups most appropriate to a given data set, VARBRUL programs are designed to carry out all statistical procedures automatically in a stepwise fashion. The first step is to find a single significant factor group by testing each factor group individually. If no group is significant, it is concluded that none of the factors considered has any influence on the choice process (the null hypothesis is confirmed (variation in this context may be random)). If there is found to be a significant factor group, the program retains it and tries to add a second factor group. The process continues until no further additions of factor groups contributes statistical significance to the model, and the program gives the optimal groups of factors that exert the most significant effects on the choice of the dependent variable at this stage. This phase of addition of factor groups is called the *stepping-up* solution. The *stepping-down* solution is based on the same principle, but in reverse. This time the program starts by calculating the likelihood of the model with all factor groups incorporated (*first stepping down run*¹³). It then discards the group whose loss least significantly reduces the likelihood, and the process continues until it exhausts all factor groups. The program, finally, finds the most significant factor group/s at this phase. In this way the program calculates the probability that each proposed factor contributes to the occurrence of the application value and displays its finding by attaching a decimal number (*factor weight*), or coefficient (*p*), to each factor. A *p* value greater than .50 indicates that the factor favors, while a *p* value less than .50 indicates that the factor disfavors choice of the dependent variable. The hierarchical order of the factors in a selected factor group is called the *constraint hierarchy*. If, in some cases, the results of the two stepping-up and stepping-down solutions do not coincide, the status of selected factor groups remains uncertain. Interaction among two or more factor groups might be the cause of this discrepancy.

¹³ Probabilities for the non-selected factor groups in our comparison of results will be reported from this run. Although non-selection of a factor group indicates the lack of its influence on variation, the hierarchy of its factors and their probabilities can be revealing when compared to those of the same factor group that is selected in other contexts (languages).

Although certain methods of handling interaction have been prescribed in the literature, only those that have been used in this study will be elaborated on in the relevant sections.

3.1.2. Comparative method

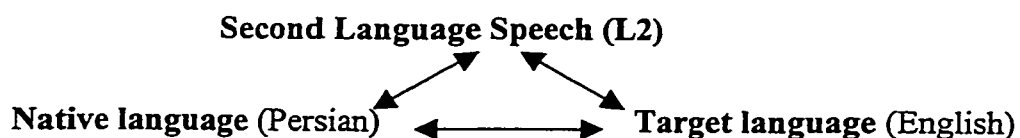
It was previously mentioned (chapter 1, section 1.2) that one of the major issues in the studies of crosslinguistic influence in SLA was the lack of an appropriate method of comparison of languages in the contact situation. Researchers working in the traditional contrastive analysis framework compared and contrasted languages, but they were concerned with only *categorical* structures of the languages and accounted for only the presence or absence of linguistic structures and their frequencies (cf. Sajavaara & Lehtonen 1981; Dechert, Bruggemeir, and Futterer 1984). This tradition of comparing languages lost ground because of its inability to account for ‘errors’ made by second language learners. Further, there were predicted errors (by contrastive studies) and hierarchies of difficulties which did not occur. One reason for this failure was, supposedly, reliance of contrastive analysis on surface grammatical rules of the languages and the frequency of their usage. As a result of variationist studies of many languages, it is now evident that neither are all linguistic structures categorical (if any), nor do the frequency counts of their presence or absence in a given language indicate their underlying patterns of usage. For instance, according to the overall distribution of data in chapter four with respect to the use of contracted *ARE*, a contrastive analyst’s interpretation would indicate that the proficient and less-proficient informants behave identically (the frequencies are similar), while the variable rule results, which rely on patterns of variability in addition to surface presence or frequency of usage, indicate the opposite.

What is recently known as ‘the comparative method’ in the new trends of variation theory is also engaged in comparison of languages. But in this new twist on the method, following the tenets of variation theory, it is the patterns of *variability* that are compared and contrasted to each other. Remember that in variation theory it is not only the presence or absence of linguistic items, nor even their frequency that must be dealt with, but also systematic conditioning of “the frequencies by co-occurring quantities” (Sankoff 1982) that need to be accounted for. The comparative method, then, deals with the inherent variable features of the languages and goes beyond surface frequency counts

and takes into consideration variability and its conditioning factors with respect to the presence/absence of linguistic features, their overall frequency, their distribution in individual linguistic (and extra-linguistic) factors, and, finally, the probabilistic influence of each factor on variable use of those features.

In the comparative method a quantitative, variationist analysis of the linguistic structures under investigation is first provided for each language. The different rates of occurrence of these structures and their conditioning systems in one language form a quantitative pattern, which is compared with that of their counterparts in other languages. In the present study my assumption is that alternate use of relative/resumptive markers and the variable contraction (deletion) of certain forms of auxiliary verbs in the native English, native Persian and English L2 of the informants are systematically conditioned by linguistic and/or extra-linguistic constraints. I use variationist methods and procedures to determine the patterns of variation in each language. Then, a three-way comparison is carried out as shown in figure 3.1, to determine the status of patterns observed in second language speech, i.e., to see what patterns L2 learners do follow: Persian patterns of variability, English patterns of variation, or their own system of variation.

Figure 3.1. *A three-way comparison of data:*



This method has recently been developed in identifying language contact phenomena, e.g., code-switching and borrowing (Meechan & Poplack 1996; Poplack & Meechan 1995, 1998, for instance). These studies have been able to demonstrate the language membership of other language items in recipient languages through a quantitative comparison of patterns of occurrence of linguistic items in donor and host languages. They were, therefore, able to present an accurate definition of code-switching and borrowing, two of the most complicated phenomena of language contact situations. Although the terms transfer and borrowing are generally used to refer to different domains of language functioning (transfer to language learning situations and borrowing

to language change, etc), there are basic similarities between the two (Corder 1992; Odlin 1989). They both involve the interaction of two linguistic systems. Wode (1984) even believes that in terms of underlying processes they appear to be the same. Similarities between borrowing and transfer have also been asserted by Odlin (1989), although he argues that their results are often different¹⁴.

Another area in which application of the comparative method has produced promising results is the debate on the origin of African American Vernacular English (AAVE) (e.g., Poplack 1999; Tagliamonte 1991, Walker 1999). Researchers have been trying for years to find what AAVE was like in earlier eras, i.e., if it derives from a prior creole or has its roots in English. Studies that have utilized the comparative method have been able to compare and contrast several variable features of AAVE (e.g., full/contracted/zero copula) and their conditionings to those of native English, Creole languages, and earlier written/oral samples of AAVE, and have been able to throw light on many controversial issues in this respect. Application of this method for my purposes in this respect is well-justified: I am concerned with both variability and comparison of patterns of variation in the languages involved in a contact (learning) situation, and my goal of finding the sources of variation in second language speech is not very much different from the one pursued in the origin studies in AAVE. In those studies researchers are trying to explain variability in AAVE by processes derived from either Creoles or English; in this study I attempt to account for L2 variability according to native or target language patterns. It is therefore appropriate to extend the variationist methodology to this area of SLA (see also Winford 1998: 114-5 for a discussion on the need for application of variationist studies on AAVE to L2).

In order to arrive at an accurate assessment of the linguistic patterns that operate in L2 data and to find out the effects of native and target language variability on L2 variation, the comparative method necessitates having samples of the speech of native and target language speakers to establish valid native/target baselines¹⁵ (not idealized

¹⁴ For Odlin, “borrowing transfer” is the influence that a second language has on L1, and “substratum transfer” is the influence of a source language (L1) on the acquisition of a second language.

¹⁵ Having target/native language data to have an analysis based on ‘more solid and less speculative grounds’ (Tarone 1988: 124) has also been emphasized by Selinker 1972, Tarone 1988, and Wolfson 1989 but not in the comparative framework.

target forms). The application of this method, which takes into consideration *patterns* of variability, to compare systems of variation in the three contexts of L1, L2 and target can demonstrate if L2 variation originates from native or target languages (when it is found that L2 patterns are similar to either of them). Moreover, introduction of this method (by applying it to L2 data) can assist SLA scholars in their search for an alternative (perhaps better) method of comparing languages involved in a transfer situation.

The focus in my comparisons will be on the following criteria, of which the last is the most revealing, since this is the place where patterns of variation are evidenced:

- Presence or absence of a variant.
- Frequency patterns of the use of each variant.
- Constraints on the variable use of each variant and their hierarchies.

3.2. The Informants

3.2.1. Selection of informants

Of the informants of this study, a total of 27 individuals (Table 3.1) who provided data for my Persian and L2 corpora were all native speakers of Persian. Attempts were made to select informants who spoke standard Persian¹⁶ (spoken in Teheran, the Iranian capital). In order to have a representative group of informants, several techniques were used in their selection. As an in-group member, I could easily contact the members of the Persian Cultural Group (a socio-cultural group with mostly educated members) and ask them for interviews. However, to give all the informants an equal chance of selection, I provided a list of educated (about 100 names, mostly graduate student members of the group) Iranians¹⁷. Then, after excluding inappropriate informants (different L1s, etc.), the remaining members were contacted. A total of 14 individuals from this group agreed to participate in the study. Another major source of contacts was either my own personal

¹⁶ Although some of the informants were from different parts of Iran, they all spoke standard Persian with their friends and family members.

¹⁷ As an elected member of the executive committee of the group for three terms, the researcher had access to the members' (and some non-members) names and addresses.

friends or 'friend of a friend'¹⁸. Ten informants who were mostly less proficient were found using this technique. Finally, several second language teaching centres (University of Ottawa, Carleton, Welcome centre, etc.), were contacted, providing three (less proficient) newcomers to Ottawa.

In order to control for the possible effect of the speakers' first language on their L2, individuals with only Persian as their first language were contacted. To obtain as much 'vernacular' speech as possible for the Persian corpus, attention was given to including individuals who were both fluent in their L1 and showed the least amount of influence from English on their native language¹⁹ (not more than only a few instances of multi-word code-switches to English, no English structures used in their Persian, etc.).

As for the L2 data, I included individuals who had all learned English as a foreign/second²⁰ language. Informants with native-like fluency in English (those who were raised in English speaking communities as bilinguals) were excluded because they are not considered as second language learners. Since different dialects of English might have different patterns of variation and different effects on L2 speech, I tried to include informants who were exposed to Canadian English (preferably in Ottawa).

L2 background of the Informants

All these informants (Persian, L2) had almost identical backgrounds in their L2 education at their high school and undergraduate levels. They all started learning English at grade seven, four hours a week until graduation from high school. They have passed two courses of general English and one or two courses of English for Specific Purposes (ESP) at the undergraduate level in Iranian universities. Textbooks (except for ESP), methods of teaching, and socio-educational context have all been almost identical for these learners throughout their education. Only one informant attended private English

¹⁸ In this technique, used in variationist studies, you ask an interviewee, perhaps a friend, to introduce (and/or to arrange an introduction of) his/her own friend(s) for an interview.

¹⁹ Otherwise I could end up with English structures in Persian data. None of my informants showed such an effect from English.

²⁰ There is a difference between learning an L2 as a foreign or second language in some SLA literature. I will however use these terms interchangeably in this study.

classes for a short period of time. The advanced speakers of English had further language education in the form of TOEFL or ESL classes mostly in Canada.

Table 3.1. Informant information

Pseudonym	Age	Sex	Ed.	SEC/YR
<i>English*</i>				
1. John Frank	III	M	S	WC/90
2. Dalton Knowles	IV	M	PS	UC/90
3. Andy Morrison	III	M	PS	MC/90
4. Jeff Finter	40	M	S	WC/94
5. Dwayne Richmond	24	M	PS	MC/94
6. Jerry Greene	24	M	PS	WC/94
7. Daniel Latherby	25	M	PS	LC/94
8. Johanna Wooster	33	F	PS	UC/94
9. Chuck Jones	20	M	PS	MC/94
10. Pamela Broado	26	F	PS	MC/94
<i>Persian</i>				
1. Ali Akbari**	42	M	S	MC/97
2. Ahmad Hadji	28	M	S	MC/97
3. Arash Kamali	32	M	PS	LC/96
4. Farah Bibar	37	F	PS	MC/98
5. Farzad Shad	36	M	PS	MC/97
6. Haron Malek	30	M	S	LC/97
7. Morad Abali	37	M	PS	MC/96
8. Mohsen Eghbali	35	M	PS	MC/96
9. Mohammad Karimi	29	M	PS	MC/96
10. Mehgran Mahdavi	33	M	PS	MC/96
11. Mehran Qorbani	45	M	S	LC/96
12. Morad Zarrin	29	M	PS	MC/96
13. Noorali Azad	34	M	PS	LC/96
14. Pira Dinrad	38	M	S	MC/96
15. Parviz Mirian	27	M	PS	MC/96
16. Ramin Farda	39	M	PS	UC/98
17. Salam Farzadi	44	M	PS	MC/96

* The English data were extracted from the *Linguistics Department Archives of Spoken Language Materials*. I gratefully acknowledge permission from Professor Shana Poplack to make use of these data housed at the Sociolinguistics Laboratory, University of Ottawa.

Table 3.1. Informant information (continued)

Pseudonym	Age	Sex	Ed.	SEC/YR	Y. in CN
<i>L2 (proficient, with TOEFL or equivalent of 550 or more)</i>					
1. Arash Kamali	32	M	PS	LC/97	5
2. Farah Bibar	37	F	PS	MC/98	6
3. Farzad Shad	36	M	PS	MC/98	6
4. Morad Abali	37	M	PS	MC/97	6
5. Marzieh Abdi	35	F	PS	MC/97	6
6. Mehdi Ghafari	25	M	PS	MC/98	5
7. Mohammad Karimi	29	M	PS	MC/98	6
8. Mehgran Mahdavi	33	M	PS	MC/97	5
9. Mazda Saradj	29	M	PS	UC/98	6
10. Morad Zarrin	29	M	PS	MC/98	6
11. Parviz Mirian	27	M	PS	MC/98	4
12. Ramin Farda	39	M	PS	UC/98	7
<i>L2 (less proficient with TOEFL or equivalent of 500 or less)</i>					
1. Ali Akbari	42	M	PS	MC/97	1;2
2. Ahmad Hadji	28	M	S	MC/97	1;11
3. Esmail Eslahi	39	M	PS	LC/97	2
4. Fati Akbari	37	F	PS	MC/97	1;2
5. Farnaz Hadji	27	F	PS	MC/97	1;1
6. Haron Malek	30	M	S	LC/97	1
7. Habib Semat	39	M	PS	LC/98	1;1
8. Karim Gorbani	31	M	PS	MC/98	0;11
9. Kobra Mirban	26	F	PS	MC/97	1;10
10. Karam Shadi	31	M	PS	MC/99	0;9

Ed. = Education*F* = female*S* = Secondary*LC* = lower class*L1* = first language of the informant*III* = 25-34*IV* = 35-44*MC* = middle class*PS* = post secondary*SEC/YR* = Socio-economic class/ year of interview*UC* = upper class*Y. in CN* = years in Canada*M* = male**** Bolded individuals were interviewed in both Persian and English**

3.2.2. Persian informants

The Persian community in Ottawa

Unofficial statistics (from the Iranian embassy, consular section) indicate that there are about 3000 Iranians living in Ottawa, the national capital of Canada. Some of the members of the Iranian community have languages other than Persian as their first language (Turkish, Kurdish, Arabic, etc.), and some others were born or raised in Canada and have learned English or French as their first language. All such speakers were excluded from the present study; therefore, the speech community studied in this thesis includes only those Iranians whose first language is claimed to be Persian.

Many of these Iranians enjoy regular get-togethers by holding various social-cultural events throughout the year. They also have a few socio-cultural groups, among which the Persian Cultural Group is one of the most active. This group, which includes the majority of the educated Iranians and graduate students at the university of Ottawa and Carleton University, has regular monthly meetings to celebrate Iranian culture, hold group discussions on scientific or cultural subjects, or have parties.

Persian Informants

The informants for the Persian corpus comprise 17 native speakers of Persian (Table 3.1). The youngest informant is 27 and the oldest is 44. They are all native speakers of Persian, having learned it as their first language. These informants (and also the L2 informants) are identified (in the examples) by the initials of the pseudonyms assigned to them by the researcher.

3.2.3. L2

Assessment of English proficiency

One of the goals of this study is to investigate the interaction between patterns of variation and acquisition/transfer; therefore, I followed a cross-sectional data collection design by gathering data from speakers at two levels of proficiency in English. Twelve speakers were chosen to form the more proficient group of informants and ten others comprise the lower proficiency group. The two groups of informants were distinguished

on the basis of their scores on the Test of English as a Foreign Language (TOEFL) or an equivalent²¹, their length of residency in English speaking communities, and judgements of two SLA experts in (mostly) cases where the informants did not have a test score.

For the purposes of this thesis, an informant was considered to be proficient if his/her TOEFL or equivalent score was at least 550, resided in (English speaking) Canada for more than four years, and used spoken and written English at school, business, and in the marketplace. A low proficiency informant was an individual with a TOEFL or equivalent score of less than 440, resided in Canada for less than two years, and categorized as less proficient by judges who listened to excerpts from his/her recorded speech.

L2: proficient informants

The advanced L2 corpus of the study includes 12 informants (Table 3.1). Nine of the L2 informants were interviewed twice: once in English to elicit L2 data, and for the second time, in a different session, in Persian to collect samples of native Persian speech. Three other informants were interviewed only in English. The twelve L2 informants range in age from 25 to 39. They are all educated, highly proficient speakers of English. They have entered Canadian universities (mostly at graduate levels) with a TOEFL or equivalent score of at least 550.

L2: less proficient informants

Speech samples from ten less proficient speakers of English comprise my less proficient L2 corpus (see Table 3.1). Five of these speakers who have taken TOEFL (or the equivalent) were not able to obtain a score over 500 (440 or less), and they have all lived in Canada (at the time of interview) for two years or less. Eight of these speakers have university degrees at the undergraduate level and only one has been registered as a special student at the graduate level (because of a language problems).

²¹ Such as the CANTest or CAEL university admission assessment tests taken by University of Ottawa and Carleton University respectively.

3.2.4. English

For the English corpus, the naturalistic data of the *Linguistic Department Archives of Spoken Language Materials*, housed at the Sociolinguistics Laboratory, University of Ottawa (I will identify this corpus as Ottawa Spoken Language Archive (OSLA)), and collected over the past 17 years by the students of LIN 3141 from native speakers of English residing in Ottawa, Canada, comprise my source of the data. Of ten native English informants (Table 3.1), eight have university degrees and two are less educated. They range in age from 24 to 44. The English informants are identified by their original numerical three-digit codes assigned to them in the Sociolinguistic Lab Archives.

3.3. The Data

The (L2) data used in this study are extracted from conversational L2 speech of Persian speakers of English as a second language. In chapter two, I argued for the advantages of using conversational data in this thesis. Further, use of spontaneous speech is well-justified in sociolinguistics (variationist studies in particular) from which the present study of variation in L2 takes its direction. The *vernacular*, “the relatively homogeneous, spontaneous speech reserved for intimate or casual situations” (Poplack 1993: 252) is taken to reflect the most “systematic form of the language acquired by the speaker, prior to any subsequent efforts at (hyper-) correction or style-shifting.” (Poplack 1993: 252). Not only have previous studies of L2 speech variation (Tarone 1983, 1985) shown that L2 learners, like their native counterparts, are sensitive to the formality of the speech situation, i.e., the more natural their speech the more systematic it is, but also, according to Selinker (1972) the only appropriate and psychologically-relevant data for SLA studies were “observable data”, by which he meant utterances produced by learners in meaningful performance situations.

Spontaneous data, unconscious, unreflecting productions (Labov 1972), can be collected in many different ways. Complete spontaneity can be assured if the learner is completely unaware that s/he is observed. The participant observation technique, where the researcher (or the one who collects the data) is a member of the speech community and is able to collect the data in its most natural context, can help the SLA researcher to obtain spontaneous data as well. In some situations, however, this is not possible;

variationists, use other techniques of collecting natural data such as sociolinguistic interviews, where the informant is involved in a relaxed, informal atmosphere, negotiating meaning without focusing on producing language forms.

3.3.1. Data collection and coding procedures

Both the above mentioned techniques were used in order to collect data used in this thesis. Data were collected by the researcher among speakers with whom he shares the same social network and fulfills the conditions for in-group membership. As an in-group member, it was not a difficult task to contact other members of the Persian community in Ottawa and ask them for informal interviews. All the informants were interested in the study and the interviews, in most cases, were performed in a very informal atmosphere. Moreover, to circumvent the Observer's Paradox -- which involves use of devices that minimize the effects of the interview situation (presence of the interviewer in particular (Labov 1984; Poplack 1989)) -- interview protocols (appendix A) were formulated so as to enable the informants to recreate some emotional experiences and consequently pay less attention to their speech.

Most of the interviews were conducted at the informants' homes and the length of the recorded interviews was about one hour for each informant, with some interviews continuing for more than 3 hours. Before recording the interview, the interviewer set the scene by first explaining the purpose of the study²² and asking a few general questions about mostly the informant's work or background, then by initiation of a topic of interest I started to record the interview. In some cases the interviews were so friendly and informal that the informants forgot they were being interviewed (and recorded), so they wanted to get up and answer a phone, get something, etc., while they had the lavalier microphone on them. After the interviews we would often continue our conversation for about 10-20 minutes.

In about five cases the data were collected via an informal group conversation with two or more participants. In these situations the microphone(s) was placed on a stand somewhere close to all the informants. Immediately after each session (interview or

²² In most cases I explained the purpose of the study when arranging the interview schedule with the informants. I usually told them that this study was to explore the similarities and differences between English spoken by speakers of different L1s and their native and target languages.

conversation) I took notes about the individuals involved, their position during the interview, or other relevant factors.

All sentences with auxiliary verbs (only the first 60 minutes) (contracted/uncontracted/zero (where auxiliary was obligatory but not present)) and all relative clauses were extracted, transcribed, and (after removing the excluded items) coded according to the factors (and factor groups) explained in the following chapters. In general, each factor in a factor group was given a code (alphabetic, numeric etc.)(e.g., 'a' for absence or 'p' for presence of an item, etc). At the end, each variant included a coding string (one code representing one factor for each factor group) that formed the input into the variable rule program. Transcriptions were made first manually, then entered into Goldvarb 2.0, a variable rule application for the Macintosh, resulting in seven separate token files: English auxiliaries and relative clauses, Persian auxiliaries and relative clauses, L2 (proficient) auxiliaries and relative clauses, and L2 (low proficient) auxiliary token files.

All ambiguous items in English or Persian contexts (if a clause is restrictive or non-restrictive, syntactic function of antecedents, type of the clause, etc.) were categorized in consultation with native speakers of the respective languages.

3.3.2. The corpora

The data of this study consist of four sets of corpora:

- A native English corpus, which I will refer to as **native English**, **English L1** or the **target** language.
- A native Persian corpus, referred to as **native Persian**, **L1** or **Persian** corpus.
- Data from proficient Persian learners of English, referred to as **proficient L2** data or **EL2**.
- Data from less proficient informants, which I will refer to as **less proficient** data or **LEL2**.

Moreover, **English contexts** will be used to refer to native English, EL2 and LEL2 contexts, and **L2** to refer to data from both proficient and less proficient learners of English.

3.4. The Variables

In order to achieve my objectives, I have selected two linguistic features whose alternate forms are variably used in the English L2 speech of Persian learners, Persian (their L1) and English (target language), namely, variable contraction of auxiliary verbs and variation in the use of relative markers and resumptive pronouns in restrictive relative clauses. Not only are these two features used variably in L1, L2 and English, but their (variable) usage is both similar in some ways and different in some other ways in Persian and English. The effects of patterns of variability of the native and target languages can therefore clearly be examined on the variable behavior of L2 learners, particularly in the contexts where the native and target language patterns contrast with each other.

Auxiliaries

Auxiliary verbs, in general, are frequently used in several constructions (negatives, indicatives, etc) in different languages. They are among the earliest (and the easiest) grammatical morphemes acquired by first and second language learners (Dulay & Burt 1973; Krashen 1977; Lightbown 1987, for English). On the one hand, in first language acquisition and studies on dialects of English, auxiliary verbs have been found to be of particular interest. In first language acquisition, variation in the rate of learning auxiliaries has been found to be related to the frequency of auxiliary verbs in input (L1) (Richards 1990); and that frequency of auxiliaries at a certain stage (of L1 acquisition) might be associated with a distinctive style or route of language development (e.g., impulsive (faster learner, more frequent auxiliaries) or reflective (less frequent but more integrated system)) (Richards 1990). In studies of English dialects, such as Black English, variation in the use and contraction/deletion of auxiliary verbs has been used as a diagnostic of the origin of this dialect of English (e.g., Edwards & Winford 1991). On the other hand, despite their early acquisition, scarcity of auxiliary contraction in L2 speech is one of the distinctive features of “foreigner talk”, easily detected by native speakers of English (Meechan 1996). However, there has not been any study to account for the variable use of contracted forms of auxiliary verbs in the English spoken by L2 learners. My focus, here, will be to empirically examine the patterns of variability in

variable contraction (and deletion, if any) of auxiliary/copula verbs in the English spoken by Persian speakers. I also submit data from *less* proficient learners of English to variable rule analysis because of the frequent use of auxiliary verbs by these speakers.

Comparison of the results among four contexts of L1, EL2, LEL2, and native English will demonstrate if the patterns of variation in L2 speech change as acquisition proceeds, and what the effects of first and target languages are on L2 variation at two levels of lower and higher proficiencies.

More specifically, this study analyzes the behavior of auxiliary verbs including:

1. An analysis of the variable contraction/deletion of auxiliary verbs in the three languages and a quantitative comparison of the constraints on variation and their hierarchy, including the following:

a. Variable contraction/deletion of copula/auxiliary²³ *BE* (present forms) in English, EL2, LEL2, and its counterpart in Persian, exemplified in the following:

2. My mom and dad *are* healthy, thank God. (OSLA 246. Ib. 154)²⁴

3. Smoking's allowed in the buildings. (EL2. MS. Ia. 355)

4. *kare man inja: ru darma:n ast; montaha ru shimi darma:ni-e* (Per. FS. Ib. 400)

work my here on treatment is; but on chemo treatment-is

My work here is on therapy; it's, however, on chemo-therapy.

In the initial clause in 4, *ast* 'be' is used in its full form, but in the next clause it is contracted to *_e*.

²³ Following other studies of copula/auxiliary contraction in English and their findings with respect to the similar behavior of these items, I do not differentiate between the two in this study, unless our results indicate otherwise.

²⁴ Examples are identified by context (Ottawa Spoken Language Archives (OSLA), Persian (per.), EL2, or LEL2), speaker code, tape and counter number. English data were taken from the *Spoken Language Archives of the Linguistic Department of University of Ottawa*; therefore the same numerical codes, assigned to the speakers as their identification codes in Sociolinguistic Lab, are used as speaker codes here. Persian and L2 speakers are identified by their pseudonym initials.

b. Variable contraction/deletion of auxiliary forms of *HAVE/HAS/HAD* in English, L2 and their counterparts in Persian as in 5, 6, 7, 8, and 9, where Persian auxiliary *ast*²⁵ has been deleted.

- | | |
|---|-----------|
| 5. She's never seen me like that. (OSLA 246. Ia. 65) | English |
| 6. We'd always fixed her car up. (OSLA 247. Ia. 134) | English |
| 7. I've never got tickets for speeding (EL2. MS. Ia. 80) | L2 speech |
| 8. I'd never had bad fever. (EL2. MG. Ia. 238) | L2 speech |
| 9. <i>albatte ye ja:-ha:i taqiira:te omdeh budeh Ø.</i> (Per. RF. Iib. 148) | Persian |
- of course one place-s changes major been be

Of course, there have been some major changes in some places.

c. Variable contraction of *WILL/WOULD* in English and L2 (Persian structures are different, with no contraction).

- | | |
|--|-----------|
| 10. I'll never do it again. (OSLA 246. Ia. 132) | English |
| 11. He will see the results of that later on. (EL2. MK. Iib.139) | L2 speech |

Relative clauses

Acquisition of relative clauses has been one the most widely investigated structures in studies on the influence of prior linguistic knowledge on L2 learning (Gass 1979; Gass & Ard 1980, 1984; Tarallo & Myhill 1983; Flynn 1989; to name a few). Unlike auxiliaries, relative clauses comprise a complex area of syntax, the acquisition of which may occur in the later stages of L2 acquisition with a slower rate of progress. Many studies, the present thesis included, have thus focused on data from only advanced L2 learners when studying the acquisition of relative clauses. Variable use of elements of these constructions (relative markers, for instance), on the other hand, have also been the topic of several variationist studies (e.g., Guy & Bayley 1995) in different dialects of native English.

²⁵ Persian uses different forms of the auxiliary verb *astan* 'to be/have' to mark perfective forms. (see chapter 4 for details).

In order to discover patterns of variation in relative clause usage in L2 and the influences of learners' native and target languages on this variation, I will perform a quantitative analysis on the variable use of relative and resumptive pronouns in restrictive relative clause constructions in the English spoken by Persian speakers, and compare the results with those obtained from similar analyses on natural speech from native Persian and English speakers. The specific contexts under investigation include the following:

1. Variable use of relative markers in native English, EL2 speech and their Persian counterparts, exemplified in the following:

a. English: alternate use of *THAT*, *ZERO* and *WH* relative pronouns:

12. There's different thing *that* you get from different people. (OSLA 241. Ib. 26)

13. There is nothing *Ø* I really want to become. (OSLA 241. Ia. 42)

14. I hate people *who* play that game. (OSLA 241.Ib. 253)

b. L2 speech: alternate use of *THAT*, *ZERO* and *WH* relative pronouns:

15. It's a question *that* I think many people ask themselves. (EL2. RF. Ib. 391)

16. The first person *Ø* I talked to was him. (EL2. MS. Ia. 167)

17. I picked a spot *which* was actually a little bit higher from the ground. (EL2. MG. Ib. 505)

c. Persian: alternate use of a general relative marker *ke* 'that' and *ZERO*:

18. *shoma: har sistemi-ro, har chizi-ro Ø mixa:y mota:leeh koni esmesh-o miza:ri sistem.* (Per. SP. Ia. 327)

you any system-OM any thing-OM want study do name-OM put system.

You call (any system) any thing you want to study a system.

19. *man avvalin peyperi-ro ke cha:p kardam natijeye ka:ra:m tu dabiresta:n-e.* (Per. FS. Ib. 380)

I first paper-OM that print do result works in high school-is
The first paper that I published was the result of my works in high school.

2. Variable use of resumptive pronouns in native English, L2 speech and native Persian, as in 20, 21 and 22 respectively.

20. The side street that I was on *it* joined another street. (OSLA 268. Ia. 226)²⁶

21. It was this stream water you would fish in *it*. (EL2 RF. Ib. 151)

22. *da:ran jensa:i ke xatti rush-e*. (EL2 ME. Ia. 73)

have things that line on-is

They have items that have some scratches on (them).

In sum, this thesis is a variationist study of the variable behavior of two grammatical features observed to occur in the English spoken by Persian speakers. Several forms of auxiliaries are learned earlier and pose no serious difficulty for many language learners. In contrast to relative clauses, they are robust in natural speech. The second feature, relative clauses, is a complex area of syntax usually acquired in later stages of L2 acquisition. It involves acquisition of a complex of various syntactic items and operations, and thus is not very often used by less proficient learners.

²⁶ Phonologically 'it' in this sentence belongs to the relative clause.

Chapter four: AUXILIARY VERBS

4.0. Introduction

One feature of English speech is the variation between full and contracted auxiliary/copula verbs, as in 23 & 24 respectively. Contraction in English is so associated with natural speech that the lack of it is considered unusual (see Meechan 1996).

23. My mom and dad *are* healthy, thank God. (OSLA 246. Ib. 154)

24. She's never seen me like that. (OSLA 246. Ia. 65)

Since Labov's 1969 study on the English copula, there has been a great deal of research on the factors conditioning auxiliary contraction/deletion in various dialects of English (Baugh 1980; McElhinny 1993; Meechan 1996; Poplack & Sankoff 1987; Walker 1999; among others), indicating the prevalence and importance of this variability in the English language.

Do second language learners acquire the highly variable contraction of auxiliary verbs of English? What we understand from research on acquisition of the English auxiliary system by second language learners (e.g., Krashen 1977; Lightbown 1987) is that copula/auxiliary *BE*, at least, are among the first grammatical morphemes that both child and adult L2 learners find easy to acquire. Despite early acquisition of auxiliary verbs, however, we know very little about if and how L2 learners acquire target language *variable patterns*, in general, and variable contraction of auxiliaries in particular. The reason for this is, on the one hand, the lack of studies concentrating on the acquisition and variable use of forms of auxiliary verbs in SLA. Much of our knowledge comes from investigations into areas in which the auxiliary was not the primary focus of attention (in SLA) (such as studies of negation, question formation, etc.). On the other hand, studies that have investigated how learners acquire English auxiliaries have treated their rules as categorical (e.g., Cacino et al. 1974; Ferguson 1968; Shapira 1978). As Ellis (1987a) asserts, they have sought to determine whether acquisition has or has not taken place, by seeing if learners supply the auxiliary in obligatory contexts.

Mention was made (chapter one) that for many researchers successful SLA involves the acquisition of target language variable rules (Ellis 1987a). Thus, use of non-

contracted auxiliaries by L2 learners in the utterances where contraction naturally occurs in native English speech does not indicate a native-like acquisition which involves acquisition of both syntactic features of the target language and its patterns of variation. The question is how can we demonstrate whether a variable feature of a target language is acquired by its L2 learners? And if variation is acquired, are the constraints on variation in the target language acquired along with it? What role does a learner's native language play in this respect?

This chapter is an attempt to answer these questions in a quantitative manner. My focus here is to compare and contrast the frequency of use and constraints on the variable contraction²⁷ of auxiliary verbs in English spoken by Persian speakers with those found in native (Persian) and target (English) languages. The results of the comparisons between L1, L2, and native English will help us determine how variable L2 is and what the sources of this variation are. Furthermore, comparison of the behavior of less proficient speakers with those of other informants will demonstrate if the patterns of variation in L2 speech change as acquisition proceeds, and what the effects of first and target languages are in this respect.

In what follows I will first present a summary of the status of auxiliary verbs in the respective languages, then the results of the variable rule analyses will follow. A discussion of the results will end chapter four.

4.1. Auxiliaries in English, Persian, and EL2

4.1.1. English auxiliaries

Quirk et al (1985) categorize English auxiliaries as primary (auxiliaries), which include forms of *BE*, *HAVE* and *DO*, and modal auxiliaries including *CAN*, *MAY*, *WILL*, *WOULD*, etc. Contraction, defined as the removal of the word-initial vowel (schwa) before a long consonant to obtain a single syllable (Labov 1972; Wolfram & Fasold 1974; also see McElhinny 1993 for a discussion on definition), applies to forms of *BE* (23, 25-27), *HAVE* (24, 28), *WILL* (29) and *WOULD* (30) in vernacular English.

²⁷ The number of deleted auxiliaries in our L2 data was not frequent enough in many contexts to be submitted to a quantitative analysis.

25. I'*m* trying to find, get on a team or whatever. (OSLA 246.Ia. 158)
26. At the moment I *am* concentrating on printmaking. (OSLA 268.Ia. 63)
27. Sometimes it's a little tough just to be a good friend. (OSLA 239. Ia. 269)
28. We'*d* always fixed her car up. (OSLA 247. Ia. 134)
29. I'*ll* never do it again. (OSLA 246. Ia. 132)
30. Jack the duck *would* follow you and it'*d* chase you, and it'*d* grab you. (OSLA 246. Ia. 11)

As exemplified in section 4.2.1, contractions can not and do not occur in some contexts such as the beginning of the clause, in tags, where the operator is the only verb in the VP or precedes ellipsis, and some phonological environments. The categorically non-contracted contexts are not included in variationist studies of contraction. In this study I will explore the variable behavior of *BE* (present forms: *AM*, *IS*, *ARE*), forms of *HAVE* (*HAVE*, *HAS*, *HAD*), *WILL* and *WOULD* in both English and L2 speech.

4.1.2. Persian auxiliaries

Persian is a SOV language, with auxiliaries following the verb. Because of the application of certain re-ordering rules (Karimi 1989), other word order patterns such as SVO, VSO, etc. may also occur in spoken Persian. This language enjoys a rich inflectional system, thus it allows null subjects. The only auxiliary verb in the language is the different forms of the auxiliary and/or copula verb *budan* 'to be' (Vahedi 1992). In the constructions with past temporal reference, different forms of *bud* 'be, past' are categorically (no contraction) used; and in sentences with present temporal reference, the infinitival *budan* appears as (*h*)*astan*²⁸ 'to be, present'. In its present form, *astan* has a double series of forms, one contracted (enclitic, Lazard 1992) and the other non-contracted, as in 31, with examples from my Persian corpus in 32 & 33. In the simple present tense *astan* appears variably in its full or contracted forms (32 & 33) for all persons. Its present progressive forms are always contracted to the verb.

²⁸ In modern Persian there is no infinitival such as (*h*)*astan* 'to be, present'; this form, however, is used by grammarians to refer to only the present form of the auxiliary. Vahedi (1992) believes that *astan* may be an archaic variant of *budan* 'to be'. /h/ in *hastan* is pronounced in emphatic environments, to refer to the

31. Non-contracted// contracted	Non-contracted// contracted
<i>Man hastam</i> // <i>man -am</i> I am	<i>Ma: hastim</i> // <i>ma-im</i> we are
<i>To hasti</i> // <i>to-i</i> you are	<i>Shoma: hastid</i> // <i>shoma-id</i> you are
<i>U hast</i> // <i>un-e</i> it is	<i>Unha: hastand</i> // <i>una:-an (d)</i> they are

32. *kare man inja: ru darma:n ast; montaha ru shimi darma:ni-e* (contracted)
(Per. FS. Ib. 400)

work my here on treatment is; but on chemo treatment-is

My work here is on therapy; it's, however, on chemo-therapy.

33. *xod-e man nemune ba:rezi-am ke 19 sa:legi, 20 sa:legi ezdeva:j kardam*
(Per. FS. IIb. 237)

myself example clear-'m that 19 years, 20 years old married did.

I ('m) myself am a clear example that got married at 19, 20.

In the present perfect, contracted forms of *astan* (here glossed as 'to have'²⁹) are used as the auxiliary, as in 34 from Vahedi (1992). Full forms of the auxiliary verb might have been used in old or middle Persian (see Honar 1992) (as in 35), but in modern Persian they are no longer in use (no variability), except for the 3rd person singular forms.

34. Present perfect: past participle + contracted forms of the aux. (spoken forms in parentheses)

rafte-am >(raft-am) I have gone *rafte-im* >(raft-im) we have gone

rafte-i >(raft-i) you have gone *rafte-id* >(raft-id) you have gone

*rafte-ast*³⁰ >(raft-Ø) s/he has gone *rafte-and* >(raft-an(d)) they have gone

35. *rafte hastam/rafte hasti/rafte hast* (I have gone/you have gone/s/he has gone)

existential connotation of the verb, or in some phonological contexts. In spoken Persian it is usually deleted.

²⁹ Identical to English auxiliary *have*, *astan* does not contribute to the meaning of the sentence, but marks the structure for its tense/aspect (Infl.) features. It is distinguished from its simple present forms (*BE* forms) on the basis of its function and preceding item, which is always a past participle in present perfect forms.

³⁰ The status of the 3rd person singular present perfect form of the auxiliary verb is not without controversy. As mentioned in the text, it is believed that contracted forms of the auxiliary are used in the modern Persian present perfect system (e.g., Lazard 1992). But as seen in 31, *ast* is the non-contracted form of the auxiliary which is used in forming present perfect tense (as in 34). There are several possible accounts for this. One account is that *ast* in the present perfect is the non-contracted form of the simple present form of the auxiliary, unlike other persons that are mostly contracted. This form is contracted to *-e* in modern spoken Persian. But, since the past participle is vowel final, the contracted form of the auxiliary is naturally deleted.

As shown (in 34), the present perfect is formed from the past participle followed by the enclitic forms of the verb *astan*. It is only in the third person singular that variability exists between deletion and non-deletion of the auxiliary verb, as in 36 (undeleted) & 37 (deleted). For this reason I will include only tokens of third person singular perfect forms in my study of Persian present perfect auxiliary system. Therefore, in the case of the variable use of forms of HAS (3rd person), we are dealing with two different systems in English and Persian: English variably contracts and Persian variably deletes. This is a site where behavior of L2 learners (either contraction or deletion) will clearly indicate what system they follow.

36. *ka:r-ha:ye elmish hatta: anjam shode ast bara: inke ru xatte tolid bere*
(Per. ME. Ia. 254)

works scientific even do become is for this-that on line production go
Even all the scientific works has been done so it can go on the production line.

37. *az se sa:l pish umade Ø tu baza:r* (Per. ME. Ia. 215)

from three years ago come in market

It has been in the market for three years.

The only conditioning factor mentioned in the prescriptive literature for deletion of auxiliary verbs in present perfect forms is presence of a following equivalent form (the same auxiliary (parallel processing)) as in 38, where because of the presence of the auxiliary in the last sentence all the previous structures are used without an overt auxiliary. Semantic redundancy (i.e., meaning (temporal reference) of the form is implied from the participle or context) has also been mentioned as another reason for the auxiliary deletion (Nobaha:r 1993).

38. *raftan mardom gereftan, ye ira:di peyda: karde Ø, rafte Ø, tamir shode Ø, bar-gashte ast.* (Per. ME. Ia. 86)

gone people got one deficiency found do gone repair become
come back is

People have gone and got it, it has been found deficient, it has gone and been repaired and has come back.

In sum, the variables analyzed in this part of the study include variable contraction of the Persian auxiliary/copula 'be' in simple present (1st, 2nd and 3rd person

singular and plural forms) and variable deletion of the auxiliary in present perfect tenses (only 3rd person singular forms). Other forms of the auxiliaries in Persian³¹ do not behave variably, and so are not included in this study.

4.1.3. Auxiliaries in L2

Different forms of auxiliary verbs are variably contracted (or deleted) in the speech of second language learners of English, as in 39-41.

- 39. Smoking's allowed in the buildings. (EL2. MS. Ia. 355)
- 40. I've never got tickets for speeding (EL2. MS. Ia. 80)
- 41. He *will* see the results of that later on. (EL2. MK. IIb.139)

This variability has been ignored in several studies on the acquisition of auxiliary verbs by second language learners (Butterworth & Hatch 1978; Cancino et al. 1974, 1975; Ferguson 1968; Heckler 1983; Shapira 1978). What we know from these studies is that learners commit systematic, consistent deviations from target language norms (Sridhar 1981), what has been called an *error* in SLA. These errors (many of which are now accounted for as variable features of L2) are claimed to be due to either learners' native language effects (transfer) (Cancino et al. 1974; Butterworth & Hatch 1978 for auxiliary deletion in L2; see also Winford (1998) on this), or assumptions made on the basis of L1 = L2 hypothesis, since they resemble errors made by children acquiring their L1 (Butterworth & Hatch 1978). Shapira's frequency counts of L2 learners' errors in auxiliary usage also inform us that of forms of *BE*, *IS* has greater chance of deletion in L2 speech, and that its use as copula is more frequent than its function as an auxiliary.

In this study, I deal with 'systematic, consistent deviations' of L2 learners not as errors, but as linguistic variations. I then examine the patterns of variable usage of non-contracted/contracted/deleted forms of auxiliary verbs in L2 speech, and finally, try to find out where those patterns originated.

³¹ See Vahedi 1992 for the claim that there is no real modal in the Persian verbal system, and Windfuhr (1979) for detailed information on Persian modals.

4.2. Analyses

4.2.1. Variable contexts and exclusions

English and L2

The variable contexts in this study, summarized in 42, include declarative sentences in which present forms of *BE* (*AM*, *IS*, *ARE*), as well as auxiliary forms of *HAVE* (present or past), *WILL* and *WOULD* were used or could have been used in English L1, English L2 , and Persian.

42.	English	L2 ³²	Persian
<i>AM/IS/ARE</i>	full/contracted	full/contracted	full/contracted
<i>HAS</i>	full/contracted	full/contracted	full/deleted ³³
<i>HAVE/HAD</i>	full/contracted	full/contracted	no variation
<i>WILL/WOULD</i>	full/contracted	full/contracted	no variation

Following other studies of English (Labov 1969; Meechan 1996; Walker 1999), environments where contraction cannot occur were not considered. These include:

Past forms of <i>BE</i>	There <i>was</i> one girl. (OSLA 241. Ia. 395)
<i>BE</i> after modals	He <i>should be</i> more respectful. (OSLA 241.Ia. 85)
Emphatic forms	Washington D.C. <i>is</i> very very fast. (OSLA 241. Ia. 192)
Yes/no questions	<i>Is</i> this person, like, mentally screwed up? (OSLA 241. Ia. 333)
Tag questions	That's about the fee, <i>isn't</i> it ? (OSLA 103. Ib. 14) (From Meechan 1996)
Phrase final instances	You just get to know how the person <i>is</i> . (OSLA 241. Ib. 8)

HAVE, *WILL* and *WOULD* were only coded in their auxiliary forms; *HAVE* and *WILL* as main verbs (also *HAVE TO*) were not included. I also excluded ambiguous temporal

³² L2 may include tokens of deleted auxiliaries as well.

³³ In this context the important point is to show what option L2 learners choose: contraction as in English or deletion , as in Persian. If they prefer to contract instances of *HAS*, as in English, it would be interesting to see what patterns they utilize.

references (and all other ambiguous forms), as in 43, since it is impossible to tell whether the deleted form would be *IS* or *WAS*, as well as the few tokens of *AIN'T*, as in 44.

43. The day I graduated, probably $\emptyset$ the day my dad and mother retires. (OSLA 246. Ib. 156)
44. I *ain't* gonna tell you me secrets. (OSLA107. Ia. 382)

A few tokens of deletion (45, 46) occurred in both native English and L2 data sets. Following Labov (1969) and Meechan (1996), these tokens of zero variants were included with contracted variants in most of the analyses³⁴.

45. She $\emptyset$ not skinny but she was thinner than she is now. (OSLA 285. Ib. 165)
46. It's on the spot memorising, you $\emptyset$ constantly memorising. (OSLA 136. Ib. 156)

Persian

All tokens of simple present forms of *astan* 'to be' were extracted from the data. Verbs having existential meanings (such as 47) and tokens of auxiliary forms of *astan* followed by *ham* 'too' were categorically non-contracted (emphatic contexts, as in 48), they were then excluded from the analyses.

47. *ala:n tu tehra:n mahalleh-ha: xeyli kam hast ke dast naxordeh.* (Per. RF.IIb. 145)

Now in Teheran neighborhood-PL very few exist that hand hit
Now, there exist very few neighborhoods in Teheran that are intact.

48. *keta:b-e xeyli xubi ham hast.* (Per. PM.Ia. 330)

book very good too is It is a very good book, too.

As mentioned above, from Persian present perfect structures I will include only 3rd person singular forms of the equivalents of *HAS* as an auxiliary verb. This is because it is the only context that shows variable use of auxiliary (deletion/insertion) in Persian present perfect forms.

³⁴ In any case as shown in Tables 4.1, the percentage of *zero* variants is small and our preliminary analyses with or without zero tokens (in some contexts) did not show any difference in the behavior of the variants. Only in one context (*HAS* in LEL2) were deleted items dealt with as an effect of Persian.

4.2.2. Coding

The factor groups incorporated in the study of variable contraction of auxiliary verbs in English contexts (English, English L2) are mostly taken from previous research on the same variation in English. These factors have been found to influence variable contraction of auxiliaries in different dialects of English, and my English contexts (native English spoken in Ottawa and English spoken by L2 learners) are supposedly variants of the same language, unless otherwise found in the results of this study, i.e., it is found that one (or more) of the contexts, e.g., LEL2, is affected by factors other than those of other studies in English (possibly Persian factors).

There is no variationist analysis of the Persian variables studied in this dissertation. Grammar books have not paid attention to the variability because it is a feature of spoken Persian (not standard written Persian), although some of them have acknowledged its presence (Nobahar 1993). Since, on the one hand, my preliminary results indicated that some of the factors incorporated in the English contexts influence variability in Persian as well, and, on the other hand, incorporation of similar factors makes comparison of the behavior of the elements under study more reliable, these factors were also used in the analysis of variation in Persian. Factors specific to the Persian context (if any) are also added to the list of factor groups and dealt with when analyzing the behavior of each variable.

In the following section I introduce the factor groups that are incorporated in the study of all auxiliaries, in general. Of course, the different forms of the auxiliaries are not hypothesized to be conditioned by the same factors in each factor group. In the analysis of each individual variant, only the relevant factors were included, which will be reported in the presentation of the results.

4.2.2.1. Phonological environment³⁵

Contraction involves the reduction and deletion of a vowel. It is then hypothesized (Labov 1969 and others) that the preceding and following phonological environments could influence auxiliary contraction/deletion. The following phonological environment was coded as consonant, vowel, and, pause,³⁶ for Persian as Persian sentences are verb-final. The preceding environment was coded as vowel or consonant for all contexts³⁷.

4.2.2.2. Type of subject

Without exception, all previous variationist studies of auxiliary contraction in English have demonstrated the significant effects of the subject type. Labov (1969), McElhinny (1993), and Meechan (1996), among others, have found that preceding pronouns favor and preceding nouns strongly disfavor *BE* contraction in standard English. Poplack & Sankoff (1987) broke down subject type into different subcategories such as here/there, that/it/what, singular/plural, personal pronouns, etc. and found differences in their effects on copula contraction/deletion in Samana English.

We initially coded subjects as singular/plural nominal, personal pronoun (we, you, they, etc.), existential (here, there), demonstrative (these, those, them, this), wh-

³⁵ Meechan (1996) has studied the effects of preceding and following stress and heaviness of the preceding subject (number of words) on auxiliary contraction. Walker (1999) also pays special attention to the prosodic features of the preceding and following phonological environments. Walker found such a strong interaction between preceding phonological, prosodic and grammatical contexts that he had to collapse these three factors together. Because of the presence of this interaction, and the fact that phonological conditioning is not as relevant to the nature of the underlying grammar as grammatical factors, I will not go into the details of the effects of phonological environment in this study.

³⁶ A few tokens of following pause in (English and L2) data sets were coded as non-count in the analyses because of their infrequent use.

³⁷ I initially coded the preceding phonological environment for the following factors:

[r], since Fasold & Nakano (1996) found that preceding [r] favors contraction, even more than vowels, in standard English.

Sibilants, since Fasold & Nakano also found the less favorable effect of preceding sibilants on contraction.

Other consonants

Vowels

Diphthongs

But for the following reasons the preceding phonological factor group was reduced to the two factors of vowels and consonants: 1) in some environments there was no significant difference between the effects of two or more factors, 2) there was interaction between preceding phonology and other factors, and collapsing preceding phonological factors reduced (or removed) the interaction effects, and 3) some factors

pronoun, zero (for Persian), and that/it/what. Depending on the analyses and behavior of each subject type they were collapsed (and excluded in some cases, for reasons such as interaction etc.) into fewer groups of:

personal pronoun	personal pronoun	pronoun
other pronoun	other pronoun	NP
NP	NP	
that/it/what	Zero	
Zero		

There is disagreement among researchers as to whether or not tokens of *IT/WHAT/THAT* (which I coded as a separate factor) should be included in variable rule analysis (of *IS*), since such forms tend to be followed by invariably contracted forms of auxiliary (Meechan 1996; Walker 1999). Blake (1997), reviewing other studies that either included or excluded tokens of *IT/WHAT/THAT* subjects in their data concludes that these forms would be considered “don’t count” because of their categorical status. On the one hand, overall distribution of the data shows that the behavior of these items in the L2 contexts is different from native dialects of English (78% and 58% contraction in EL2 & LEL2 respectively vs. 88% in native English). On the other hand, as suggested by Meechan (1996), I excluded frozen (categorical) expressions like, ‘that’s right’, ‘it’s ok’, or ‘that’s it’ from my data. Therefore, following Poplack & Sankoff (1987), I included tokens of *IT/WHAT/THAT* subjects in my analyses. This will be elaborated more when reporting the results of variable rule analyses.

4.2.2.3. Following grammatical category

The effect of the following grammatical category on contraction/deletion of auxiliaries (*BE*, in particular) has received a great deal of attention since Labov (1969). Labov (1972), McElhinny (1993), and Meechan (1996) have all found that the following grammatical category contributes significantly to variable contraction of *IS*, with NPs with the least effect and *gonna*, locative and *V+ ing* with the most. In McElhinny’s study, this factor group was not selected as significant for *ARE*-contraction, while in Meechan

were not justified by the number of the tokens in some contexts; then for the sake of comparison, those factors were collapsed with other related factors in all contexts.

(1996) it was selected with locatives as the most favorable and NPs as the least favorable contexts for contraction. In light of the claims made for the effects of following grammatical category, I included this factor group in the current study, with the following factors (Labov 1972; McElhinny 1993; Meechan 1996; Poplack & Sankoff 1987; Poplack & Tagliamonte 1991; Walker 1999).

- | | |
|---|--|
| 49. He is <i>acting</i> younger actually (OSLA 268. Ia. 403) | Following V-ing |
| 50. Oh, my mom's <i>gonna</i> kill me. (OSLA 107. Ib. 298) | Following gonna (going to) |
| 51. He's much <i>happier</i> now. (OSLA 268. Ia. 404) | Following adjective ³⁸ |
| 52. The street itself isn't <i>repaired</i> quite as often as some of these city streets are. (OSLA 133.Ia. 38) | Following participle |
| 53. There's also <i>stories</i> about my grandfather. (OSLA 107. Ib. 136) | Following NP |
| 54. Most of my work is <i>at the airport</i> . (OSLA 285. Ia. 323) | Following locative |
| 55. That's <i>how they picked it up</i> . (OSLA 285. Ib. 379) | Following wh-clause |

4.2.2.4. Underlying form of the auxiliary

Some previous research has examined the behavior of forms of *BE* (*AM*, *IS*, *ARE*) together (e.g., Poplack & Sankoff 1987), and other work has already shown that morpholexical properties of each underlying form of *BE* have different effects (Meechan 1996, Walker 1998). In some studies (as in Labov 1969; Meechan 1996) *AM* contracts so much that it is excluded from the analysis, while in Poplack & Sankoff (1987) *AM* contracts less than *IS*.

I analyze the behavior of each form separately. Type of verb was coded as *AM*, *ARE*, or *IS* (for *BE*); *HAVE*, *HAS*, *HAD*; *WILL* and *WOULD* (and equivalent forms of *AM*, *IS*, *ARE* and *HAS* in Persian).

In addition to the above factor groups³⁹, the following were added to the coding of Persian auxiliaries:

³⁸ Following adjectives and participles were combined as a single factor in some analyses because of their similar behavior; following NPs and wh-clauses were also collapsed for the same reason.

³⁹ When examining the effects of following complement after *WILL/WOULD* + *BE* on contraction/deletion, McElhinny asserts that likelihood of a grammatical effect on a phonological process from such a distant environment (AUX + be + (ADV) COMP) might seem implausible. She, however, finds that adjective is the least favorable environment for contraction of *WILL/WOULD* + *BE*. This observation might also hold

- Since Persian is a pro-drop language, a factor of zero subjects was added to the subject type factor group.
- As an SOV language, Persian auxiliary verbs are mostly sentence-final. The following grammatical category for Persian includes a zero factor as well. For this reason, following phonological environment for Persian includes pause as an additional factor.
- A preceding grammatical factor group, with factors of preceding NP, adjective, adverb, pronoun, negative, participle, and other, is added and applies to the Persian context only.

If the L2 is behaving like English with respect to the contraction of any of the auxiliary verbs, the factors introduced above should have similar effects both in English and L2. If it is behaving like Persian, the patterns of variation should mirror those of Persian. If it is found that the system of variability in L2 is different from both English and Persian, its independent status as an interlanguage will be quantitatively shown.

4.2.3. Data

All structures in which forms of auxiliary verbs were used (full, contracted, zero) were extracted from the four corpora representing: native English, EL2, LEL2, and native Persian (about 60 minutes of speech per informant). All tokens of auxiliaries were coded according to the phonological and syntactic factors. The variable rule program, Goldvarb 2.0, an application for the Macintosh (Rand & Sankoff 1990) was used to perform (where possible) separate analyses on different types of auxiliaries. Overall distribution of the data is represented in Table 4.1. Overall distribution of auxiliary verbs is almost identical in all contexts: *IS* with the most and *WILL* with the least frequency of occurrence. As shown, *AM* is the most favorable form for contraction and *WOULD* has the fewest contracted forms in English contexts (English, EL2, LEL2).

for cases where there is some other element between *BE/HAVE/WILL/WOULD* and following grammatical category, as in (i), where 'always' intervenes between *HAVE* and following participle.

i. I've always liked it. (OSLA 285 Ib.301)

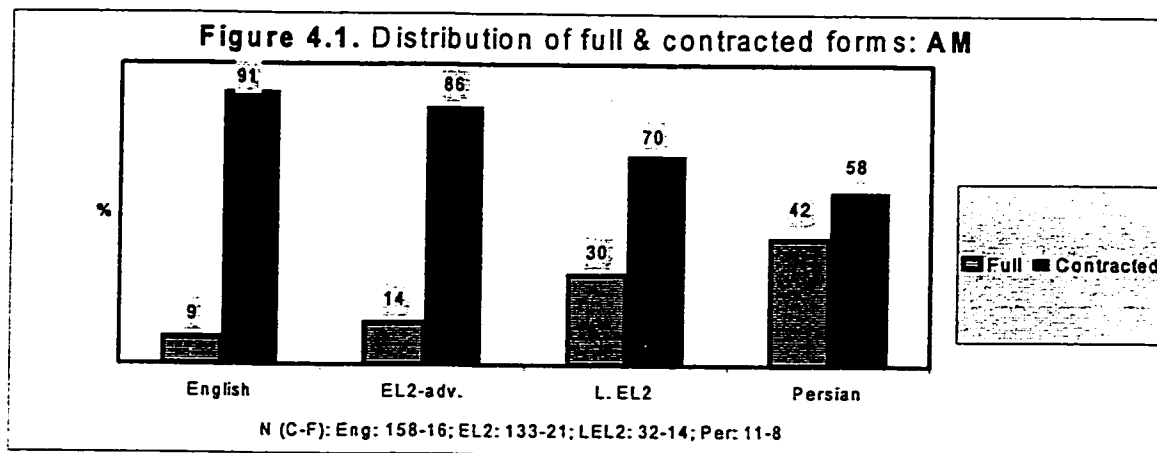
In order to find the effect of these intervening items, I coded for presence of an intervening item between the auxiliary and following grammatical category. This factor group was found to have no effect whatsoever on contraction of auxiliary verbs; therefore it will not be further discussed here.

Table 4.1. Overall distribution of forms of auxiliary verbs in English, EL2, LEL2 and Persian.

Form of auxiliary		English		EL2-adv.		LEL2		Persian	
		N	%	N	%	N	%	N	%
AM	Full	16	9	21	14	14	30	8	42
	Contracted	157	90	130	84	29	63	11	58
	Zero	1	1	3	2	3	7	Ø	Ø
	Total*	174/8		154/8		46/7		19/1	
IS	Full	237	25	447	46	269	66	261	32
	Contracted	688	73	519	53	87	21	539	66
	Zero	19	2	9	1	49	12	20	2
	Total	944/44		975/52		405/65		820/48	
ARE	Full	132	34	266	79	80	82	35	47
	Contracted	234	60	51	15	3	3	39	53
	Zero	25	6	18	5	15	15	Ø	Ø
	Total	391/18		335/18		98/16		74/5	
HAVE/HAS/HAD	Full	62	22	98	75	8	53	7	8
	Contracted	219	77	33	25	4	27	Ø	Ø
	Zero	4	1	Ø	Ø	3	20	86	92
	Total	285/13		131/6		15/2		93/5	
WILL	Full	22	25	79	89	40	87	NA	
	Contracted	66	75	9	10	2	5		
	Zero	Ø	Ø	1	1	3	7		
	Total	88/4		89/5		45/8			
WOULD	Full	140	55	176	99	10	83	NA	
	Contracted	116	45	2	1	1	8		
	Zero	Ø	Ø	Ø	Ø	1	8		
	Total	256/12		178/10		12/2			
Total	All forms	2150		1861		621		1006	

*Total of each auxiliary. Percentages calculated out of the total number of auxiliaries.

Before discussing the analyses, mention should be made that when one variant is overwhelming in the data (e.g., in case of almost categorical contraction of *AM* in English), variable rule analysis was not undertaken.



4.2.4. Results

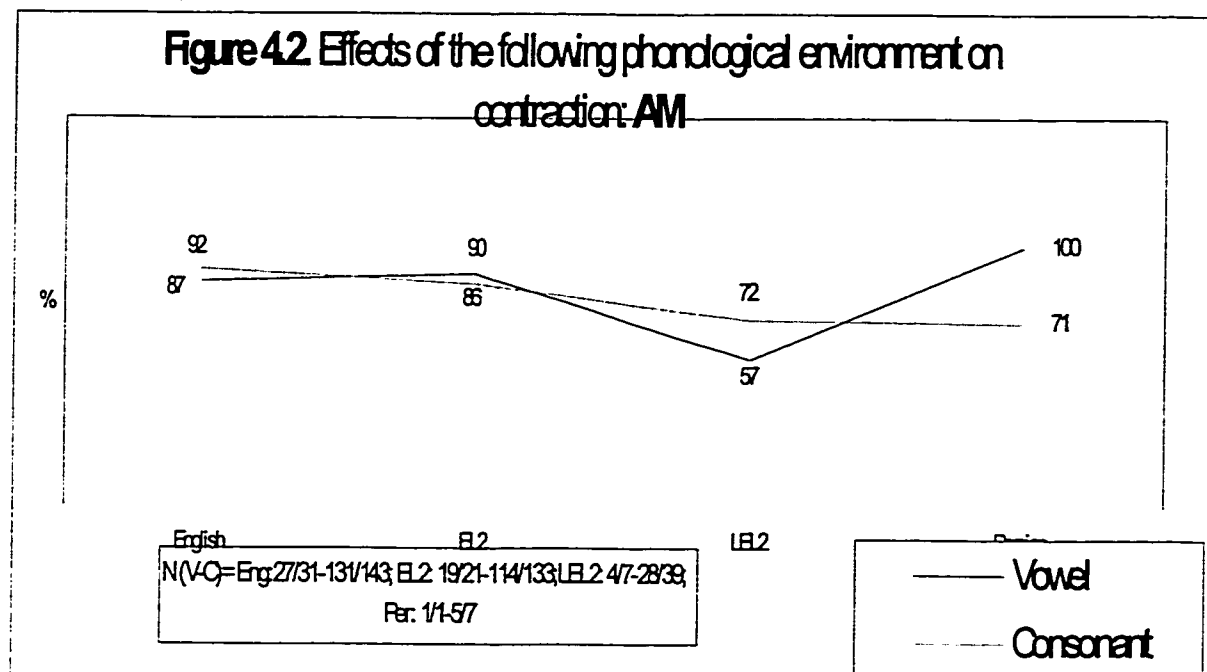
AM

91% of tokens of *AM* in the English corpus are contracted (including one zero form) (Figure 4.1)⁴⁰; therefore, variable rule analysis was not performed. There is a correspondingly high percentage of contraction in EL2 (86%), as well. In Persian, on the other hand, contraction and non-contraction are almost equal. As shown in Figure 4.1, contraction is greater in English and less in Persian. Frequency of contraction in LEL2 is closer to that of Persian (70% and 58% respectively). There seems to be a gradual increase of contraction from Persian to English.

Since *AM* is always preceded by a personal pronoun (*I*) as the subject of the sentence and a diphthong as preceding phonological environment in all English contexts (EL2, LEL2, English), I report the effects of the following grammatical and phonological environments on the data. Persian is different in this regard since the subject may be either a personal pronoun (*man* 'I') or zero.

⁴⁰ These abbreviations are used in the figures: N = number, C = contracted, F = full, / = out of.

As shown in Figure 4.2, English and EL2 do not differ concerning the effect of following phonological segments. LEL2, on the other hand, behaves differently. After vowels (V), *AM* contracts less in LEL2, while in Persian⁴¹ contraction is categorical. Following consonants (C), however, have the same effects in both Persian and LEL2.



The effects of the following grammatical items on *AM*-contraction are reported in Table 4.2. As seen, except for locatives, the percentages for the effect of the following grammatical category are rather similar in all English contexts. The different effect of following locatives on contraction in LEL2 might be attributed to the scarcity of the tokens in this context (only one token); otherwise L2 learners do not seem that much different from native English speakers. Persian is totally different, since as a SOV language with auxiliary in sentence-final location, only some verbs are followed by re-ordered grammatical items⁴² or subordinate clauses. In these cases contraction is encouraged in Persian (75%). Moreover, my results for the effect of the preceding phonological environment (Table 4.3) indicate that in Persian, a preceding vowel

⁴¹ Although the number of the tokens in Persian is small.

⁴² In Persian it is possible for some items such as direct/indirect objects, adverbs, etc. to appear post-verbally.

categorically disfavors contraction and zero subjects favor contraction 80% of the time, unlike English contexts where the effects of a preceding diphthong is not categorical (70% in LEL2 to 91% in native English), but in favor of contraction. It is noteworthy that L2 learners did not have any null-subject sentences in their L2 speech.

Table 4.2. *Effects of the following grammatical item on contraction : AM*

	English		EL2		LEL2		Persian	
	N	%	N	%	N	%	N	%
Gonna	9	100	5	100	3	100	-----	
Locative	15	100	7	78	1	33	-----	
V+ ing	49	92	61	87	9	75	-----	
Adjective	51	91	52	88	13	68	-----	
NP	34	83	8	73	6	67	-----	
Zero	-----		-----		-----		5	45
Other	-----		-----		-----		6	75
Total	158	91	133	86	32	70	11	58

Table 4.3. *Effects of the preceding environment on contraction: AM*

	English		EL2		LEL2		Persian	
	N	%	N	%	N	%	N	%
Preceding diphthong/vowel	158	91	133	86	32	70	Ø	Ø
Preceding zero	-----		-----		-----		4	80
Preceding consonant	-----		-----		-----		11	65

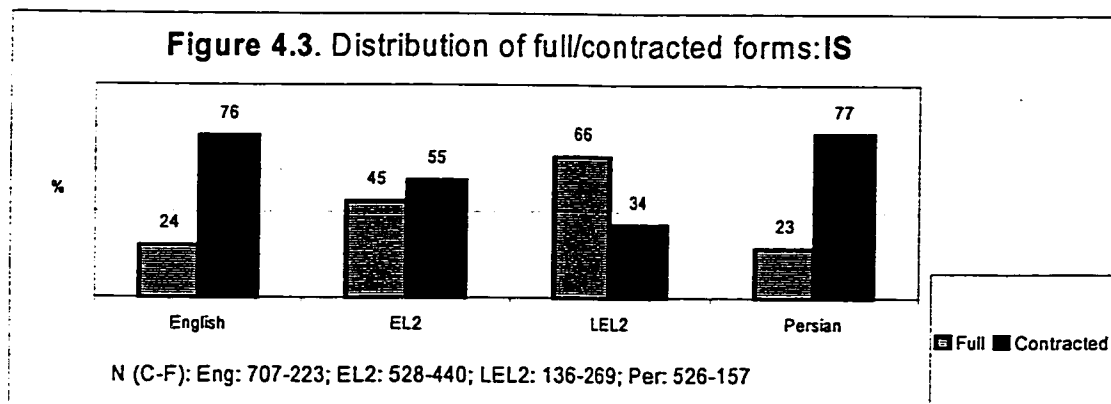
In sum, although the overall rate of contraction in LEL2 might suggest effects of native L1 at the early stages of acquisition (Figure 4.1), the effects of the following grammatical environment do not prove it to be so, specially for *gonna*, *V+ing* and adjectives (Table 4.2). Here, as we see, accounting for factors that condition the variable use of contraction (although only percentages are reported) is more revealing than the overall rate of occurrence of contracted *AM*. As mentioned, accounting for conditioning

is an important advantage of the comparative method over contrastive analysis that dealt with only rates of occurrence of linguistic features in its comparisons.

Preceding vowels categorically co-occur with full forms of the Persian equivalent to *AM*, while in English and the L2 this context strongly prefers contraction. It is only the following phonological factor of consonants (Figure 4.2) that indicates a slight tendency of LEL2 learners to follow their native language when contracting *AM*. But the effect of grammatical factors is of more importance here, since it is most relevant to the nature of the underlying grammar of the learners. Therefore, these results indicate that L2 learners all appear to follow the English patterns of contraction when using *AM* in their L2 speech.

IS

Unlike *AM*, overall distribution of data for *IS*-contraction does not show gradual increase from the L1 to the target language. Persian and English have the same rate in contracting *IS*, while L2 learners show an in-between behavior. Unlike Persian,



contraction is minimal in LEL2, even with 12% (49) of deletion added to contracted forms in Figure 4.3. This may be an indication of the lack of L1 influence on LEL2 in this context. According to the overall rate of contraction, Figure 4.3, L2 learners seem to have started from an L2 specific rate of *IS*-contraction (fewer contractions in LEL2), gradually moving toward the target or their native language systems (since the rates of contraction are identical in both Persian and English), as seen for EL2 learners with a higher rate of contraction. If I stop at this level of analysis (rates of contraction) I can

‘empirically’ argue for both L1 transfer effects on L2 and/or ‘gradual approximation’ (Nemser 1972) of L2 learners to the target language norms. Only discovering the *patterns* of variation through a variable rule analysis can help us choose between the two options. Table 4.4 presents the results of the variable rule analysis.

As seen in Table 4.4, *patterns* of contraction in EL2 and LEL2 are not very different from those of native English. In all three English contexts preceding subject and preceding phonological environment are selected as significant, with similar constraint hierarchies (note, however, that in some cases (LEL2) the ranges⁴³ are quite different). But in Persian, subject type was not selected (the direction of the effect, however, is the same). The difference (range) between preceding phonological factors in Persian is much bigger than that of other contexts; and the range for subject type (not selected), unlike English contexts, is the lowest in this language.

In other studies of contraction in standard English, preceding vowels have always favored contraction more than consonants (Labov 1972; McElhinny 1993; Meechan 1996), but why is it the opposite in my data in Table 4.4? Moreover, following grammatical category, selected as significant in other studies and in my native English context, was not selected for EL2/LEL2 data. In a closer examination of the data it was found that an overwhelming number of subjects in my data are tokens of *THAT/WHAT/IT* (e.g. 474 tokens in English), which are all consonant-final. This naturally caused an interaction between preceding phonological and grammatical (subject type) environments⁴⁴. Although I excluded frozen expressions from the data, there still remained enough tokens of *THAT/WHAT/IT* to possibly affect the findings. I, like Labov (1972) and Meechan (1996), had to exclude these tokens from my analyses and ran the program (for English contexts, because of the interaction problem and their almost categorical favorable tendency to contraction).

⁴³ Factor groups with higher ranges exert the most effect.

⁴⁴ Some interaction was present between preceding vowel and preceding personal pronouns (vowel-final), but higher frequency of data, higher number of pronominals other than personal pronouns that were collapsed together, and presence of tokens of vowel-final nouns considerably reduced the interaction effect. This interaction between preceding factor groups is one of the abiding problems of variationist studies of auxiliaries.

Table 4.4. *Contribution of factors selected as significant to the probability of contraction in English, EL2, LEL2, and Persian (THAT/IT/WHAT included): IS*

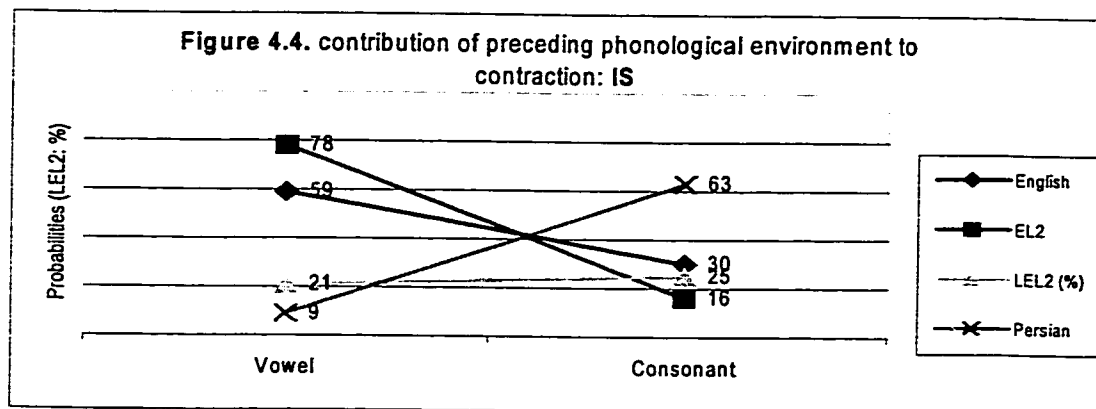
	English	EL2	LEL2	Persian
Total N	930	968	405	683
Log likelihood	-413.013	-485.484	-243.859	-258.095
Significance	.529	.078	.260	.130
Subject type	Prob. N/%	Prob. N/%	Prob. N/%	Prob. N/%
Pronoun	.61 648/85	.70 520/68	.58 96/41	[.55]* 245/79
Noun	.11 59/36	.04 8/4	.39 40/24	[.46] 281/76
Range	.50	.66	.19	[.09]
Preceding phonological environment				
Consonant	.58 497/81	.59 435/61	.58 106/41	.63 489/88
Vowel	.35 210/67	.27 93/36	.35 30/21	.09 37/29
Range	.23	.32	.23	.54
Following phonological environment				
Consonant	[.50] 506/76	[.50] 388/55	[.49] 106/33	.52 104/67
Vowel	[.50] 201/75	[.49] 140/54	[.54] 29/37	.32 8/62
Range				.20
Following grammatical category				
Gonna	.70 9/82	[.42] 5/45	[.48] 1/33	[NA]
V-ing	.65 51/76	[.62] 49/52	[.50] 5/28	[NA]
Locative	.61 27/69	[.39] 20/38	[.29] 2/13	[NA]
Adjective	.54 222/80	[.51] 179/58	[.56] 75/38	[NA]
Nominal	.46 384/75	[.50] 258/56	[.45] 53/30	[.56]
Participle	.25 14/56	[.32] 17/38	---	NA
Other	---	---	---	[.49]
Range	.45			
Preceding grammatical category**				
Adjective				.65 220/87
Nominal				.50 208/78
Pronoun				.44 51/77
Other				.20 47/48
Range				.44
Factor not selected:				
Subject type				X
Following gram. category		X	X	X
Following phono. environment X	X	X	X	
Intervening items X	X	X	X	NA

*In variable rule results, probabilities in square brackets, reported from the first stepping down run, were not selected as significant.

** not relevant to English contexts

00 Shadings show similarities

As shown in Figure 4.4 (also Table 4.5), this time preceding vowels favor contraction in both English and EL2. In LEL2, after excluding 113 tokens of *THAT/WHAT/IT* subjects, no factor was selected as significant (mainly because of the lack of enough data in some factors). I use percentages of contraction (not probabilities) in Figure 4.4, which show no important difference between the effects of vowels or consonants⁴⁵ in LEL2. Persian, on the other hand, shows opposite effects. Vowels almost



categorically disfavor and consonants dramatically favor contraction. This is an ideal site to demonstrate which pattern L2 speech mirrors because patterns of variation in the native and target languages contrast. These findings, in this respect, show that EL2 learners (and LEL2 learners after consonants) follow native English patterns of *IS*-contraction⁴⁶.

Following grammatical category was selected as significant in EL2 after exclusion of *THAT/WHAT/IT* tokens⁴⁷. As shown in Table 4.5 (and also Table 4.4),

⁴⁵ For Persian, the probabilities reported in Table 4.4 are used here.

⁴⁶ Some might argue that this is a specious finding. The reason that preceding consonants favor contraction in Persian is because the contracted forms start with a vowel. In English contraction removes the vowel. So the contrast merely shows that both languages prefer universally-favored CVCV structures. On the one hand the English rule (contraction after vowel) is not as categorical as non-contraction after vowels in Persian. The application of the universally-favored structure in English and Persian is therefore not completely identical. On the other hand, I interpret this pattern, and all the other patterns in all my analyses in this study, together with the patterns demonstrated for other contexts such as the effects of subject, following grammatical item, etc. This is also applied to contexts in which the number of tokens is not quantitatively justified or the numbers are reduced due to exclusion of certain factors.

⁴⁷ This might be due to the interaction between this factor group and *IT/WHAT/THAT* subjects, as well. These subjects are usually followed by nominal and adjectival items; their exclusion then reduced the number of nominal and adjectival items in the following grammatical factor groups in EL2 from 459 to 209 and from 306 to 138 respectively.

except for the effects of *gonna* in EL2 (maybe because of the scarcity of data), the constraint hierarchy is identical in both EL2 and English. This is of more interest here, because unlike phonological factors that may be subject to universal phonological processes (e.g., preference for CVC syllable structure), or other non-linguistic constraints (e.g., difficulty of producing some phonological segments e.g., production of English /r/ or /ə/ (as in *think*) by some L2 speakers), grammatical factors are related to the grammar of the respective languages. Despite dissimilarity of constraint hierarchies, the less proficient speakers prefer contraction before adjectives with almost identical probabilities as in English or EL2 (Table 4.4), while in the other contexts their behavior is different from English⁴⁸, but closer to EL2 (before *gonna* and locatives). At the same time they do not resemble the Persian system, where the following grammatical category does not contribute significantly to the contraction.

Table 4.5. *Contribution of factors selected as significant to the probability of contraction in English and EL2: (THAT/IT/WHAT excluded): IS*

	English		EL2	
Total N	456		448	
Subject type	Prob.	N/%	Prob.	N/%
Pronoun	.58	194/67	.68	88/35
Noun	.36	59/36	.28	8/4
Range	.22		.40	
Preceding phonological environment				
Vowel	.59	210/67	.78	93/36
Consonant	.30	43/30	.16	3/2
Range	.29		.62	
Following grammatical category				
V-ing	.73	33/70	.71	23/37
Gonna	.68	4/67	.49	1/17
Locative	.64	18/60	.62	7/22
Adjective	.49	50/47	.51	25/18
Nominal	.44	148/55	.41	40/19
Range	.29		.30	

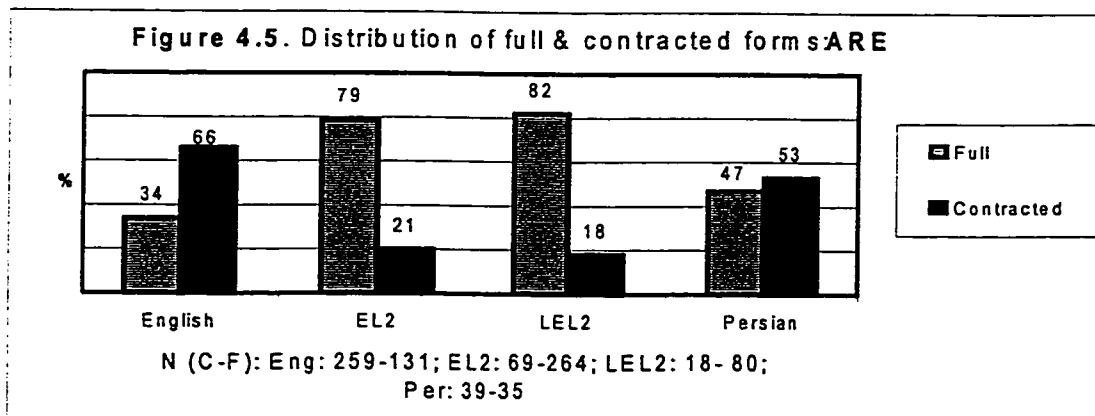
⁴⁸ This might be due to the scarcity of tokens in these contexts (which is the result of exclusion of some factors). As shown before adjectives and nominals, where there are enough instances of auxiliaries, their behavior is similar to EL2. Although small number of tokens in some contexts is problematic, possibility of performing the variable rule analysis and the results obtained show the reliability of the findings, at least in the contexts studied here.

Of other factor groups incorporated, Persian preceding grammatical items and following phonological environment contribute significantly to the contraction (Table 4.4), but in English contexts neither was following phonology selected as significant, nor are L2 constraint hierarchies similar to Persian in this factor group. Preceding grammar is not relevant to English contexts.

In sum, although overall rate of *IS*-contraction indicated differences between L2 learners' and native/target language speakers' behavior, the results of variable rule analyses of the conditioning of variability proved otherwise. Factors that contribute significantly to the contraction of *IS* in L2 (EL2 in particular) are similar to those that constrain *IS*-contraction in English, different from Persian. Discrepancy between low and high proficiency learners with respect to the constraint hierarchy in the following grammatical category might be an indication of change in the patterns of variability as acquisition proceeds.

ARE

Overall distribution of contracted/full forms of *ARE*, shown in figure 4.5, indicate that contraction is prevalent in English, though more evenly divided in Persian (with a slight preference for contraction). According to the traditional contrastive analysis claims, L2 acquisition is eased when native and target languages share a similar property. But, as shown in the figure, despite similarity between Persian and English in the rate of *ARE*-contraction, the majority of L2 learners, even the advanced ones, do not contract their *ARE* tokens. Have they created their own pattern of contraction in this context, different



from Persian or native English? Or do they differ only in the rate of contraction and not in its patterns?

Table 4.6. *Contribution of factors selected as significant to the probability of contraction in English, EL2, LEL2, and Persian: ARE*

	English		EL2		LEL2		Persian⁴⁹	
Total N	390		333		98		74	
Log likelihood	-173.918		-157.361		-41.120		-48.407	
Significance	.199		.138		.212		.188	
Subject type	Prob.	N/%	Prob.	N/%	Prob.	N/%	Prob.	N/%
Pronoun	.55	244/77	[.50]	63/22				24/51
Noun	.30	15/20	[.51]	6/13				15/56
<i>Range</i>		.25			.22			
Preceding phonological environment								
Vowel	.64	241/81	[.53]	61/23				7/32
Consonant	.13	18/20	[.38]	8/12				32/62
<i>Range</i>		.51			.16		.19	
Following grammatical category								
V-ing	.69	82/81	.59	23/26	[.48]	2/13		
Gonna	.61	19/79	.92	11/73	[NA]			
Locative	.50	24/62	.52	4/21	[.56]	1/20		
Adjective	.48	92/63	.41	17/15	[.67]	9/35		
Nominal	.28	42/52	.42	14/15	[.40]	5/10		
<i>Range</i>		.41	.51					
Following phonological environment								
Consonant	.54	201/69	[.49]	54/20	[.57]	15/21	[.64]	14/56
Vowel	.38	58/59	[.59]	15/24	[.33]	3/11	[.28]	2/40
Pause							[.45]	23/52
<i>Range</i>		.16						
Factors not selected:								
Subject type			X				X	
Preceding phono. environment			X					
Preceding gram. environment			—		—		X	
Following gram. category					X		—	
Following phono. environment			X		X		X	
Intervening items	X		X		X		X	

⁴⁹ In Persian, unlike English, the second person singular form of the verb is different from its plural forms. In this table and figure 5.3 only the plural forms are reported. When, in some analyses, the tokens of second person singular forms were added to other plural forms the results were almost identical.

The results of variable rule analyses are presented in Table 4.6. In native English the constraints on *ARE*-contraction are somewhat equivalent to those affecting *IS*. Pronominal subjects and preceding vowels favor contraction, as in Meechan (1996) and McElhinny (1993). Following grammatical category was selected as significant with the identical hierarchy as that of *IS* (Table 4.4). The following phonological environment was selected as significant, with consonants having a favorable effect.

EL2 and LEL2, which showed similar behavior when reporting surface frequency counts (Figure 4.5), have two distinct behaviors with respect to conditioning. EL2 shows similarities to native English in the preceding phonological environment (same hierarchy) and some parallels in the following grammatical category, although it does not totally parallel the constraint hierarchy. In the two contexts of the following *gonna* and nominals where EL2 constraint hierarchy is not parallel to English, higher probabilities in EL2 indicate application of the contraction rule in EL2, although differently. Moreover its behavior in these contexts is quite different from Persian where following grammatical category was not selected. LEL2 looks more like Persian; in both Persian and LEL2 preceding phonology exerts a significant effect on *ARE*-contraction with the same hierarchy. Preceding subjects, although not selected in Persian, show similar hierarchies, as well. In the following phonological environments, LEL2 has the same hierarchy as English (or Persian).

In sum, according to the selection of the factor groups, the operative effect on contraction of *ARE* in EL2 is following grammatical category, which is also selected in English. LEL2, on the other hand, is conditioned by preceding phonological environment (identical to Persian and its hierarchy) and preceding subjects (with the same hierarchy in Persian, different from English). Grammatically speaking, EL2 learners have acquired English patterns of variation. The similarity in the contribution of the phonological factors in English and EL2 (although phonological effects are more superficial) emphasizes the acquisition of these patterns. Less proficient learners, however, do not seem to have acquired the English system of *ARE*-contraction yet.

Summarizing the results in this section:

Overall surface distribution of contraction in English, EL2, LEL2, and Persian is misleading in the sense that it shows identical patterns for L2 learners, different from

either English or Persian, while analyses of conditioning indicate a number of differences among them. If I considered frequency rates only, the sole conclusion I could draw is that L2 learners of this study (proficient or not) have not acquired target language norms of contraction (no native-like acquisition), although they know how to use non-contracted *ARE* in grammatical sentences⁵⁰. The study of conditioning of contraction, however, clearly indicated that proficient learners behave totally differently from less proficient ones. They have acquired English patterns of contraction, although their rate of contraction is much lower than that of their English counterparts. But, LEL2 learners have yet to acquire patterns of contraction, although they produced so many grammatically correct sentences with *ARE*.

An issue that requires elaboration is the lower rate of *ARE*-contraction in L2 (particularly in EL2) contexts and the different behavior of EL2 learners in following phonological vowel and preceding nominal subject contexts. This difference may be attributed to the difference in the articulation of [r] in Persian and English. English [r] (like English diphthongs) involves both raising and retroflexion of the tongue tip (Lindau (1985), in Meechan (1996)). In Persian, on the other hand, [r] is a trilled phoneme articulated by the tip of the tongue in a flapped articulation against the alveolar ridge (Samareh 1985). This difference in the quality of the [r]⁵¹ in English and Persian has some effects (perhaps a preventive effect since it is difficult for Persian speakers to produce English [r]) on contraction of *ARE* before and after some phonological environments, the effects of which need further exploration.

Another explanation for the lower rate of *ARE*-contraction in EL2 is to assume that acquisition of this feature has already taken place and that learners will not change in so major a way their mastery of the structure, but rather their sociolinguistic knowledge as found by Regan (1996) for her L2 learners. However, LEL2 learners should acquire both the rule and its application.

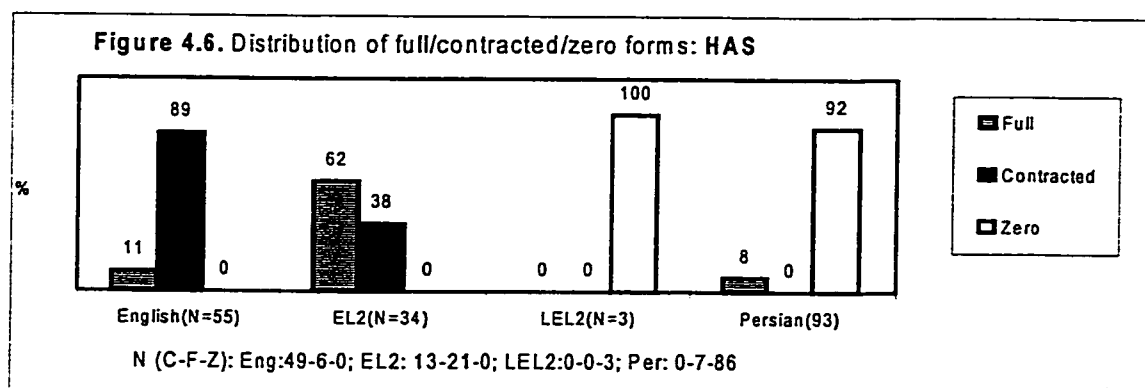
⁵⁰ Of all transcribed sentences in which *ARE* is used as an auxiliary/copula in EL2 or LEL2, only a few were used in ungrammatical sentences (mostly due to non-grammatical reasons (false starts, ...)).

⁵¹ This does not hold for production of *IS* and its contraction, so the rate of contraction is higher for *IS*.

HAVE/HAS/HAD

As was mentioned in section 4.1.2, the present perfect tense in Persian is formed from the past participle followed by the enclitic forms of the verb *astan* 'have' (see note 3). It is only in the 3rd person singular forms that variability exists between deletion and use of the full form of the auxiliary verb, *ast* 'has'. In this section, therefore, I will compare the behavior of Persian 3rd person singular present perfect forms to those of *HAS* (3rd person) in English and L2. In all other cases (*HAVE* & *HAD*), only English and L2 contexts will be taken into consideration.

Overall distribution of tokens of *HAS*⁵² in all contexts is presented in Figure 4.6.



While Persian and LEL2 almost categorically delete auxiliary verbs⁵³, there is not a single token of deleted *HAS* in English or EL2. There are then two distinct patterns of *HAS* usage in this context: deletion, which is favored by Persian and less proficient learners of English who, apparently, rely on their native language; and contraction, used by native English speakers and advanced learners of English. There is a difference between English and EL2 learners with respect to the frequency of contraction, i.e., EL2 learners contract *HAS* less than native speakers of English.

⁵² I will use *HAS* as a general word to refer to the auxiliary used in forming present perfect (3rd person) in all contexts.

⁵³ Mention was made that the only conditioning factors mentioned in the prescriptive literature for deletion of auxiliary verbs in present perfect forms in Persian were presence of a following equivalent form as in 38, where because of the presence of the auxiliary in the last sentence all the previous structures are used without an overt auxiliary. In our Persian corpus, 38 was the only instance of this condition. In all other cases (7 tokens) where two or more present perfect forms were used together, auxiliary verbs were all deleted regardless of the preceding or following perfective structures.

Because of the scarcity of the data and the large number of environments in which the contraction cannot occur, a more sophisticated analysis is not possible here (see also McElhinny (1993) for the same problem). Table 4.7 shows the effects of preceding grammatical and phonological environments in English and EL2 (Persian and LEL2 are not included because of their categorical behavior).

Table 4.7. *Distribution of contracted forms of HAS in English and EL2.*

	English		EL2	
	N	%	N	%
Preceding subject				
Pronouns	39/43	91	13/23	57
Nouns	10/12	83	Ø/11	Ø
Preceding phono.				
Vowels	26/28	93	3/13	23
Consonants	23/27	85	10/21	48
Total	49/55	89	13/34	38

Preceding pronouns favor contraction both in English and EL2, but with different rates, while preceding vowel co-occurs more with contracted verbs only in English. One reason for the different behavior of EL2 data here might be the effect of interaction between subject type and preceding phonological environment on contraction. Therefore,

Table 4.8. *Distribution of contracted forms of HAS in English and EL2; Preceding subject and phonology combined.*

Preceding subject/phonology	English		EL2	
	N	%	N	%
Consonant-final nouns	5/6	83	Ø/7	Ø
Vowel-final nouns	5/6	83	Ø/4	Ø
Consonant-final pronouns	18/21	86	10/14	71
Vowel final pronouns	21/22	95	3/9	33
Total	49/55	89	13/34	38

I combined the preceding subject type with the preceding phonological environment to factor out the interacting effect. Table 4.8 presents the distribution of the data after

collapsing preceding factor groups. As seen, EL2 shows similarity to English in contracting forms of *HAS* after pronominals, consonant-final pronouns in particular, while EL2 nouns behave differently with no contraction at all.

In sum, the findings reported with respect to *HAS* indicate that: a) EL2 learners do not follow their L1 system, since there is no instance of *HAS*-deletion in their data. On the other hand, LEL2 learners follow their L1 in this context by deleting all instances of *HAS* (as mentioned in footnote 46, I interpret my findings in a certain context together with the patterns demonstrated for other contexts, this does not exclude other interpretations if one context (e.g., *HAS* deletion in LEL2) is considered individually); b) EL2 learners contract *HAS* more after pronouns, although the rate of contraction is quite different from that of English; and c) the lower rate of contraction and distribution of the data in Tables 4.7 & 4.8 indicate that EL2 learners may have partially acquired patterns of *HAS*-contraction in English, but LEL2 learners rely more on their L1.

The finding for *HAS* is further confirmed when examining the distribution of the contracted forms of *HAVE* in the two contexts (Table 4.9). Here, because of the interaction effect as for *HAS*, I have collapsed the preceding subject and phonological factors together. Contracted forms of *HAVE* are correlated with preceding pronouns both in English and EL2 (only vowel-final pronouns, though). Less proficient learners have also used 12 tokens of *HAVE* in their speech of which four tokens are contracted after pronouns. Interestingly, unlike *HAS*, they did not delete any tokens of *HAVE* auxiliaries, most likely because in their L1, equivalents of *HAVE* are never deleted.

Table 4.9. *Distribution of contracted forms of HAVE in English, EL2 and LEL2, Preceding subject and phonology combined.*

Preceding subject/phonology	English		EL2		LEL2	
	N	%	N	%	N	%
Consonant-final nouns	6/12	50	Ø	Ø	Ø	Ø
Vowel-final nouns	Ø	Ø	Ø	Ø	Ø	Ø
Consonant-final pronouns	29/42	69	Ø	Ø	Ø	Ø
Vowel-final pronouns	127/149	85	19/59	32	4/7	57
Total	162/204	79	19/73	26	4/12	33

There are no tokens of *HAD* used by LEL2 learners and tokens of *HAD* are categorically non-contracted in EL2. In the native English context, the contracted form of *HAD* occurs 48% (13/27) of the time which is conditioned only by preceding vowels (100%) and preceding pronouns. Summarizing the findings on *HAS/HAVE/HAD*, EL2 informants seem to have acquired (partially) patterns of contraction from English, although they need to increase their sociolinguistic knowledge (see Reagan 1996) in this respect. LEL2 learners, however, rely on (most likely) their native language rules in these contexts.

*WILL/WOULD*⁵⁴

I submitted 88 tokens of usage of auxiliary *WILL* in native English to variable rule analysis, and only preceding grammatical and phonological factors were selected as significant. Since variable rule analysis was not possible in the EL2 and LEL2 contexts, I will compare only the marginal results here, as shown in Table 4.10. All contexts favor contraction after vowel-final pronouns (preceding factors are combined as for *HAS*, for the same reason). Frequency of contraction by L2 learners is much lower than that of native speakers. Of five tokens coded as contracted in LEL2, three tokens are actually deleted; this makes their behavior different from the others.

Table 4.10. *Distribution of contracted forms of WILL in English, EL2 and LEL2, Preceding subject and phonology combined.*

Preceding subject/phonology	English		EL2		LEL2	
	N	%	N	%	N	%
Consonant-final nouns	1/9	11	Ø	Ø	Ø	Ø
Vowel-final nouns	2/6	33	Ø	Ø	Ø	Ø
Consonant-final pronouns	2/4	50	Ø	Ø	Ø	Ø
Vowel-final pronouns	61/69	88	10/62	16	5/34	15
Total	66/88	75	10/89	11	5/45	11

⁵⁴ Mention was made in chapter three that Persian modal auxiliaries, equivalents of *WILL/WOULD*, are totally different from those of English; and they are thus not dealt with here.

Table 4.11 presents the findings for *WOULD*. As shown, second language learners almost categorically do not contract tokens of *WOULD* in their natural speech, although it might be argued that their rate of contraction decreased because the same thing happens in native English, where there is only 45% contraction for *WOULD* while the rate of contraction for *WILL* was 75%. Native-like use of auxiliaries involves not only their grammaticality but also their patterns of variable contraction, otherwise, as observed by Meechan (1996), an L2 learner will always sound like a ‘foreign’ no matter how grammatical his/her auxiliary usage is in the L2. The EL2 informants, here, show no sign of native-like acquisition.

Table 4.11. *Distribution of contracted forms of WOULD in English, EL2 and LEL2, Preceding subject and phonology combined.*

Preceding subject/phonology	English		EL2		LEL2	
	N	%	N	%	N	%
Consonant-final nouns	3/31	10	Ø	Ø	Ø	Ø
Vowel-final nouns	2/11	18	Ø	Ø	Ø	Ø
Consonant-final pronouns	6/24	25	Ø	Ø	Ø	Ø
Vowel-final pronouns	105/189	56	2/128	2	2/10	20
Total	116/256	45	2/178	1	2/12	17

The results reported for *WILL/WOULD* are then not different from those of *HAVE*: contraction occurs only after vowel-final pronouns in L2. But, as was mentioned in a few places, phonological factors may be subject to factors not relevant to the grammatical repertoire of the informants. One possible reason for contraction of forms of *HAVE/WILL/WOULD* after vowel-final pronouns may be that the informants prefer universally-favored phonological structures (e.g., CVCV); therefore the behavior of L2 learners does not indicate that they follow their English counterparts, but they use a natural phonological process. Although this possibility holds, the different behavior of vowel-final nouns (i.e., no contraction after vowels, although only a few tokens in each context, even in English) and the difference in the behavior of EL2 and LEL2 learners shows that proficient learners may have partially acquired patterns of contraction from

English (in some contexts only) but, the less proficient group still relies on their L1 patterns of usage.

4.3. Discussion

Whatever the frequency of contraction, results (marginal, variable rule) found for advanced second language learners in the previous section (excluding *WOULD/HAD*) indicate that these learners have acquired the system of variability that operates in their target language. Of course, as advanced L2 speakers, they were expected to have acquired *syntactic* knowledge of auxiliaries in well-formed sentences. But, regarding the acquisition of patterns of variability, surface frequency rates observed in their speech (that is in most cases lower than that of native English speakers) suggested otherwise. For instance, in the case of *ARE*, rates of contraction do not indicate that EL2 learners know English rules of contracting. However, studying patterns of variation revealed that L2 learners had acquired the variable rules of the target language (contrary to Gregg 1989).

For *AM* and *IS*, factor weights (percentages and/or probabilities) and constraint hierarchies showed a rather complete match between English and EL2, both phonologically and grammatically (Tables 4.2-4.5; Figures 4.1-4.4). In the case of *ARE*, too, the constraint hierarchies were very similar (Table 4.6). Factors conditioning the variable contraction of forms of *HAS* (i.e., preceding subject and phonological environment) have a similar effect both in English and EL2; although the rate of contraction in EL2 is quite lower than that of English. However, in the case of *HAS*, unlike Persian, EL2 learners did not delete auxiliaries at all. For instances of *WILL*, we observed rather the same behavior as that of *HAVE*. Despite frequent use of *WOULD* in EL2, the rate of contraction in this context is very low. Closer examination of the data of individual informants revealed that nearly half of all instances of *WOULD* (83/178 tokens) were produced by only one speaker.

The most revealing of all contexts are those where both native and target languages demonstrate variation but with different patterns of variability. We witnessed such contrasting patterns for the effect of phonological environment before *IS* and *ARE*, subject type on *ARE*-contraction, following grammatical category, and contrast between deletion and contraction in the case of *HAS*. In these contexts, EL2 learners were following patterns operating in native English. In sum, patterns of variation found in EL2,

in most cases, are parallel to the patterns that operate in native English. There is no evidence that they are created by L2 learners or transferred from their native language.

Less proficient learners, on the other hand, demonstrate a different system. When contracting *AM* and *IS*, their patterns are closer to English. For instance, the effects of following grammatical category on *AM*-contraction is not that different from English or EL2 (except for the lower effect of locatives). Contribution of factors selected as significant for the probability of *IS*-contraction, and their constraint hierarchy for LEL2 mirrors those of EL2 (or English)(Table 4.4), as well. Influence of Persian, however, is slightly evident in some contexts such as preceding phonological environment, after tokens of *THAT/WHAT/IT* subjects were excluded from the data (e.g., vowels in Figure 4.4). However, contraction of *ARE* and constraints on it draw a distinct line between higher and lower proficient learners. In this context, LEL2 patterns reflect the factor weights and constraint hierarchies of Persian (Table 4.6). This native language-like pattern is even more evident in the contexts where *HAS* is used as auxiliary. LEL2 speakers categorically delete instances of *HAS* (although only a few of them), following their Persian counterparts. Less proficient learners, in sum, rely more on their native language in some contexts (e.g., *HAS*), and approximate the patterns of target language variability in some others (e.g., *IS*). This, on the one hand, indicates that even less proficient learners acquire patterns of variability from the target language (as for *AM*, *IS*) and, on the other hand, shows that each auxiliary verb behaves differently in L2 development. Some auxiliaries and their patterns of variability are acquired earlier, regardless of their lower frequency in the input (target language), as for *AM* with only 8% frequency, while others are acquired later, no matter how frequent they are, as in the case of *ARE* with 18% frequency in English⁵⁵.

What light do these findings shed on the questions I asked at the beginning of this chapter? The first question was whether SLA is exclusively seen as acquisition of categorical rules. The results of this study indicate that advanced second language learners (and less proficient informants in the cases of *AM* and *IS*) have acquired not only

⁵⁵ Explaining the reasons for the difference in the behavior of forms of auxiliary verbs is not my concern here, but syntactic complexity of contexts where different forms of auxiliaries are used and learners' 'needs' to use some forms of auxiliaries in their everyday conversations might be helpful speculations in this respect.

the grammatical features of the target language, but also the patterns of their variable use, contrary to Gregg (1989) who asserts that variable patterns cannot be acquired at all. It was also shown that frequency of occurrence of a target language item in L2 (either low (*ARE*) or high (*WOULD*)) is not evidence that an L2 learner has acquired socio/linguistic knowledge with respect to that certain item. Native-like second language acquisition involves acquisition of both target language syntactic and variable patterns; otherwise an L2 learner would sound like a foreigner regardless of his/her frequent (grammatical) production of the target language structures. An example may make this more transparent.

One of my advanced L2 speakers used 83 instances of *WOULD* (47% of all instances of *WOULD* in the EL2 corpus) in his speech. Frequent and grammatical use of *WOULD* by this informant apparently indicates that he has acquired this feature of English. However, native English speakers contract *WOULD* 45% of the time, while contraction for this informant is only 3%. The same informant contracts *IS* 52%, *HAVE* 75%, and *HAS* 67% of the time, almost the same as an average native speaker. This informant used *WILL* only once in his interview, and he has contracted this one token. Can we say that this certain speaker has gained linguistic knowledge of the use of *WOULD*, but has yet to acquire *WILL* in his L2 speech? As far as the English proficiency of this person is concerned (according to his interview), he will correctly judge/answer any sentence/question involving use of *WILL* and *WOULD* in English, i.e., if he is given a grammaticality judgment test, the results will indicate his mastery of English syntax in this respect. But as the results of this study indicate this does not mean that he will demonstrate native-like behavior when actually using this knowledge in natural situations. In other words, he might have acquired a categorical rule for *WOULD*, but native English utilizes a variable rule that needs to be acquired by this informant, as he did demonstrate in his use of auxiliaries such as *AM*, *ARE*, or *IS*.

L2 patterns fluctuate between rules applied in the native and target languages due to lack of complete acquisition of variable patterns of the target language, as is the case for the less proficient learners when contracting *ARE*. In some contexts they follow English patterns of variation (where patterns have been acquired), e.g., effects of the following phonological environment where their constraint hierarchy is more similar to

that of English speakers than Persian (Table 4.6); and in some other cases they mirror L1 patterns of variation (where the pattern has not been acquired), such as the effects of preceding phonological factors on *ARE*-contraction.

How is it possible to determine if the constraints on variability have been acquired? The comparative method used in this study and the results obtained from it demonstrate that variationist means and methods are not only applicable to the SLA situation but can also quantitatively describe patterns of variability that operate in L2 speech. We saw that variable items and their conditioning factors are acquired by the advanced L2 learners; and that we may call native-like acquisition.

Table 4.12. *Factors conditioning variable contraction/deletion of auxiliaries in English, EL2, LEL2 and Persian*

Contexts	Language			
	English	EL2	LEL2	Persian
<i>AM</i>	Eng.=EL2=LEL2=Per.			
<i>IS</i> Eng.=EL2=LEL2 Per.	<u>Subject type</u> <u>Preceding phono.</u> <u>Following gram.</u>	<u>Subject type</u> <u>Preceding phono.</u> <u>Following gram.</u>	<u>Subject type</u> <u>Preceding phono.</u> Not selected	Preceding phono. Following phono. Preceding gram.
<i>ARE</i> Eng.=EL2=LEL2=Per. (same constraint hierarchy in LEL2 & per.)	Subject type Preceding phono. <u>Following gram.</u> Following phono.	Not selected Not selected <u>Following gram.</u> Not selected	Subject type <u>Preceding phono.</u> Not selected Not selected	Not selected <u>Preceding phono.</u> Not selected Not selected
<i>HAS</i>	Eng.=EL2=LEL2=Per.			
<i>HAVE/HAD</i>	Eng. (more or less) =EL2=LEL2=Per.			
<i>WILL</i>	Eng. (more or less) =EL2=LEL2			
<i>WOULD</i>	Eng.=EL2=LEL2			

Finally, what are the sources of variation in second language speech? A simple answer is either the native or the target languages. Findings of this study evidenced that variation in advanced second language speech is, in most contexts, conditioned by similar factors, factor weights, and constraint hierarchies as those of the target language, summarized in Table 4.12 (in this table only the effect of factors selected by the variable

rule analyses are shown). This means that of the general hypotheses with respect to the sources of variation in L2 (chapter 1), hypothesis two is confirmed, i.e., the target language is the source of variation in EL2. Sources of variation in the speech of less proficient speakers, on the other hand, are either L1 (e.g., *ARE*) or the target (*AM*, *IS*). The English system of variation is utilized when L2 learners have sufficient knowledge of the patterns of variation of the target language (hypothesis two), but in cases where they are still far from the target language system, their patterns mirror those of the native language (hypothesis three). Similar to the processes of acquisition that involve acquisition of variable patterns, use of native language patterns in L2 is coupled with the transfer of variable patterns of L1, as well. This means that transfer of native language features need not be limited to transfer of categorical rules, but can be the influence of the patterns of variability in the L1 on the second language. Although my results revealed systematicity in all aspects of L2 speech studied here (hypothesis one), we did not witness any L2-specific patterns of variation in any context (hypothesis four).

Some other general patterns that emerged from the results of the analyses in the previous sections are interesting enough to be discussed. First, based on the findings reported in Table 4.1, a hierarchy of the frequency of occurrence of each auxiliary (contracted or not) across all contexts looks like the following:

English	EL2	LEL2	Persian
is (44%)	is (52%)	is (65%)	is (48%)
are (18%)	are (18%)	are (16%)	are (5%)
have (13%)	would (10%)	will (8%)	have (5%)
would (12%)	am (8%)	am (7%)	am (1%)
am (8%)	have (6%)	have (2%)	
will (4%)	will (5%)	would (2%)	

Regardless of the similar hierarchies in different contexts, according to those who consider a major role for the frequency of items in L1 or the target language in their accounts of language transfer (e.g., Andersen 1983; Zobl 1980, 1983; or what Adamson & Regan (1991) call “frequency of input hypothesis”), we expect a more target-like pattern of variation in L2 for more frequent items of the target language, such as *IS* or *ARE*. But this is not always the case. For instance, *WOULD* is more frequent than *AM* in

both English and EL2, while it is one of the least variable auxiliaries of second language speech (Table 4.11); and at the same time there is no matching item in the native language to affect its use in L2. *AM* on the other hand has only 8% occurrence in English, but patterns of variation for *AM* in EL2 are in no way different from those of English (see figures 4.1, 4.2 & Tables 4.2, 4.3). The frequencies of contraction (not only the frequency of use, contracted or not) show similar effects as well. Both *WILL* and *IS* are contracted 75% of the time in English, while EL2 learners contract *IS* 54% and their contraction of *WILL* is only 11% (Table 4.1).

If frequency of L1 items had an influence on language transfer, *IS* as the most frequent auxiliary of Persian (even more frequent than English) should affect patterns of variation in second language speech. This prediction is not born out either. According to the findings of Table 4.4, 4.5 and Figure 4.3, L2 learners, advanced speakers in particular, follow the same patterns of variation used by native English speakers.

The second pattern (or rather lack of pattern) is the relation between the overall distribution (frequency counts) of data and patterns of variability (regardless of frequency) found through variable rule analysis. Mention was made in chapter one that one of the main problems of the traditional contrastive analysis was that researchers following that tradition used surface language structures or, at most, their frequency of occurrences as the basis for their comparisons among languages. This did not lead to a thorough understanding of the phenomenon of cross-linguistic influence (Odlin 1989). Findings of the present study indicate that surface frequency counts are not revealing of the effects of native or target language features on second language acquisition. A clear example is the case of variable use of *ARE* (full/contracted) in my data. Overall distribution of data, shown in Figure 4.5, indicates an identical behavior for both advanced and less proficient learners of English that is different from both Persian and native English. The variable rule results, on the contrary, demonstrate that patterns of variation for advanced learners are completely different from those of less proficient learners (Table 4.6). EL2 learners follow the same patterns of variation as those of native English, while LEL2 learners rely more on their native language patterns.

Dependence on solely the prescriptive or grammatical accounts of linguistic features of languages involved in contact/SLA situation is also misleading. Mention

was made, for instance, that one of the conditioning factors mentioned in the prescriptive literature for deletion of auxiliary verbs in present perfect forms of Persian is presence of a following equivalent form (as in 38)(Nobahari 1993). However, in my Persian data, of 93 instances of *HAS*, only seven are followed by another equivalent form, of which only one is full (while according to Persian grammar all should have been used overtly). There is no other account for deletion of almost 95% of instances of the auxiliary verbs in present perfect forms of the language.

A third pattern that emerges from the results is a sort of sequence of acquisition of patterns of variability in auxiliary usage (as in 56). These results indicate that acquisition of patterns of variation is maximal for *AM* and *IS* in second language speech (EL2 and LEL2), i.e., L2 learners use native-like constraints on *IS/AM* contraction. *ARE*⁵⁶ is a turning point for less proficient speakers. Here they are separated from advanced learners, who continue to follow target language patterns, by relying more on their native Persian system of contraction. When it comes to the use of *HAS*, there are two clearly distinct behaviors for the two groups of learners: EL2 learners contract, like native English speakers, and LEL2 learners delete, the same as Persian speakers. This pattern continues through other auxiliaries for EL2 learners, while for LEL2 learners there are not enough tokens of contracted/deleted forms of other auxiliaries to extract a pattern.

56. *AM* $\Rightarrow$ *IS* $\Rightarrow$ *ARE* $\Rightarrow$ *HAVE* (has, have, had) $\Rightarrow$ *WILL* $\Rightarrow$ *WOULD*

A question that might be raised here is the high frequency of tokens of *WOULD* in EL2 data. If *WOULD* is the last auxiliary whose pattern of variation is acquired, why are there then so many tokens of it in the data? To answer this question, I mentioned above that 83 instances of *WOULD* were used by only one individual. This is not that surprising since, as asserted by Gass & Selinker (1994), each individual L2 learner might create his/her own language system.

A final consideration is to see if the patterns of variation in L2 speech change as acquisition proceeds. It was discussed before that less proficient learners rely more on their native language system of variation before they acquire patterns that operate in the

target language. After acquisition of the patterns of variability of the target language, they are applied by L2 learners in their speech, regardless of the frequency of applications. My less proficient informants, for instance, followed Persian patterns of variation when contracting tokens of *ARE*, while proficient informants used English patterns in the same situations. Patterns of variability, therefore, change as acquisition proceeds (from one level to a higher level of proficiency). But after acquisition of the target language patterns, it is the overall rate of contraction (rule application) that changes. The effect of the native language on L2 is also maximal at the early stages of acquisition; this effect is weakened when patterns of variation are acquired from the target language.

4.4. Summary of chapter four

In this chapter, I compared the patterns of variability in the use of auxiliary verbs in second language speech at two levels of proficiency to those of Persian, and English, in order to find out what patterns L2 learners follow. The findings indicate that less proficient learners rely more on their native language patterns, while advanced learners follow the system of variability that they have acquired from the target language. Furthermore, these results revealed no L2-specific patterns, different from the native and target languages, operating in the contexts I examined in this chapter.

⁵⁶ See my discussion on *ARE* for its lower rate of contraction in L2.

Chapter five: RELATIVE CLAUSES: Relative Pronouns

5.0. Introduction

English relative clauses are variably marked by three relative marker options: a *WH*-word, *THAT*, and *ZERO*, as in 57-59 (from Guy & Bayley, 1995):

- 57. This is the house *which* I told you about.
- 58. This is the house *that* I told you about.
- 59. This is the house $\emptyset$ I told you about.

A substantial amount of research has already been devoted to the study of this syntactic variation in English. The most common questions asked in this respect are: how do native speakers of English choose a relative pronoun, and what factors, linguistic or extra-linguistic, constrain the variable use of these options. A number of suggestions have been presented as the answers to these questions and a number of factors have been singled out as contributing to the choice of relative markers.

How do second language learners choose a relative pronoun? Is use of relative pronouns in L2 speech variable; if it is, is it systematic? Are the constraints on relative pronoun choice acquired from the target language, transferred from the native language or created by L2 speakers themselves? These are the kinds of questions that this part of the present study attempts to answer.

In order to answer these questions, I will perform a quantitative variationist analysis on conversational EL2 data from Persian speakers; then, similar analyses will be done on data from native Persian and native English. Comparison of the results of my three-way analyses will hopefully shed light on the sources of variation in EL2. This method will yield evidence as to whether the EL2 system follows a system of variable relative marking that operates in native English, relies on Persian, or has its own specific system of relative marking. My L2 corpus in this chapter is limited to data from advanced L2 learners of English since relative clause constructions are not frequently produced by less proficient learners.

I first present a summary of the literature on the grammar and use of relative markers in English, Persian and EL2. Then I establish my coding procedures and method

of analysis. The results of the quantitative analyses and a discussion of the results comprise the final sections of this chapter.

5.1. Relative markers⁵⁷ in English, Persian, and EL2

Relative pronouns, in post-nominal languages, are used to introduce a post-modifying clause within a noun phrase (e.g., Persian, English). The post-modifying clauses are usually modifiers of the head noun (or antecedent) of the whole NP (Keenan 1985). The relative pronoun is coreferential with the head noun, and may have different syntactic functions in the relative clause. The whole NP including the antecedent, relative marker and the relative clause (relative complex (Tottie 1995)) can syntactically function as the subject, object, etc. of the matrix sentence.

5.1.1. Relative markers in English

There have been many studies concerned with the alternate use of relative markers in English⁵⁸. Quirk (1957) is based on data from educated spoken British English and its results show that non-personal antecedents are followed by *ZERO* relatives more and that *ZERO* relatives mainly function as the object of relative clauses that are mostly adjacent to their head nouns. Quirk et al (1985) indicate that subject relative clauses are marked with *WHO*, *WHICH*, and *THAT*, while *ZERO* relatives are used only in informal speech with indefinite pronouns. They also assert that objective *THAT* and *ZERO* are preferred with non-personal antecedents, whereas *WHO(M)* is mostly used with personal ones. Proximity of the relative clause and its head noun, and the degree of complexity of the subject of the relative clause influence selection of the relative pronoun, too. Huddleston (1971) found definiteness of the antecedent and Taglicht (1973) found personal pronoun subjects to be important factors for the selection of *ZERO*. Olofsson (1981) found definiteness and modification of the antecedent NP and subjects as influential factors in relative pronoun usage.

⁵⁷ Relative marker, relativizer, or relative pronoun all are used interchangeably in the literature. Relative markers/pronouns will be used in this thesis.

⁵⁸ Recently there has been an increasing amount of interest in the study of variation in relative clause construction in different varieties of English, African American English in particular (Harvie 1998, Tottie & Rey 1997, for instance). I will benefit from their findings and methodology, in particular, in this study.

Biesenbach-Lucas (1987) and Kikai, Schleppegrell, and Tagliamonte (1987) were among the first to apply variable rule analysis to the study of relative marking in English. The former found that *WHO* is used exclusively with human antecedents in restrictive (99%) and non-restrictive relative clauses. *WHICH* is used only with non-human head nouns and it is almost always confined to non-restrictive relative clauses. *THAT* and *ZERO* mark only restrictive clauses and accompany mostly non-human head nouns. Kikai et al. examine the effects of syntactic position on relative pronoun selection. Their study is based on data from speech and writing in modern American English. They pay special attention to the interaction of syntactic position with other factors of humanness and definiteness. From a methodological point of view, they are the first to deal with the interaction effect in relative clause studies.

Adamson's (1992) main objective is the investigation of *ZERO* marking and the role of social class on the selection of *ZERO* relatives. His results indicate that two factors, the position of the gap within the relative clause and the socioeconomic class, are significant factors in the choice of *ZERO* as relative marker. Tottie (1995) found subject of the relative clause and the function of the relative clause in the matrix sentence as significant factors conditioning choice of *ZERO* in present-day written standard British and American English. And finally, Guy and Bayley (1995), dealing with spoken as well as written data, report that animacy of the antecedent, the channel of communication (written, spoken), the syntactic position of the relativized element in the relative clause, and the adjacency of the antecedent and the relativized element all have significant effects on the choice of relative marker.

In sum, being fairly frequent in language, being subject to strong constraints, and having several clearly distinct variants, relative marking has been an interesting site for investigation in English. Most analyses of variation have focused on animacy and definiteness of the antecedent, adjacency of the antecedent and the relative clause, and syntactic function of the relative marker in the relative clause as the most important factors conditioning relative marker choice in English. The debate, however, still continues concerning the specific role of each factor and the interaction among several significant factors. My study is an attempt to examine the influence of the factors conditioning relative marker usage in English on the acquisition of these markers by L2

learners. I categorize various factors as syntactic, semantic and discourse, and try to demonstrate their effect on L2 variation.

5.1.2. Relative markers in Persian

Persian relative clauses are mostly introduced by a dummy element *ke* ‘that’. *ke* is a general complementizer; it never functions as a noun or pronoun, it does not mark gender and number and does not have WH features; witness 60.

60. *kesi ke bacheh da:re ba:yad tu in ha:vs-a: beshineh.* (Per. PD. Ib. 34)

someone that kid has must in this houses sit

Someone who has kids should live in these houses.

There is no general consensus among Persian grammarians on the function and use of this relative marker. Hashemipour (1983) and Sadighi (1982) both believe that *ke* is obligatory in Persian relative clause constructions. Tarallo & Myhill (1983), however, assert that the Persian relativizer may be deleted under some conditions, but Persian grammarians have never been able to present a satisfactory account of this condition. Windfuhr (1979), emphasizing the obligatoriness of *ke* in most types of relative clauses, lists some exceptional cases in which *ke* is optional, as in the following:

- in generalised relative clauses introduced by *har* ‘each, every’ (Windfuhr, P. 67)
e.g., *har kesi (ke) sarba:z bude...* whoever was a soldier once...
- temporal clauses introduced by (*har*) + *vaqti* ‘every time/whenever’
e.g., (*har*) *vaqti (ke) residi...* when(ever) you arrive...
- clauses introduced by conjunctive phrases like *az bas (ke)* ‘so’ (literally ‘enough’)⁵⁹
e.g., *as bas (ke) xaste shod...* he became so tired that...

He relates the optionality of *ke* in these cases to the semantic properties of the head nouns, which he believes are [-definite] [-specific]. Then, he asserts, this indicates that *ke*

⁵⁹ This construction is rarely used in natural Persian speech and no tokens of this kind were found in our Persian corpus.

is not dummy but implies a semantic differentiation (definiteness/specificity) yet to be determined.

Lazard (1992) defines three functions for Persian relative clauses:

a) The relative clause makes the head noun definite (restrictive relative clauses), as in 61⁶⁰.

61. *mardi ke a:mad bara:dar-e man ast.*

man that came brother-EZ my is The man who came is my brother.

b) The relative clause has, in relation to the head noun, the value of an apposition (non-restrictive relative clauses), (62).

62. *diruz mahdi, ke pesare xubi ast, ba man bud.*

yesterday Mahdi that boy good is with me was

Yesterday, Mahdi, who is a nice boy, was with me.

c) The relative clause adds necessary information to the meaning of the sentence, without making the head definite (an intermediary type (treated as restrictive relative clauses by Lazard and many others in the literature)); see 63.

63. a. *do ta: ma:shino ke be ham xordeh budand didam.*

two cars that to each other collide were saw

I saw two cars which had collided.

b. *do ta: ma:shino didam (ke) be ham xordeh budand.*

He adds that a relative of the last type may be separated from its head noun by the verb, as in 63.b and that, when the relative clause follows the verb, the particle *ke* is often omitted in colloquial language. Another site for the optional use of Persian relative marker, according to Lazard, are indefinite relative clauses that are all introduced by *har* 'each, every'. Lazard's indefinite relative clauses are equivalent to Windfuhr's generalized and temporal clauses, examples of which were cited above from Windfuhr.

⁶⁰ Examples 61-63 are from Lazard (1992).

In sum, Persian restrictive relative clauses, according to the grammatical accounts of the language, are all marked by *ke*, except for the following contexts where the relative clauses are optionally marked with the relative marker:

- Indefinite relative clauses introduced by *har* 'each, every'.
- Relative clauses separated from their head nouns by an intervening verb.

Does the Persian system of relative marking have any effect on the behavior of Persians learning English relative clauses? For instance, since the behavior of Persian *ke* is so similar to English *THAT*, does it facilitate learning patterns of *THAT*-usage by L2 learners? A comparison of the systems of relative marking in Persian and EL2 will shed light on these and other relevant questions.

5.1.3. Relative markers in EL2

Relative pronoun marking in the speech of second language learners had not attracted much attention until recent years. The reason for this lack of interest may be the dominance of the research on the availability of the Accessibility Hierarchy (see chapter 6) in second language acquisition, particularly order of the frequency of relative clause types and presence of resumptive pronouns. In 1979, Gass incorporated a rather small section on relative markers in her study on the acquisition of relative clauses by learners from nine different linguistic backgrounds. In order to determine the existence of language transfer in this area, she tested two groups of learners: one group was comprised of speakers with an invariant relative marker in their native language (Arab, Persian⁶¹, Thai, Chinese, and Korean) and another group with a variable relative marker. She noted no significant differences with regard to errors in the EL2 speech of these learners, suggesting that language transfer was not a relevant factor in the acquisition of English relative markers by L2 learners. Myhill (1982) examines the results of grammaticality judgement tests of Spanish and Chinese EL2 learners and concludes that learners of a second language pick and use the constructions which are most easily accessible. Among these accessible options are the native language constructions that learners mistakenly equate with ones in the target language. If the L2 constructions, he adds, are so opaque or foreign that there is nothing in the native language that is even close to them, no transfer

⁶¹ Contra Gass, it was shown above that Persian relative marking is not invariant.

effects will be produced. He uses this assumption to conclude that English relative pronoun deletion⁶² is not even remotely connected in its domain of application to anything in Chinese or Spanish or any other language (including Persian) examined in Ioup & Kruse (1977). Then, according to Myhill, this is why neither Ioup & Kruse nor Myhill himself found any interference from their informants' native languages in their EL2 speech with respect to the use of *ZERO* relatives.

Tushyeh (1985) has looked into relative pronoun marking, among other things, by non-native speakers of English. She finds relative pronoun omission to be the second most frequent error (as she calls it; see ft. 62) in her data of 102 Arabic learners. She claims that all these errors are due to the transfer of Arabic relative pronoun deletion in contexts where the antecedent noun is indefinite. Yoroza (1995) focuses, for the first time most probably, only on relative pronoun usage and explores the differences and similarities in relative pronoun usage by native and non-native speakers of English. He found that non-natives employ more relative markers than native speakers and that beginning level non-native speakers used more relative pronouns than those in the advanced group. He claims that the frequency of relative pronoun use characterizes the current developmental stage of the learner's interlanguage system.

There are only two studies (that I am aware of) of relative pronoun marking in L2 speech that have applied variable rule analysis. In a study of 73 relative clauses (66 restrictive and 7 non-restrictive), Karstadt (1996) found that Swedish-Americans use *THAT/dat* (phonetic substitution for *THAT*) to mark both subjective/objective and animate/inanimate antecedents. She relates this to transfer from the speakers' first language. In the second study that concentrates on the use of objective relative pronouns, Flanigan and Inal (1996) collected data from native and non-native speakers of American English via a written preference task (what forms the speakers prefer in writing and in speech) and a sentence combining task. They divided their non-native speakers into two groups with length of residency (in the US) of less or more than two years. Their findings

⁶² Unfortunately there is no distinction between deletion (first insertion and then deletion) and *ZERO* in the literature. Following other studies on relative marking in English I use *ZERO* as a relative marker in English restrictive relative clauses. Restrictive relative clauses without an overt relative marker are therefore not dealt with as errors. Tushyeh (1985)(and others) do not distinguish between *ZERO*-relatives and relatives that are most likely ungrammatical due to the lack of a relative marker (e.g., markers after a preposition).

indicate that non-native speakers use more relative pronouns than native speakers, due to the influence of prescriptive grammar teaching in English classes abroad. They also found that exposure to native speaking environments leads to a gradual adoption of native speaker norms. Their findings evidenced variable use of relative pronouns both by native and non-native speakers. Native speakers favor use of *ZERO* in oral speech and *THAT* in written discourse, but non-natives preferred *WH*-forms in both written and oral forms. Native speakers preferred *ZERO* with both human and non-human referents, while non-natives with more than two years of residency favored *WH*-forms with human referents and those with less residency preferred *WH*-pronouns with both human and non-human referents. Function of the relative marker as direct object or object of preposition had significant effects, as well. Native speakers preferred *ZERO* with all object types while non-natives with less residency favored *WH*-forms in this context. Non-native speakers with more than two years of residency used *WH*-forms with only human direct objects or object of prepositions.

In sum, systematicity of the variable use of relative pronouns in second language speech and sources of this variation have not been dealt with in enough detail in any of these studies (Even Flanigan & Inal's study is limited to only objective relatives, and their data are collected in unnatural situations). Nor is there a clear idea of the effects of L1 on the variable use of relative markers in L2. Moreover, except for Flanigan and Inal (1996) no other study has dealt with the influence of the target language patterns of variation on the acquisition of relative clause constructions by L2 learners. The present study attempts to tackle these issues in the following sections.

5.2. Analyses

5.2.1. Variable context, Exclusions and Coding

When the relative clause places a restriction on the referent of the head noun and modifies its content, it is called a restrictive relative clause as in 64⁶³, where the relative clause confines the potential reference of the 'things'. In non-restrictive relative clauses⁶⁴

⁶³ For ease of understanding only examples from our English corpus will be used here. Whenever there is any difference between Persian and English contexts, Persian examples will be provided as well.

(65), on the other hand, the relative clause is inessential to the linguistic identity of the noun, it only gives more information about the head noun already identified previously.

64. There is certain *things* Ø you don't do. (OSLA 241.Ia. 266)

65. It was on *Percy*, which is almost downtown. (OSLA 246.Ia. 286)

The alternate use of relative markers in English, EL2 and Persian restrictive relative clauses is the main concern of my study in this chapter. The proportion of non-restrictive relative clauses is low in my data, as in all vernaculars (Tottie et al. 1997). Moreover, in English non-restrictive relative clauses, variability in relative marking does not exist since they strongly favor *WH*-pronouns (Ball 1996; Guy & Bayley 1995; Kikai et al. 1987); they were therefore excluded from my analyses. Among the other constructions not considered here are the instances of *relative marker* + *BE* deletions, as in 66. These constructions are not normally dealt with as relative clauses and they lack variability in relative marking. Headless relative clauses, such as 67, were excluded because there is no variability in this context, either.

66. We all went to university, we all graduated, just like a lot of you Ø [*who are*] listening to this. (OSLA 246.Ib. 101)

67. The chip wagon was Ø *where* I met my girl friend. (OSLA 247.Ib. 125)

Ambiguous constructions, such as 68 and 69 where the relative clause does not restrict the referent of the head noun and the clause is ambiguous between a relative clause and a subordinate clause, were all excluded from the analyses. In 68, for instance, the utterance could be read as "I know that I can call someone", where the *that clause* is a complement of the matrix verb; or as "I know enough people that I can call (someone)"

68. I know enough people *that I can call someone*, ... (OSLA 268. Ia. 175)

⁶⁴ The distinction between restrictive and non-restrictive relative clauses is sometimes problematic if it is based only on semantic factors. In oral discourse, access to the audiotapes of the speech is essential. Quirk (1957: 101) points out that non-restrictive relative clauses are characterized by "a fresh intonation contour, and a change in the degree of loudness". Like Tottie and Rey (1997) I relied on both semantic and phonological criteria to distinguish between the two types. Persian relative clauses, among other criteria, share the same semantic/phonological features as English.

69. That brings me to another thing *that I don't really believe in God*. (OSLA 285.IIa.31)

Adverbial relative clauses are included only in the presentation of overall distributions; they were all excluded from the variable rule analyses because there is not only less variability in the choice of relative marker in these constructions (e.g., Persian adverbial relative clauses categorically favor *ke*, and English adverbs of cause and manner are mostly marked with *ZERO*), but also because the antecedents are most often restricted to certain words like *time* for temporal adverbials or *the reason* for adverbs of cause, etc. Adverbial relatives, also, interacted with other factor groups like humanness and subject of the relative clause in my data; and the only way to remove interaction was to exclude them.

Finally, as is customary, I excluded all tokens whose analysis is unclear (restrictive/ non-restrictive; relative/non-relative, etc.).

5.2.2. Factor groups

I extracted all tokens of restrictive relative clauses from my English, Persian and EL2 corpora; then, every one of them was coded for several factor groups that are assumed to affect the choice of relative markers. For the English and EL2 data of this study, I coded the materials for the factors whose effects have been found to be most revealing in previous research and for which I had enough data to base my quantitative analyses on. For the Persian corpus, in addition to the English factors, factors whose influence on relative marker use has been mentioned in the grammatical literature of the language were used. English factors were used for the Persian data because of the close similarity of the relative clause constructions in both contexts: relative clauses are headed, finite and post-nominal in both languages; they both enjoy variability in relative marking, etc. My inventory of factor groups is listed in Table 5.1. After coding, separate analyses were performed for each of the variants (*WH*-forms, *THAT*, *ZERO*, and *ke* (for Persian)) using Goldvarb 2.0, a variable rule application for the Macintosh (Rand & Sankoff 1990).

Table 5.1. Factors and factor groups incorporated in the study of relative marking.

		Factor Group	Factors	Previously studied in
<i>Semantic factors</i>	1	Definiteness of the head noun	Definite, indefinite, pronominal	Huddleston 1971, Olofsson 1981, Kikai et al. 1987, Fox & Thompson 1990, Adamson 1992, Tottie 1995, Tottie & Rey 1997,
	2	Humanness of the head noun	Human, non-human	Quirk 1957, Kikai et al. 1987, Beisenbach-Luca 1987, Fox & Thompson 1990, Adamson 1992, Guy & Bayley, 1995, Ball 1996, Tottie & Rey 1997
<i>Syntactic factors</i>	3	Premodification of the head noun	Premodified, not-premodified, non-premodifiable	Olofsson 1981, Tottie & Rey 1997
	4	Syntactic function of the NP head in the matrix clause	Subject, direct obj., indirect obj./object of prep., existential subj., subj. complement, other	Fox & Thompson 1990, Adamson 1992, Guy & Bayley 1995, Tottie 1995, Tottie & Rey 1997
	5	Syntactic function of the relative marker in the relative clause	Subject, direct object, object of preposition/indirect object, other	Quirk 1957, Kikai et al. 1987, Fox & Thompson 1990, Adamson 1992, Guy & Bayley 1995, Ball 1996, Tottie & Rey 1997,
	6	Category membership of the subject of the relative clause	noun, personal pronoun, relative pronoun	Quirk 1957, Taglicht 1973, Olofsson 1981, Tottie 1995, Tottie & Rey 1997
<i>Formal factors</i>	7	Adjacency of the head noun and the relative marker	Adjacent, non-adjacent	Quirk 1957, Olofsson 1981, Guy & Bayley 1995, Tottie & Rey 1997
	8	Length of the relative clause	Six words or less, seven words or more	Quirk 1957, Quirk 1968 (in Ball 1996)
	9	Position of the relative clause	Sentence-final, sentence-medial	Quirk 1957,

5.2.2.1. Semantic factor groups

Definiteness of the head noun

Restrictive relative clauses restrict the reference of the head noun. A major factor often mentioned as conditioning the choice of relative marker is the degree of definiteness of the head noun (Kikai et al. 1987). Tottie (1995) and Tottie & Rey (1997) have found that pronominal heads favor selection of *ZERO*, while Huddleston (1971) and Olofsson (1981) indicated the significance of definite heads in selection of *ZERO*. In

Kikai et al. (1987), too, the tendency is toward a greater rate of choice of *ZERO* as the degree of definiteness for direct and indirect object positions increases. Adamson (1992), on the other hand, did not find definiteness to be an important factor in the selection of *ZERO* marking.

The factors in this factor group include definite nouns, as in 70, indefinite nouns, as in 71, and pronominals, as in 72.

70. It was *the first death* that I'd ever encountered. (OSLA 278.IIa. 240)

71. If it's *a decision* that her and her husband made that, it is not a problem. (OSLA 285.Ia. 426)

72. There's *nothing* ØI really want to become. (OSLA 241.Ia. 42)

Humanness of the head noun

Many analyses have found that the humanness of the head noun exerts a major effect on the choice of relative marker. Their results show that whether the antecedent is human is clearly the principal constraint affecting the choice between *WHO* (human) and *WHICH/THAT* (non-human) (Ball 1996; Guy & Bayley 1995; Kikai et al. 1987; Quirk 1957, to name a few). Flanigan and Inal's (1996) results, on the other hand, indicate that native speakers of English mark objective relative clauses with human or non-human head nouns more often with the *ZERO* relative marker, whereas non-native speakers prefer *WH*-forms for human head nouns in the same contexts. In order to examine the effects of humanness of the head noun on the data, all antecedent head nouns were coded as either human, as in 73 or non-human, as in 74.

73. I'm not like one of these *people* that'll have a couple 'f drinks a day. (OSLA 278.IIa. 378)

74. There's an Italian *program* that's on Sundays. (OSLA 285.Ia. 48)

Humanness of the antecedent plays an important role in English relative pronoun usage, while in the Persian grammatical literature there is no mention of this factor with respect to relative marking. On the other hand, Persian grammarians assert that all NPs

containing relative clauses are obligatorily⁶⁵ marked by an affixal item ‘-i’ (demonstrative -i (Windfuhr 1979)) that marks the nouns for definiteness. If it is found that humanness exerts a significant effect on relative marking in EL2 in this study but definiteness plays a trivial role, the results will confirm acquisition of target language system by EL2 learners. If it is found otherwise, transfer from L1 will be argued to play a role in EL2 patterns of relative marking.

5.2.2.2. Syntactic factor groups

Syntactic factors are more relevant to the underlying grammar of the speakers. Therefore, if the results indicate that syntactic factors significantly affect relative marker choice in EL2 with similar rates and conditionings as English, and if these simultaneously differ from Persian, I will argue that patterns of relative marking are part of EL2 grammar and that these patterns are acquired from English. If EL2 patterns parallel the Persian system of relative marking, transfer will be argued to be the most important factor in this respect.

Premodification of the head noun

Although not selected as a significant factor group, Tottie & Rey (1997) found that premodification of the head noun asserts some effects on the selection of relative markers. They found that premodified NPs have a tendency toward *ZERO* relatives and not-premodified NPs favor *THAT* more than premodified or not-premodifiable head Ns. I included, following Tottie & Rey, premodified, not-premodified, and not-premodifiable head nouns as the three factors of this factor group, exemplified in 75, 76 and 77 respectively. Premodified nouns are those antecedents containing items such as indefinite pronouns (e.g., all, any, every), modifiers (e.g., superlative, the only), etc. Not-premodifiable nouns include personal pronouns and other words such as *anybody* and *everything*.

75. It's *the first time* Ø I've actually done a romantic comedy on the stage. (278.Ib. 159)

76. We've had *people* who've come to the border and had to arrest them. (285.Ia. 300)

⁶⁵ In my data NPs are not obligatorily marked with this item.

77. *Everything* that they make is for their children. (285.Ib. 307)

Syntactic function of the NP head in the matrix clause

Among those who studied the effects of the syntactic function of the head noun in the matrix clause, Guy and Bayley (1995) and Adamson (1992) did not find this factor group to contribute significantly to the selection of either *ZERO* or other relative markers. Only Tottie & Rey (1997) evidenced a significant effect for this factor group in their Ex-Slave data. They found that subject complements and subjects in existential sentences strongly favor choice of *ZERO* as the relative marker in Early African American English. The factors in this group, in this study, include Subject, as in 78, Direct Object (79), Object of Preposition⁶⁶ (80), Subject Complement (81), Existential Subject, as in 82, and Other syntactic functions (including adverbs and no syntactic functions, where it is not possible to tell what the syntactic function of the head is, as in 83).

78. Usually *the rules* Ø they set up are pretty basic. (OSLA 278.Ia. 379)

79. I just met *one guy* that I really recognized. (OSLA 136.IIa. 362)

80. I may have dreamed *about things* that've had already happened. (OSLA 136.IIa. 101)

81. It's *one of the most modern cities* Ø I've been in. (OSLA 133.Ia. 339)

82. There's *lots of stuff*, like, here on records that I try and play. (OSLA 239.IIb. 73)

83. *Your sister* that you can almost tell anything to. (OSLA 241.Ib. 369)

⁶⁶ I treated direct objects and object of prepositions as one factor. Persian indirect objects are almost always governed by prepositional items, and English and EL2 indirect objects were collapsed with object of prepositions because of their related syntactic functions (as objects), their analogous behavior in English relative clauses (Gass, 1979), and infrequent usage of indirect objects. Keenan & Comrie (1977) notice that indirect objects do not have a clear typological status in terms of relative clause formation, i.e., they tend to behave as either direct objects or objects of preposition. Pavesi (1986) also asserts that from a more general point of view, indirect objects are widely realized as adverbs of movement. English is a typical example of a language in which indirect objects are not distinguishable from other oblique objects; therefore they are treated as simple objects of preposition even by L2 learners of English (Pavesi 1986).

Syntactic function of the relative marker in the relative clause

This factor group concerns the grammatical role of the relative marker in the relative clause. The relative marker may function as the subject of the relative clause, as in 84, direct object (85), or object of preposition, as in 86⁶⁷.

- 84. I hate people *who* play the game. (OSLA 241.Ib. 253)
- 85. It was basically the biggest, cheapest house $\emptyset$ we could find. (OSLA 268.Ia. 5)
- 86. You've lost a whole part of something $\emptyset$ you've worked so hard for. (OSLA 278.Ib. 207)

Almost all studies on relative clauses in English have considered the syntactic function of the relative marker in the relative clause an important determinant of various grammatical features (e.g., Harvie 1998; Kikai et al. 1987), the choice of the relative marker, in particular. Quirk (1957, also Quirk et al. 1985) found that in standard English subjective relative clauses do not favor a *ZERO* marker at all. This finding has been supported by later studies, as in Guy & Bayley (1995). In Tottie & Rey's (1997) study, nevertheless, of 57 subjective relative clauses, 23 (40%) were marked by *ZERO*. Direct objects and objects of preposition have also been found to favor *ZERO* relatives, and disfavor *WH*- or *THAT* forms in these studies.

Category membership of the subject of the relative clause

The subject of the relative clause in English is either a nominal item (definite/indefinite/proper name), a pronominal (mostly personal pronouns), or the relative pronoun when the relative marker syntactically functions as the subject of the relative clause. In the latter case there exists strong interaction between this factor group and the syntactic function of the relative marker in the relative clause. Tottie (1995) excludes subjective relative clauses from her study and finds subject of the relative clause as a significant factor contributing to the choice of relative markers, with personal pronouns favoring *ZERO* while definite NPs favor *THAT* or *WH*-forms. In Tottie & Rey (1997), where subjective relatives were also included in the study, this factor group was not selected as significant. However, a tendency of *ZERO* to occur with personal

pronouns has been found in the earlier studies in British (Quirk 1957; Taglicht 1973) and American English (Olofsson 1981).

The grammatical category of the subject of the relative clause was broken down into three factors: relative pronoun, as in 87, other pronouns, as in 88, and nouns, as in 89.

87. He's doing good stuff for all the crazies *that* are in there. (OSLA 239.Ib. 405)
 88. She actually worked for the same printer $\emptyset$ *I* worked for at one time. (OSLA 239.Ib. 321)
 89. I have a model which *one of my friends* is building right now. (OSLA 241.IIa. 88)

Persian, as a pro-drop language, utilizes zero subjects in relative clauses; zero as the subject of the relative clause was added as the fourth factor in this factor group (for Persian), exemplified in 90.

90. *Un chiza:i ke Ø tu ira:n xundeh budam, qabl az umadan, hameash gera:mer bud* (Per. FB.Ib.136)

Those things that in Iran study be before from come all
 grammar was

The stuff that I had studied in Iran, before coming, were all grammar.

5.2.2.3. Formal factor groups

Adjacency of the head noun and the relative marker

In Guy & Bayley (1995) and Tottie & Rey (1997) adjacency has been found to significantly influence the choice of relative markers, *ZERO* in particular. They found that non-adjacent relative clauses were usually marked by an overt relativizer rather than *ZERO*. Quirk (1957) had also found adjacency of the head noun and the relative clause as an important factor in the choice of *ZERO* relatives in standard (British) English. According to Lazard (1992), in Persian, on the contrary, non-adjacent relative clauses are the only contexts in which relative markers may be deleted. I incorporated the two factors of adjacent (91) and non-adjacent (92) relative clauses to examine their effect on the

⁶⁷ Other grammatical function of the relative marker (adverbials, etc.) were all excluded from the variable rule analyses.

choice of relative markers. EL2 learners will mark non-adjacent relative clauses more with an overt relative marker if they follow the English system of relative marking. But, if their non-adjacent relative clauses are marked with *ZERO* more, they are transferring the L1 patterns of marking into their L2 (If Lazard's assumption is empirically confirmed).

91. The hoist is *the thing that lifts the car up*. (OSLA 247.Ia. 294)

92. *The front end* of the car on the driver side **which had hit the ground first** was twisted. (OSLA 247.Ia. 408)

Length and position of the relative clause

Quirk (1957) identified a number of other properties of relative clauses which may affect the selection of the relative marker, including length of the relative clause and its position in the matrix clause. Quirk (1968, cited in Ball 1996) suggested that the length of a relative clause influences the choice of *WHICH* versus *THAT* for non-personal subjects when the clause is sentence-final. These factors were either overlooked in the other studies of relative clauses, particularly in American English, or were never selected as significant where studied (e.g. Olofsson 1981).

To further investigate any possible effect of these factors on the data, relative clauses with six⁶⁸ words or less were coded as short and seven words or more were coded as long (see 93 & 94). Position factor group included two factors of sentence-medial and sentence-final relative clauses, exemplified in 95 and 96 respectively.

93. Most industrialized countries *ØI've been in* have a value added tax system. (OSLA 133.Ia. 276)

94. There's lots of little *things that , you know, you kind of say well, you know, you don't really have to do it that way*. (OSLA 247.IIa. 41)

95. Everything *that they make* is for their children. (OSLA 285.Ib. 307)

96. There's a lot of shows on it *that I find children shouldn't be watching*. (OSLA 285.Ib. 20)

⁶⁸ Considering the length of an average English sentence and claims in psychology with respect to the capacity of short-term memory to memorize 6-7 words or numbers at a time.

5.2.3. Data

All tokens of restrictive relative clauses were extracted and coded for the factors described in the previous section. After excluding tokens that were not suitable for variable rule analysis, three sets of corpora, English, Persian, and EL2, were submitted to Goldvarb 2.0.

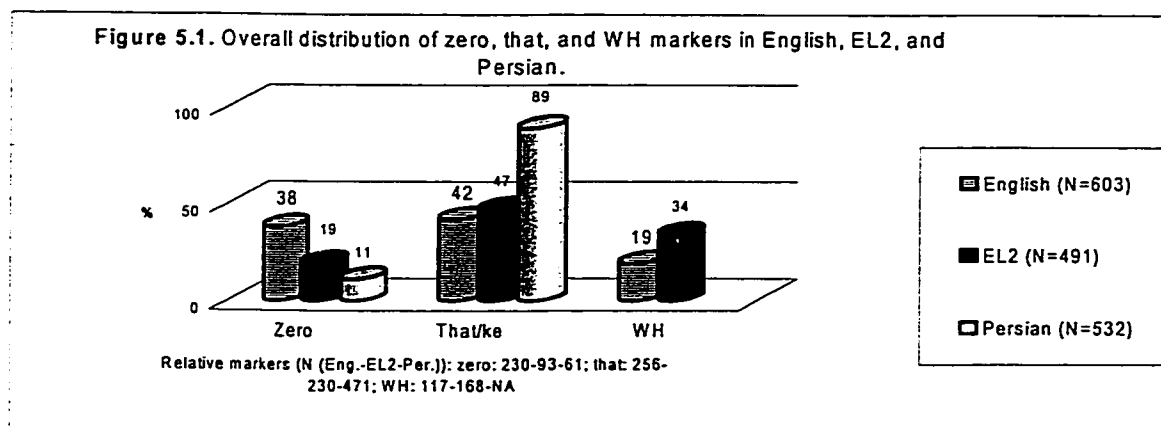
5.3. Results

Figure 5.1 presents the overall distribution of variants in the three contexts. As shown in the figure, Persian relative clauses are marked with *ke* 'that' more, but unlike what has been mentioned in some grammatical accounts (e.g., Hashmipour 1989), and like what was reported in others (Lazard 1992), *ZERO* relatives are also used by Persian speakers. Actually, taking the grammatical accounts into consideration, I did not expect such a high preference for *ZERO* in Persian (N=61, 11%). There are only eight tokens of indefinite relative clauses introduced with particle *har* 'every' (the only potential site for *ZERO* relatives in Persian, asserted by Lazard and Windfuhr) in the data, of which only four tokens lack the relative marker *ke*, as in 97.

97. *har chi Ø da:shtam niga: kardan.* (Per. MP.Ia. 258)

every thing had-I saw did They saw whatever I had.

Later it will be shown what factors condition the choice of zero relative in Persian.



The first observation, based on Figure 5.1 is that EL2 learners do not behave differently from native English speakers regarding their tendency to use *THAT*-relatives

(47% & 42%). With respect to the use of *ZERO*, their behavior looks more like Persian than English, where *ZERO* is used 38%. But the frequency of the use of *ZERO* in other studies like Kikai et al. and Guy & Bayley's spoken data is not that different from that of EL2 (16% and 26% vs. 19% here). Therefore, the assumption that the transfer of Persian lowers the preference for *ZERO* in the EL2 spoken by Persian speakers is still premature. The results of *WH*-marking in EL2 are, again, different from that of the English data (34% vs. 19% for English), and almost identical to that of Guy & Bayley and Kikai et al. (35% and 34% vs. 34% here). Despite this very close similarity, since informants of this study were exposed to mainly Canadian English spoken in the Ottawa area, the difference in their behavior from that of native English speakers needs more explanation.

When the data on *WH*-forms are broken down into different forms of *WH*-pronouns, as in Table 5.2, we observe that EL2 speakers' use of *WHERE*, *WHEN*, and *WHAT* is not different from that of English. The only differences are in the use of *WHO* and *WHICH*, with the largest difference for *WHICH* (10%). There might be two

Table 5.2. *Use of WH- relative pronouns in English and L2:
(Persian not applicable)*

Relative Marker	English		EL2	
	N	%	N	%
who	63	10	88	18
which	17	3	64	13
where	22	4	9	2
when	10	2	4	1
what	5	1	3	1
Total	117/603	19	168/491	34

explanations for the EL2 speakers' preference for use of *WHO* and *WHICH*. In Guy & Bayley's study *ZERO* relatives are preferred more in the spoken data⁶⁹. Looking at the percentage of *WH*-forms in their written data (and also that of Kikai et al.), we observe a rather sharp increase for preference for *WH*-forms, 58% *WH* vs. 26% for *ZERO*. It may

⁶⁹ I do not have data on the frequency of the use of individual *WH*-forms in Guy & Bayley or Kikai et al.

be concluded that in written discourse (in formal speech, in other words), the preference for relative marking use shifts from *ZERO* to *WH*-forms. Therefore, one might conclude, EL2 speakers' behaviour in this study indicates not a difference in behaviour but a difference in the style of the speech, with EL2 data being more formal than spoken English. Regarding my explanations on the method of collecting (via conversations or sociolinguistic interviews) and handling the data (chapter three) this explanation seems untenable.

The second explanation involves some elaboration on the transfer phenomenon. We cannot assume any effect of Persian on EL2 with respect to the use of *WH*-forms since there are no such forms in Persian. We cannot, either, suggest that the lower tendency of Persian speakers to use *ZERO* has affected EL2, hence the lower preference for *ZERO* in EL2 has caused a shift from *ZERO* to *WH*. Any effect of Persian on EL2 (according to contrastive analysis studies, for instance) should cause a shift from *ZERO* to *THAT*, because *THAT* is the most prevalent item in both Persian (89%) and English (42%). However, as shown in Figure 5.1, use of *THAT* in EL2 (47%) is almost identical to its use in English and much different from Persian.

Earlier, I mentioned that, based on the claims made by Flanigan et al (1996) (also Yoroza (1995)), transfer of the teaching materials (transfer of training (Selinker 1972)) used in English classes abroad may have significant effects on the EL2 of second language learners. To explain the different behaviour of the EL2 speakers from that of English in *WH*-usage, I had to have a look at the materials used in the general English classes at Iranian universities. My EL2 informants have all graduated from Iranian universities and had to pass two general English courses (General English I and II). The textbook⁷⁰ used in these classes is the same all over the country, and is specified by a national organising committee (known as SAMT). Following are the grammatical notes from the textbook, which are followed by patterns to be practiced in the form of exercises, with more emphasis on standard rules (i.e., use of *WHO*, and *WHICH*).

⁷⁰ "English Sentence Structure" by Robert Krohn, first published in 1970, is the only textbook used in these classes. This book relies mostly on the old structural approach to teaching languages and tries to automatize the structural patterns by giving extensive amounts of pattern practices and exercises. First the patterns are introduced in the form of notes, then practices follow. See appendix B, for a copy of the notes and a sample of exercises used to teach relative clauses.

(Krohn 1998: p. 180 (Appendix B))

- *who* and *whom* are used for persons
- *which* is used for things
- *that* is used for persons, things, and animals

The EL2 speakers' use of *WH*-forms, *WHICH* in particular (64/13% use for things), mirrors the prescriptive rules taught in the English classes⁷¹.

In sum, according to the findings in Figure 5.1, the overall distribution of the data in EL2 looks more similar to English, particularly where *THAT* and *WH*-forms are involved. More frequent use of *WH*-forms in EL2 was argued to be due to the transfer of training materials used in English classes in Iran.

5.3.1. Marginals

As was mentioned before (Chapter three), knowledge of the overall distribution of the variants in different contexts gives us enough information to realise only the surface behaviour of each variant. Variable rule analysis, on the other hand, helps us determine which factors contribute a significant effect to the choice process when a number of them are considered simultaneously. In order to show the relation between surface distribution of the data and the actual patterns of variation and also, to report the effects of some factors that are not selected by variable rule analysis, but might be revealing in some of the contexts, I will present the results in two sections: a short section on marginal results (of only a few contexts), and another section that includes the results of the variable rule analyses.

5.3.1.1. Definiteness of the head noun

The degree of definiteness of the antecedent appears to affect the choice of the *ZERO*-relative markers in English (Kikai et al.1987; Olofsson 1981). Tottie (1995), also, shows that pronouns prefer *ZERO* relatives more than definite or indefinite nouns. The

⁷¹ It might be interesting to explore the difference in *WH*- usage in the speech of speakers with different amounts of exposure to native English. Most of the informants of this study (only EL2 speakers are included here) had almost the same length of residency in Canada, however, two of the speakers whose proficiency in English was a little lower than the others used more *WHICH*. Further research needs to be done to clarify this issue.

findings in Table 5.3 indicate that English pronouns and definite nouns co-occur more with *ZERO* relatives (57% and 50%), thereby confirming previous studies. In the Persian, context definite nouns almost never co-occur with *ZERO*, only 1%, contra English; while pronominal heads favor *ZERO* 23% of the time.

Table 5.3. Relative markers and definiteness of the antecedent.

Relative Marker	Definite						Indefinite						Pronoun					
	English		L2		Persian		English		L2		Persian		English		L2		Persian	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
<i>Zero</i>	125	50	53	20	1	1	55	24	26	13	44	13	50	57	14	41	16	23
<i>That</i>	94	37	138	53	127	99	130	49	82	41	289	87	32	36	10	29	55	77
<i>Wh</i>	33	13	68	26			78	30	90	45			6	7	10	29		
Total	252	42	259	53	128	24	263	44	198	40	333	63	88	15	34	7	71	13

EL2 behaves more like native English when the antecedents are indefinite or pronominal. There is a preference for *WH*- and *THAT* with indefinite NP heads, with 45% and 41% respectively; and the proportion of *ZERO* with pronominal heads is more than the other contexts. When the head noun is definite, EL2 speakers choose *THAT* more than any other marker (53%), getting closer to Persian, where the only choice is *THAT* (*ke*). *ZERO* relative usage with only 20% of occurrence is much less than that of English (50%); and interestingly enough *WH*-forms are selected less in this context compared to indefinite or pronominal heads. Why is this so? Do the EL2 speakers transfer the Persian system to EL2 when the NP head is definite? This is a point where surface similar/different rates of frequencies lead to unreliable results. We will see later that, despite similarities between Persian and EL2 in definiteness context, their patterns of variation are completely different from each other. Another point where surface rate of occurrence is not a true representation of actual patterns of variation is the apparent prevalence of *WH*-forms in EL2 that makes this context look different from English. This surface difference, again, does not reveal the actual system of variable *WH*-form usage in EL2, a system that mirrors nothing but the native English system of variation.

Table 5.4. Relative markers and adjacency.

Relative Marker	Adjacent						Not-adjacent					
	English		L2		Persian		English		L2		Persian	
	N	%	N	%	N	%	N	%	N	%	N	%
<i>Zero</i>	221	40	86	19	47	12	9	17	7	17	14	11
<i>That</i>	230	42	217	48	356	88	26	51	13	31	115	89
<i>Wh</i>	101	18	146	33			16	32	22	52		
Total	552	92	449	91	403	76	51	8	42	9	129	24

5.3.1.2. Adjacency of the head noun and the relative marker

Table 5.4 presents the overall distribution of variants in the adjacency factor group. Since this factor group was not selected as significant in any context, I report only the surface distributional patterns. As shown, in Persian relative clauses the proportion of relative marking is exactly identical in adjacent and non-adjacent contexts, 88/89% for *ke* 'that' respectively. Lazard's (1992) assertion for the possibility of deletion of the relative marker when the relative clause follows the verb (non-adjacent) is thus confirmed by the results here. At the same time, there is a 12% frequency of relative marker deletion in adjacent relative clauses, something that has not been mentioned (or accounted for) in the grammatical literature of the language.

The findings for the English data indicate preference for *ZERO* relatives in adjacent (40%) vs. non-adjacent contexts (18%). This is on par with what has been found by Quirk (1957) for British English and by Guy & Bayley (1995) for American English. Also on par with the findings of Guy & Bayley, English speakers are more likely to use a *WH*-form (32% vs. 18% (adjacent)) or *THAT* (51% vs. 42% (adjacent)) in non-adjacent than in adjacent contexts. For the non-adjacent factor, EL2 learners' use of *ZERO* is lowest, as in English (17%⁷² and 18% respectively), and they prefer more *WH*-relatives for non-adjacent relative clauses than adjacent ones, just like English. Unlike English or Persian, their use of *THAT* is reduced to only 31% for the non-adjacent factor versus 48% for adjacent relative clauses. Why are *WH*-forms more frequent in this (non-adjacent) context? Of 22 *WH*-forms used with nonadjacent relative clauses in EL2, 10 are instances

of *WHICH* and 3 instances of *WHERE*, while in the English data there are only 4 tokens of *WHICH* and no tokens with *WHERE*. Here, the strong effect of the instructional material is clearly seen in the EL2 of Persian speakers; they act according to the rules (emphasis on overt markers) they have been taught. Why then, do more *WH*-forms (not *THAT*-relatives) replace *ZERO* relatives in this context? We saw that *ZERO* is not favored in non-adjacent relative clauses, so *THAT* or *WH* should be favored in this context. In selecting between these two options (in addition to the effects of instructional material), the distance between head and modifying clause causes EL2 speakers to make the meaning of their utterances more transparent by inserting a relative pronoun that specifies more of the features of the NP head. They, therefore, prefer a *WH*-form with person and gender features.

Where the relative clause is adjacent to its antecedent, EL2 speakers' use of *THAT* is not very much different from English, the highest in both contexts. EL2 prefers more *WH*-forms than *ZERO* relatives for the reasons stated above.

5.3.1.3. Humanness of the head noun

Since the prescriptive grammatical rules taught in English classes in Iran emphasize use of *WHO* with human antecedents and *WHICH* or *THAT* with non-human nouns, we expect to see more *WH*-forms with human NP heads and more *WH*- (*WHICH*) or *THAT* forms with non-human antecedents. This is what we see in Table 5.5. *WH*-forms are selected most for EL2 human head nouns. As shown and also found by Guy & Bayley (1995, (for English)), both English and EL2 non-human antecedents co-occur with *ZERO* more than human NPs.

It is possible that EL2 departs from English in the humanness factor group more because of the effects of the teaching materials. Does this surface discrepancy affect the patterns of variation in the two contexts? As will be shown, EL2 and English utilise strikingly similar patterns of variation in this respect.

⁷² All non-adjacent relative clauses in EL2 behave exactly like English, separated by modifying material while all Persian non-adjacent relative clauses are separated by the verb.

Table 5.5. Relative markers and the humanness of the head noun.

Relative Marker	Human						Non-human					
	English		EL2		Persian		English		EL2		Persian	
	N	%	N	%	N	%	N	%	N	%	N	%
<i>Zero</i>	29	16	14	11	19	20	201	47	79	22	42	10
<i>That</i>	82	46	27	20	78	80	174	41	203	57	393	90
<i>Wh</i>	66	37	91	69			51	12	77	21		
Total	177	29	132	27	97	18	426	71	359	73	435	82

Persian, on the other hand, does not show any specific effect for humanness. Human heads only slightly favour *ZERO* more than non-human nouns, contra English and EL2. What is more obvious in this factor group is the effect of the teaching material and the absence of Persian influence on the EL2 system of relative marking.

5.3.2. Variable rule analyses

In variable rule analysis the basic assumption is that each factor group exerts an independent effect on the selection process. But what happens if, for instance, 80% of premodified NPs are also definite? Is variant selection due to the effects of premodification or definiteness of the head noun in our relative clauses? Here, it is said, these two factors are interacting. Interaction between the factors contaminates results obtained from variable rule analysis.

Variable rule analyses for relativization in previous studies (e.g., Harvie 1998; Kikai et al. 1987; Tottie & Harvie 1999) show that there is interaction among different factor groups, and this is indeed what I found in my data. In the following sub-sections I will first report my findings regarding the interacting factor groups and the methods utilized to eliminate those effects. Then, the results of variable rule analyses will be presented. Three separate sets of analyses were performed for each variant (*ZERO*, *THAT* (*ke*) and *WH*-) in each context (English, Persian and EL2).

5.3.2.1. Interactions

English data, only, are used to present the interacting factor groups. EL2 was found to have similar effects. Interactions among Persian factors (if different from

English) will be discussed when presenting results of variable rule analyses on Persian data.

** Interaction between humanness and function of the relative marker in the relative clause:*

As shown in Table 5.6, human NP heads mostly co-occur with subjective relative clauses while non-human antecedents are mostly followed by objective (direct object) ones (see also Ball 1996, Harvie 1998, Tottie & Rey 1997, Qurik et al. 1985). Relative markers coded as *other*, the majority of which are adverbials, never occurred with human antecedents.

Table 5.6. *Interaction between humanness and function of the relative marker in the relative clause.*

Antecedent	Relative marker in the relative clause		
	Subject	Direct Object	Other
Human	129	32	Ø
Non-human	92	174	88

**Interaction between humanness and function of the NP head in the matrix clause*

Adverbial relative clauses are naturally headed by non-human antecedents, thus, of 35 antecedents coded as having *other* function (adverbial) in the matrix clause, none were human.

**Interaction between humanness and subject of the relative clause*

As shown in Table 5.7, non-human antecedents co-occur more with personal pronouns and nominal items as the subject of the relative clause, while human antecedents co-occur with relative pronouns more as the subject of the relative construction.

Table 5.7. *Interaction between humanness and subject of the relative clause.*

Antecedent	Subject of the relative clause		
	Relative pronoun	Personal pronoun	Noun
Human	129	47	2
Non-human	90	219	25

**Interaction between subject of the relative clause and function of the relative marker in the relative clause*

It is obvious that in subjective relative clauses the relative marker itself is the subject of the relative clause (Table 5.8). Objective relative clauses, on the other hand, usually have a personal pronoun as the subject of the relative clause. There is a possibility that interactions presented in Tables 5.6 & 5.7 are actually a result of interaction between subject of the relative clause and function of the relative marker. I will deal with this more in presenting the results for each relative marker.

Table 5.8. *Interaction between subject of the relative clause and function of the relative marker in the relative clause.*

Subject of the relative clause	Relative marker in the relative clause	
	Subject	Direct Object
Relative pronoun	218	1
Personal pronoun	3	188
Noun	Ø	16

**Interaction between premodification and definiteness factor groups*

Non- premodifiable nouns, in the premodification factor group, include personal pronouns and other words such as *anybody*, *everything*, etc. This means that almost all pronominal items in the definiteness factor group (with definite nouns, indefinite nouns and pronominals as its factors) are the same as non-premodifiable items in the premodification, see Table 5.9. On the other hand, all definite and indefinite nouns are either premodified or non-premodified.

Table 5.9. *Interaction between premodification and definiteness.*

Premodification	Definiteness		
	Definite	Indefinite	Pronominal
Premodified	110	128	3
Not-premodified	78	109	6
Non-premodifiable	Ø	1	76

As a general solution to the interaction problem, first, adverbial relative clauses were all excluded from further analysis, for the reasons stated above (section 5.2.1). As a second solution, variable rule analyses were performed in two separate phases. First, after running the program with all factor groups to see the effects of interactions on the probabilities obtained at different levels, subject of the relative clause factor group was excluded from later analyses. This factor group interacted with several other factor groups such as humanness, function of the relative marker in the relative clause and function of the NP in the matrix clause. Moreover, collapsing subject of the relative clause and some other factor groups like function of the relative marker in the relative clause was not linguistically justified; and efforts to remove this interaction resulted in further interaction among the new factor group and other factors (see also Harvie 1998; Tottie & Harvie 1999). Collapsing relative pronoun and personal pronoun factors into a single factor of pronominal subjects also did not help remove the interaction. Therefore, in the first phase of the analysis I did not consider the effects of subject of the relative clause on relative marker choice. And, to factor out other interaction effects, for some analyses not-premodified and non-premodifiable antecedents were collapsed into one factor (not-premodified), because they are both not-premodified anyway. Since pronominal and definite nouns are both semantically speaker/hearer-known and also their behaviour was similar in various analyses, they were combined as a single factor (definite). The interaction between humanness and other factor groups, when affecting the results, was handled by collapsing or excluding (either of) them from the analyses.

In the second phase of the analyses, all tokens of subjective relative clauses (when relative marker functions as the subject of the relative clause) were excluded. In this way I was able to examine the effects of all factor groups, including subject of the relative

clause. This also helped to reduce interaction between some other factor groups like humanness and subject of the relative clause or humanness and function of the relative marker in the relative clause.

5.3.2.2. Variants and contexts

In this section a sketch of the system of variable relative marking in second language speech will be presented. Then the comparison and contrast of the results obtained for EL2 with those of English and Persian will follow.

5.3.2.2.1. EL2 system of relative marking

The first thing the findings show is that variable use of relative markers in EL2 is conditioned by certain linguistic factors. Now, let us see what factors play a prominent role in conditioning EL2 relative marker choice. As shown in Table 5.10, for the *ZERO* relatives, only one factor group, function of the relative marker, was found to make a significant contribution, even for non-subjective relative clauses (not reported in this table, see Table 5.12). Direct objects favor *ZERO* more than other functions and subjects favor *ZERO* the least.

The analysis of variation in the use of *THAT* relatives in EL2 demonstrated more interaction and greater complexity, while I expected otherwise because of its prevalent and systematic usage in both English and Persian. After many analyses to factor out interaction effects, the semantic factor of humanness and two syntactic factor groups were found to constrain the use of *THAT* in EL2. *THAT* is much more likely with non-human antecedents (.60) than human NP heads (.31). Objective relative markers correlate with *THAT* more than those with a subjective function (.62 for direct objects and .69 for objects of preposition vs. .34 for subjects), while subjective antecedents, with a probability of .56, favor *THAT* more than objective ones, direct objects in particular (.41)

Table 5.10. *Contribution of factors selected as significant to the selection of relative markers in EL2 relative clauses (subjective relative clauses included):*

<i>Total N=403</i>	<i>n=157</i>	<i>n=183</i>	<i>n=63</i>
Definiteness of the antecedent	wh-	that	zero
Indefinite NP	.58 86/49%	[]	[]
Definite NP	.44 71/31%	[]	[]
Humanness of the antecedent			
Human	.70 91/69%	.31 27/20%	[]
Non-human	.40 66/24%	.60 156/58%	[]
Function of the relative marker in the relative clause			
Subject	.77 130/67%	.34 50/26%	.35 15/8%
Other	.60 3/30%	.54 6/60%	.42 1/10%
Object of preposition	.35 16/18%	.69 59/67%	.53 13/15%
Direct object	.16 8/7%	.62 68/62%	.74 34/31%
Function of the antecedent in the matrix clause			
Subject complement	[]	.66 41/59%	[]
Subject	[]	.56 36/49%	[]
Object of preposition	[]	.50 40/48%	[]
Other	[]	.47 10/43%	[]
Direct object	[]	.41 39/37%	[]
Existential subject	[]	.39 17/35%	[]

Effects of teaching materials considerably increased frequency of *WH*- use in EL2. Rules of grammar taught to EL2 learners, however, affected the frequency of usage only, not the system of variable marking. As seen in Table 5.10, human antecedents favor *WH*-forms with a probability of .70 and non-human NPs disfavor them with a probability of .40; and this is what has been found for English in previous studies. Definiteness of the head noun and function of the relative marker also condition selection of *WH*- markers. Indefinite antecedents favor *WH*- forms, while objective relative markers do not have a strong preference for *WH*- relatives.

Now that we know what constraints condition relative marker choice in EL2, we are at a point where we are able to see if the EL2 system of marking originates from EL2-internal factors, follows the English system of relative marker usage, or relies on the system of the informants' native language. In the following I will compare and contrast weight and hierarchy of factors that have been found to condition relative marker choice in EL2 with those of Persian and English in every individual context. This will help to shed light on the system that EL2 utilizes in its relative clause constructions.

5.3.2.3. Variants and contexts: EL2 compared to Persian and English

In this section I compare the behaviour of relative markers in EL2 to those of Persian and English.

5.3.2.3.1. Zero

Table 5.11 includes factor groups that were found to condition the selection of *ZERO* in English, EL2, and Persian. Selection of *ZERO* in EL2 is conditioned by only one syntactic factor group, the function of the relative marker in the relative clause. Objective relative clauses correlate with *ZERO* more than subjective ones (.74 and .53 vs. .35). This is exactly what has been found for English in this (.81 and .72 for objects and .15 for subjects) and all other studies. Even according to the ranges, this factor group exerts the strongest effect in both English and EL2. It not only was not selected in Persian, but also the constraint hierarchy in Persian is noticeably different. Subjective clauses co-occur more with *ZERO* (.52) and object of preposition, in particular, disfavours *ZERO* much more than EL2 (only .16). Definiteness was not selected in EL2; however, the constraint hierarchy in English and EL2 is identical⁷³, contrary to Persian. Function of the antecedent in the matrix clause was not selected in EL2, while it was selected in both English and Persian (with strongest effect in Persian: range = 61). Persian exerts some influence here; therefore, the EL2 constraint hierarchy (not selected, though) is similar to Persian in this context (direct object > subject > object of preposition).

Remember that variable rule analyses were performed in two phases: first with subjective relative clauses (the type of the subject of the relative clause factor group excluded) and then without them. This helped us to both incorporate subjective relative clauses in the study and find their effect in the first phase, and to explore the effects of the type of subject of the relative clause in the second phase, and at the same time reduce the influence of interaction between these and other factor groups.

⁷³ Probabilities for non-selected factor groups were reported from the first stepping down run of the program, where all factor groups are analyzed together.

Table 5.11. Contribution of factors selected as significant to the selection of ZERO in English, EL2, and Persian relative clauses (all) ⁷⁴.

	<i>n=169/511</i>	<i>n=63/403</i>	<i>n=58/488</i>
	English	EL2	Persian
Function of the relative marker in the relative clause			
Direct object	.81 117/57%	.74 34/31%	[.59]*
Object of preposition	.72 40/47%	.53 13/15%	[.16]
Subject	.15 12/5%	.35 15/8%	[.52]
Other	NA	.42 1/10%	[.45]
Range	.66	.39	
Definiteness of the antecedent			
Pronominal	.71 47/55%	[NA]	[NA]
Definite NP	.53 75/40%	[.56]	[.45]
Indefinite NP	.40 47/20%	[.42]	[.53]
Range	.31		
Function of the antecedent in the matrix clause			
Subject	.74 36/55%	[.46]	.43 15/6%
Subject complement	.73 61/57%	[.43]	NA
Other	.50 9/29%	[.63]	.45 2/7%
Object of preposition	.47 23/28%	[.40]	.20 2/2%
Existential subject	.34 11/13%	[.67]	NA
Direct object	.32 29/20%	[.54]	.81 39/30%
Range	.42		.61
Premodification of the antecedent			
Not-premodified	[.49]	[.52]	.63 53/16%
Premodified	[.51]	[.47]	.25 5/3%
Range			.41
Factors not selected:			
Adjacency	X	X	X
Humanness	NA	X	NA
Premodification	X	X	
Definiteness		X	X
Function of the antecedent in the matrix			X
Function of the relative marker in the relative clause			X

* Probabilities in the square brackets were not selected; they are reported from the first stepping down run of the program.

⁷⁴ In the variable rule analyses I do not consider position and length factor groups anymore since, as for the previous research, these factors were found not to have any effects on variation here. In EL2 the definiteness factor group, pronominal and definite NPs are collapsed. Humanness in English and Persian (also in Table 5.12) was not considered because of its interaction with other factors. However, in other analyses, with or without humanness, the results were the same as those of Table 5.11.

The results for non-subjective relative clauses are presented in Table 5.12. As expected, in EL2, the function of the relative marker is selected as significant with a constraint hierarchy identical to that of English (although this factor group was not chosen in English). The Persian constraint hierarchy parallels English and EL2, but its objects of preposition disfavor *ZERO* more than those of both English and EL2. The similarity between English and EL2 is more striking in other contexts such as subject of the relative clause and definiteness, where constraint hierarchies are identical, contra Persian. Another Persian selected factor group, premodification, seems not to have an effect on EL2, either. There is only one site (in Tables 5.11 and 5.12) where EL2 seems to be affected by Persian: syntactic function of the antecedent in the matrix clause. There is a possibility that, since this factor group has the strongest effect in Persian (range = .74) and direct objects almost categorically favor *ZERO* marking (.90), EL2 learners tend to be affected by Persian direct objects, which causes a change in constraint hierarchies shown in 98.

98.	English	Persian	EL2
	Subject	Direct object	Subject
	Object of preposition	Object of preposition	Direct object
	Direct object	Subject	Object of preposition

A probable explanation is that since this syntactic factor group is selected, with the highest ranges, in both English and Persian, and its effect is neither totally different nor completely identical in both languages (subjects behave differently and objects of preposition have similar effects), EL2 learners face difficulty in their behavior in this respect. They follow English in some contexts (e.g., object of preposition, subjects) and are affected by Persian in others (direct objects).

Generally speaking, factors that constrain choice of *ZERO* relative marker in EL2, their weight and their constraint hierarchy are mostly similar to those of English and different from Persian. This is particularly true in contexts where the constraint hierarchies contrast with each other in Persian and English, such as subject of the relative clause and definiteness. Therefore, at least in these contexts no EL2-specific system of *ZERO* marking was found. Persian did show some effects on EL2, as well.

Table 5.12. *Contribution of factors selected as significant to the selection of ZERO in English, EL2 and Persian relative clauses: subjective relative clauses excluded.*

	<i>n=157/291</i>	<i>n=46/206</i>	<i>n=37/263</i>
Definiteness of the antecedent	English	EL2	Persian
Definite NP	.57 114/64%	[.56]	[.50]
Indefinite NP	.39 43/38%	[.40]	[.50]
<i>Range</i>	.31		
Function of the antecedent in the matrix clause			
Subject	.75 35/81%	[.48]	.16 1/1%
Subject complement	.72 57/77%	[.48]	NA
Other	.43 8/50%	[.76]	NA
Object of preposition	.41 21/44%	[.43]	.37 1/3%
Direct object	.32 28/36%	[.51]	.90 34/39%
Existential subject	.26 8/25%	[.53]	NA
<i>Range</i>	.52		.74
Subject of the relative clause			
Personal pronoun	.53 149/56%	[.51]	NA
Noun	.26 8/30%	[.44]	.34 7/7%
Zero	NA	[NA]	.63 29/23%
<i>Range</i>	.28		.29
Function of the relative marker in the relative clause			
Direct object	[.53]	.61 32/30%	[.56]
Object of preposition	[.43]	.39 13/15%	[.27]
Other	[NA]	.29 1/10%	
<i>Range</i>		.32	
Premodification of the antecedent			
Not-premodified	[.48]	[.51]	.63 34/22%
Premodified	[.52]	[.48]	.25 2/3%
<i>Range</i>			.38
Factors not selected:			
Adjacency	X	X	X
Humanness	X	X	NA*
Premodification	X	X	
Definiteness		X	X
Function of the antecedent in the matrix		X	
Function of the relative marker in the relative clause	X		X
Subject of the relative clause		X	

* See footnote 74.

5.3.2.3.2. That

According to the traditional contrastive analysis, when both native and target languages share a linguistic feature, say a grammatical structure, that very feature is one of the easiest to learn for second language learners. For the proponents of markedness theory in SLA, also, an unmarked (the most frequent, as one criteria for unmarkedness) feature in both native and target languages is assumed to be easily acquired by L2 learners (e.g., Eckmann 1977; Kellerman 1983; Jordens 1992). The case of *THAT*-relatives in Persian and English suitably fits both these assumptions. English restrictive relative clauses are marked with *THAT* the most (42%) and the proportion of *ke* 'that' in Persian is 89%. Therefore, a natural expectation is to see more frequent and systematic use of *THAT* in EL2 compared to *ZERO* or *WH*-forms which are not unmarked options in Persian.

On the one hand, frequency of *THAT* tokens in EL2, although more frequent than *WH*- or *ZERO*, is almost identical to that of English ((47%), as shown in figure 5.1), much less than its frequency in Persian (89%). And, on the other hand, in my analyses, the system of *THAT* marking in EL2 was found to be much more complicated than that of *ZERO* or *WH*-forms. Severe dissimilarity of systems of *THAT* marking in Persian and English, as seen in Table 5.13, may lead to confusion in EL2 speech, particularly when the two systems are similar at surface level but different in their constraint hierarchies.

In Table 5.13, premodification of the antecedent only contributes an effect in Persian, differing from English. L2 learners pick up the system of the target language, with the same weight and hierarchy for factors in this context (and non-selection of this factor group). Humanness operates in English, but not in Persian⁷⁵; EL2 informants show identical behavior to that of English speakers for this factor group, as well. But when it comes to a factor group that is operational in both systems, but with different hierarchies of effects, EL2 learners come across difficulties in applying the target language system. For the syntactic function of the antecedent, for instance, constraint hierarchies for the three languages are shown in 99 (only factors common in 3 contexts). As seen, direct objects (and subjects) behave similarly in Persian and EL2, while EL2 objects of

⁷⁵ Humanness was excluded in Persian because of its strong interaction with other factor groups. However, our marginals and preliminary analyses indicate that human antecedents do not exert significant effects on the use of *THAT*.

Table 5.13. *Contribution of factors selected as significant to the selection of THAT in English, EL2 and Persian relative clauses (all).*

	<i>n</i> =246/511	<i>n</i> =183/403	<i>n</i> =430/488
Humanness of the antecedent	English	EL2	Persian
Non-human	.56 164/49%	.60 156/58%	[NA]
Human	.40 82/46%	.31 27/20%	[NA]
Range	.16	.29	
Function of the antecedent in the matrix clause			
Existential subject	.64 53/65%	.39 17/35%	NA
Direct object	.58 80/56%	.41 39/37%	.19 92/70%
Object of preposition	.51 39/48%	.50 40/48%	.80 80/98%
Other	.42 13/42%	.47 10/43%	.55 28/93%
Subject complement	.41 41/38%	.66 41/59%	NA
Subject	.34 20/30%	.56 36/49%	.57 230/94%
Range	.30	.27	.61
Function of the relative marker in the relative clause			
Subject	.64 131/60%	.34 50/26%	[.48]
Object of preposition	.40 35/41%	.69 59/67%	[.84]
Direct object	.39 80/39%	.62 68/62%	[.41]
Other	NA	.54 6/60%	[.55]
Range	.25	.35	.43
Premodification of the antecedent			
Premodified	[.48]	[.49]	.75 158/97%
Not-premodified	[.52]	[.51]	.37 272/84%
Range	.4	.2	.38
Factors not selected:			
Adjacency	X	X	X
Premodification	X	X	
Definiteness	X	X	X
Function of the relative marker in the relative clause			X

99.	English	Persian	EL2
	Direct object	Object of preposition	Subject
	Object of preposition	Subject	Object of preposition
	Other	Other	Other
	Subject	Direct object	Direct object

preposition follow English patterns. The other syntactic factor group, function of the relative marker, is selected in both English and EL2 (Table 5.13). English direct objects and objects of prepositions disfavor *THAT*, while subjective relative clauses, here, prefer

THAT marking the most. Contra English, Persian (this factor group was not selected in Persian) objective relative clauses show a very different pattern (.84 probability for objects of prep. and .41 for direct objects) and Persian subjective relative clauses co-occur with *THAT* less than those of English (.48 vs. .64 in English). In EL2, subjects disfavor *THAT* (.34) and objects have the same hierarchy as in English, but with different rates. In this factor group, too, Persian influence is seen in two contexts: subjects, that disfavor *THAT* in both Persian and EL2 and objects of preposition that show a higher probability like Persian.

Results of the second phase of the analyses, subjective relative clauses excluded, are reported in Table 5.14. In EL2, only subject of the relative clause was selected as significant. There were no tokens of pro-drop (a property of Persian) in EL2, indicating no influence from Persian in this respect. Pronominal subjects slightly favor *THAT*, (much more than Persian, similar to English) and nominal subjects disfavor *THAT*, unlike English or Persian (maybe due to scarcity of tokens, only 13 of them). Premodification shows similar behavior in English and EL2, unlike Persian, as before. The constraint hierarchies for the function of the antecedent are shown in 100.

100.	English	Persian	EL2 ⁷⁶
	Direct object	Subject	Direct object
	Object of preposition	Object of preposition	Object of preposition
	Subject	Direct object	Subject

As seen this time the constraint hierarchies in English and EL2 are identical.

The question is why is there such a fluctuation in the behavior of EL2 speakers? When subjective relative clauses are included they behave more like Persian and when they are excluded, their behavior is more like English. One explanation is that subjective relative clauses had some effects on the results obtained in the first analysis (Table 5.13). For instance, EL2 subjective relative clauses disfavored *THAT* marking more than other contexts (a possible effect of Persian), and its exclusion not only causes non-selection

⁷⁶ This factor group was not selected in EL2, it is used here because it had a significant effect before (Table 5.13), it has the second largest range (.12) in EL2, and, also, to show its behavior compared to other languages, as in 91.

Table 5.14. Contribution of factors selected as significant to the selection of *THAT* in English, EL2, and Persian relative clauses: subjective relative clauses excluded

	<i>n</i> =114/291 English	<i>n</i> =133/206 EL2	<i>n</i> =195/231 Persian
Function of the antecedent in the matrix clause			
Existential subject	.74 20/63%	[.49]	NA
Direct object	[.65] 41/53%	[.51]	.10 54/61%
Other	.63 8/50%	[.44]	NA
Object of preposition	[.59] 22/46%	[.48]	.63 34/97%
Subject complement	.33 17/23%	[.56]	NA
Subject	[.22] 6/14%	[.47]	.84 107/99%
Range	.52	.12	.74
Premodification of the antecedent			
Premodified	[.47]	[.49]	.75 73/97%
Not-premodified	[.53]	[.51]	.37 122/78%
Range			.38
Subject of the relative clause			
Noun	[.65]	.32 13/46%	.66 96/93%
Zero (pro)	[NA]	[0%]	.37 99/77%
Pronoun	[.48]	.53 120/67%	[1%]
Range		.21	.29
Factors not selected:			
Adjacency	X	X	X
Humanness	X	X	NA
Premodification	X	X	
Definiteness	X	X	X
Function of the antecedent in the matrix		X	
Function of the rel. marker in the rel. clause	X	X	X
Subject of the relative clause	X		

of this factor group in Table 5.13, but also affects the findings in a way that influence of Persian in other factor groups is reduced as well. Therefore it may be concluded that Persian affects only one context, that of subjective relative clauses, and this, apparently, results in other L1 effects. Another explanation deals with the differences and similarities in the patterns of variation between English and Persian. Whenever the patterns are totally different (or similar) in the two systems involved in the learning situation, L2 learners face no important problem in adopting the target language system (e.g., premodification and humanness). But when a factor group has significant effects in both native and target languages, with different constraint hierarchies (some factors show

similar effects, while others behave differently) the acquisition of target language patterns gets complicated. L2 learners fluctuate between the two systems of L1 and target language. The influence of native language is only seen in these contexts.

5.3.2.3.3. *WH*-forms

Persian relative clause constructions are not marked by a relative marker equivalent to English *WH*-forms; therefore, this section is devoted to the comparison of the use of *WH*-relative pronouns in EL2 and English only⁷⁷. Since Persian and English utilise a similar general relative marker (*THAT* and *ke* in respective languages⁷⁸) and *WH*-forms comprise a new system of relative marking confronted by Persian learners, our expectation was to witness a large majority of EL2 relative clauses marked with *THAT*, and only a minimum number of *WH*- tokens in EL2. In contrast to this expectation there are 168 (34%) tokens of *WH*- usage in the EL2 data, more than English (117/19%)⁷⁹. We related this frequent use of *WH*-forms to the effects of instructional materials in English classes. But this is only one side of the coin. As seen in Table 5.15, Persian EL2 learners have not only learned how to mark their relative clauses with *WH*-forms, but they have also acquired some portions of the patterns of variable usage of *WH*-marking of their target language. Selected factor groups in EL2 are identical to those of English, except for the function of antecedent that is selected only in English. Constraint hierarchies are parallel for all selected factor groups.

⁷⁷ I will also report only the findings of the first phase of the variable rule analyses (all tokens, subjective relative clauses included, subject of the relative clause excluded). Findings of the second phase of the analyses were not that different from the first phase.

⁷⁸ They are both general relative markers without gender, number, and person features.

⁷⁹ Numbers and percentages based on overall distribution of the data, adverbials included.

Table 5.15. *Contribution of factors selected as significant to the selection of WH-forms in English and EL2 relative clauses (Persian not applicable) (all).*

	<i>n</i> =96/511	<i>n</i> =157/403
Definiteness of the antecedent	English	EL2
Indefinite NP	.61 66/28%	.58 86/49%
Definite NP	.41 30/11%	.44 71/31%
Range	.20	.14
Humanness of the antecedent		
Human	.69 66/37%	.70 91/69%
Non-human	.40 30/9%	.40 66/24%
Range	.29	.30
Function of the relative marker in the relative clause		
Subject	.71 76/35%	.77 130/67%
Other	NA	.60 3/30%
Object of preposition	.56 11/13%	.35 16/18%
Direct object	.26 9/4%	.16 8/7%
Range	.45	.61
Function of the antecedent in the matrix clause		
Other	.66 9/29%	[.46]
Direct object	.62 34/24%	[.58]
Object of preposition	.60 20/24%	[.57]
Subject	.52 10/15%	[.45]
Existential subject	.47 18/22	[.49]
Subject complement	.25 5/5%	[.36]
Range	.41	.22
Factors not selected:		
Adjacency	X	X
Premodification	X	X
Function of the antecedent in the matrix		X

What we understand from these findings is that EL2 learners have acquired both the specific linguistic feature and its system of variable use. This might be used as a criterion to distinguish if a target language system is actually acquired by L2 learners or if tokens of the target language forms are only an indication of rote learning, not their native-like acquisition. Findings reported in Table 5.15, also indicate the effectiveness of language instruction and training material on the rates of rule application, despite claims made against the usefulness of instruction (Krashen 1981).

5.4. Discussion

Of the findings of the variable rule analyses an important one is that the second language system is indeed systematic. It was shown that variation in the use of relative pronouns in EL2 is conditioned by some of the linguistic factors incorporated in this study. Now we know that syntactic function of the relative marker in the relative clause is the only linguistic factor that significantly contributes to the selection of *ZERO* in EL2 restrictive relative clauses. Use of *THAT* is conditioned by the humanness and syntactic function of the head noun and also syntactic function of the relative marker. Variation in the use of *WH*-forms is significantly constrained by definiteness and humanness of the head noun and syntactic function of the relative marker. There was no evidence for any kind of unsystematic (lack of any conditioning factor) variation in the data, disproving the claims made by Ellis (1987) and others in this respect.

Table 5.16. *Factors selected as significant for all variants in English, EL2, and Persian (extracted from Tables 4.11, 4.13, and 4.15)*

Variant	Factor groups selected as significant		
	English	EL2	Persian
<i>ZERO</i>	*Syntactic function of the relative marker *Definiteness of the antecedent *Syntactic function of the head noun *Not selected	*Syntactic function of the relative marker *Not selected *Not selected *Not selected	*Not selected *Not selected *Syntactic function of the head noun *Premodification of the head noun
<i>THAT</i>	*Humanness of the head N *Syntactic function of the head noun *Syntactic function of the relative marker *Not selected	*Humanness of the head N *Syntactic function of the head noun *Syntactic function of the relative marker *Not selected	*Not selected *Syntactic function of the head noun *Not selected *Premodification the head noun
<i>WH-</i>	*Definiteness of the head noun *Humanness of the head N *Syntactic function of the relative marker *Syntactic function of the head	*Definiteness of the head noun *Humanness of the head N *Syntactic function of the relative marker *Not selected	

Were constraints on variable relative marker usage in EL2 independent of the learners L1 and/or their target language? Table 5.16 presents factor groups that contribute the most significant effect to the selection of *ZERO*, *THAT* and *WH*- in the three languages. As shown in this table, there is not a single factor group found to be significant in EL2 that is not significant in English. This indicates that the EL2 system of relative marking is in many cases parallel to the one that operates in English.

It may be supposed that the EL2 might be conditioned by factors other than those incorporated in this study (social factors, for example). Although this assumption holds (for both English and Persian context, as well), the combination of selected factor groups, and the hierarchy of constraints and similarity of factor weights in some contexts all emphasize the fact that the EL2 system is related to English as the learners' target language. Another claim would be that EL2 and Persian are similar in non-selection of some factor groups and EL2 and English are different for the same reason. According to Table 5.16, only definiteness was not selected in either Persian or EL2 when using *ZERO* relatives. As shown in Table 5.11, the constraint hierarchy for this factor group in EL2 is in the opposite direction of Persian. EL2 behaves identically to English in this respect. Non-selected factor groups in EL2 that are, at the same time, selected in the English context include definiteness and syntactic function of head noun for *ZERO* and syntactic function of the head for *WH*-forms. Tables 5.11 and 5.15, indicate that EL2 constraint hierarchies are closer to English in these contexts, as well (although different in some cases).

EL2 patterns, on the other hand, are different from those of Persian, although there is evidence for effect of Persian conditioning factors on EL2 relative usage. In one instance the selected factor group is identical in EL2 and Persian. Syntactic function of the NP head in the matrix clause had a significant effect in selection of *THAT* as the relative marker in English, EL2 and Persian. Mention was made that Persian patterns may exert some effect on EL2 in this context because of the selection and strong effect of this factor group in Persian⁸⁰. This influence of the L1 is more negative in nature (negative transfer or interference) in the sense that it prohibits the learner from using (acquiring)

⁸⁰ Second phase of the analyses (without subjective relative clauses), however, indicated similarity of EL2 with English for this factor group.

the target language pattern with the same ease s/he uses (acquires) other systems where this complexity does not exist. According to Myhill (1982) one reason for this is that in these situations (when the L1 structures are similar but not identical to those of the target) the learners are “deluded into thinking that they know more than they do know, and so they will ‘turn off’ their learning process” (p.199).

The findings of this chapter showed that, although variation in EL2 is systematic, the patterns of variation are not L2-specific in any sense. Not only are the selected factor groups, in many contexts, identical to those of English, but the relative weight of factors and constraint hierarchies are similar to English as well. In contexts where EL2 patterns are different from English, Persian hierarchies are found to be effective. These findings shed doubt on the so-called interlanguage hypothesis where independent patterns of language use are claimed to be present in L2 speech.

A question that deserves consideration at this stage is to see if the findings reported here can help us induce what features of L1 have greater chance to affect EL2 patterns and what features of the target language are acquired without any L1 disturbance. There might be two responses to this question according to either *selection* of factor groups or *internal* patterns of variation in each factor group. First, based on the findings in Tables 5.11, 5.13, and 5.15, the following (summarized in Table 5.17) may hold, at least for the contexts studied here:

Table 5.17. Effects of L1 on EL2: selection of factor groups

Context			Effect of L1
English	Persian	EL2	
✓	X	✓	No effect
X	✓	X	No effect
✓	?	✓	Some effect

- Factor groups that do not exert a significant effect in L1 (not selected): EL2 learners easily follow the patterns of the target language in these contexts. This includes syntactic function of relative marker and definiteness for *ZERO* (Table 5.11), humanness for *THAT* (Table 5.13), and all selected factors for *WH*-forms (Table 5.15). Non-significant factors of L1, then, have little chance of affecting L2 acquisition.

- Factors that are significant in Persian but not significant in English (e.g., premodification for both *ZERO* and *THAT* (Tables 5.11, 5.13)): These factor groups were not selected as significant in EL2, and they had a slight influence on the factor weights in only one context, that of *ZERO*. Different internal patterns (i.e., if constraint hierarchies in L1 and the target are identical or different) might have different effects, though. In the cases studied here, the patterns are contrasting (opposite hierarchies in English and Persian), so there is no important effect.
- Factors that are selected as significant in both English and Persian: these contexts show the greatest influence from L1 on EL2 patterns, as is the case for syntactic function of the antecedent in the matrix clause for *THAT* and *ZERO* (Tables 5.11, 5.13).

As discussed before, when factor group-internal patterns (hierarchies) are completely similar or different in native and target languages (e.g., definiteness and function of the relative marker for *ZERO*, premodification for *THAT*), it is easier for the L2 learners to acquire and follow the system of the target language. But, when a certain factor group is selected both in native and target languages with complex hierarchies, i.e., some factors behave similarly and some others differently (e.g., function of the antecedent in the matrix), L2 learners show their confusion by fluctuating between the two systems. In these cases some factors resemble patterns of L1 and others follow those of the target language. The target language system is yet to be totally applied by the learners in these contexts.

In section 5.3.1, when discussing the marginal results, it was found that surface patterns of variation may suggest influence of the Persian system on EL2. In those marginal results, moreover, prevalence of *THAT* (*ke*) relatives in Persian had influenced use of this marker in EL2 at least in four contexts of humanness, definiteness, syntactic function of the head, and syntactic function of the relative marker. If I had stopped at that level of analysis, I could ‘quantitatively’ (according to the similarities in the frequency of occurrences) argue for crosslinguistic influence from Persian to EL2 and claim to have found evidence for language transfer in almost all these contexts. This is what was usually done in many contrastive analysis studies of second language speech. My

findings reveal that surface similarities of native, L2, and target languages, be it surface structural similarities or similar frequency of occurrence of those structures, are of little indication of any equivalence in patterns of variation. What is the most important in these contact/learning situations is the underlying systematic patterns of variable use of linguistic features that reveals the basic knowledge of the speakers with respect to that feature.

Chapter six: RELATIVE CLAUSES: Resumptive Pronouns

6.0. Introduction

One of the most common patterns in relative clauses is to have within the relative clause a gap with the same reference as the head noun in the matrix clause (Keenan 1985) (as in 102.a). In some languages like Persian (and English) a pronoun, a so-called resumptive pronoun⁸¹, may be found in the place where this gap is expected. Example 101, where *her* is the resumptive pronoun with the same reference as *woman* in the main clause, is a typical Persian sentence usually cited in the literature. An English example, from the English corpus (OSLA), is presented in 102.b.

101. *man zan-i ; ra: ke John be u ; sibe zamini dad mishena:sam.*

I the woman that John to her potato gave know-1SG

I know the woman that John gave a potato (to her)

102. a. There is certain *things_i* Ø you don't do [*gap_i*]. (OSLA 241.Ia. 266)

b. It's all *the same thing_i* we keep doing *it_i*. (OSLA 107. IIb. 77)

Presence of resumptive pronouns in the English spoken by second language learners, particularly learners whose native languages mark relative clauses with resumptive pronouns, has initiated a wide variety of research in SLA. Traditionally, many of these studies have focused on properties of the learners' first language to discover the sources of learners' "errors" in marking English relative clauses with resumptive pronouns. This chapter is devoted to the exploration of the nature of resumptive marking in the EL2 of Persian speakers. More specifically I want to know if resumptive marking in EL2 is variable, if variability (provided that it is variable) is conditioned, and, finally, what the sources of that variability are. First, a summary of the literature on the grammatical status of resumptive pronouns in English, Persian and EL2 is presented. Then, after an introduction to the coding system and method, results of the variable rule analyses on native Persian, native English, and conversational EL2 data will be reported. A discussion of the results will end this chapter.

⁸¹ Epenthetic pronoun, retained pronoun, pronominal reflex, pronominal copy, intrusive pronoun, and replacive pronoun are the other names for the same phenomenon

6.1. Resumptive pronouns in English, Persian, and EL2

6.1.1. Resumptive Pronouns in English

Resumptive pronouns have not attracted much attention in English linguistics. This is mainly because use of resumptive pronouns is not common in this language. Grammar books have only a sentence or two about the mostly unacceptable use of these ‘alien’ pronouns in older English by those speakers ‘without grammatical conscience’ (Curme, 1931). Sells (1984), however, asserts that, in English, a pronoun may be found in the place where we expect a gap as in 103, where *them* fills an expected gap after *read*.

103. The *books*_i which I don’t remember who read *them*_i.

But, he adds, the acceptability of these sentences varies according to dialect and register. Sells names these resumptive pronouns as ‘intrusive’ and Jespersen (1928) calls these structures ‘exhausted relative clauses’⁸²; both terms indicate that the pronoun in English sentences of this kind is seen as somehow syntactically ‘alien’ (Asher 1994), a filler inserted as an aid to comprehension. Kroch (1981), based on native English examples with resumptive pronouns (extracted from naturalistic data), argues for a syntactic rule for insertion/base-generation of resumptive pronouns in some structures of English; and Prince (1995), makes an attempt to show that some of the sentences in Kroch’s data with resumptive pronouns, which she calls *Kind*-sentences,⁸³ are not actually relative clauses. Generally speaking, resumptive pronoun usage is not common⁸⁴, and acceptability of their usage varies in different dialects of English.

6.1.2. Resumptive Pronouns in Persian

Mention was made in the previous chapter that in Persian relative clause constructions there is no relative pronoun; relative clauses are usually introduced by a dummy element *ke* ‘that’, as in 104:

⁸² Jespersen calls them exhaustive because he asserts “not infrequently a relative clause, which has been begun in the ordinary way, is continued irregularly as if the power of the relative were exhausted, a personal pronoun being substituted for it” (p. 107)

⁸³ As in ‘He’s the kind of guy that he gets into a lot of fights.’.

104. *ma: hamun reshte-i ke mixa:stim qabul shodim* (Per. FS. Ib. 287)
 we that field that wanted admit became

We were admitted to the field that we wanted.

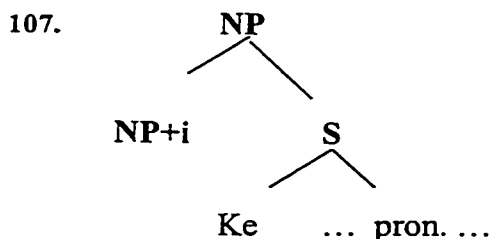
According to the prescriptive grammars, Persian relative clauses are marked by resumptive pronouns, optionally for subject and direct objects, and obligatorily for the other functions of the relative marker (or a pronoun in the relative clause that refers to the head noun)(Windfuhr 1979, Sadighi 1982, Lazard 1992). Windfuhr asserts that since there are no (specific) relative pronouns in Persian, instead, the related identical noun occurs as a personal (resumptive) pronoun, separate or suffixed as in 105: (Windfuhr p. 62)

105. *man ke peser-e man / man ke pesar-am*
 I that son-EZ I I that son-I I whose son.

He adds that in contemporary Persian, except for emphatic instances, the resumptive pronoun is elided as the subject or direct object of the relative clause. Lazard (1992) also believes that the resumptive pronoun in subject or direct object functions may be omitted when there is no risk of ambiguity in the sentence. They are the only grammarians who also refer to ellipses of resumptive pronouns in adverbial relative clauses in Persian (106). The conditions of this phenomenon have not been determined yet.

106. *ja:-i ke be ja: residi / ja:-i ke (be a:n) residi*
 place that to place arrive (to that)
 the place that (to it) you came to/arrived at

Sadighi (1982) presents the following as the Persian relative clause structure in which a resumptive pronoun is an obligatory element (in deep structure).



⁸⁴ According to Kroch (cited in Tarallo & Myhill 1983), in his study only 2% of the data included resumptive pronouns.

6.1.3. Resumptive Pronouns in the English of Persian speakers

Traditionally, many of the second language studies within the domain of restrictive relative clause constructions have focused on differences between the first and second languages in terms of their different properties to discover in what ways, and at which levels, knowledge of the first language may mediate the second language acquisition process. The effect of Persian relative clause constructions on the acquisition of English by Persian speakers, therefore, has been the focus of several studies. Schachter (1974, 1983) analysed the written performance of Persian, Arabic, Chinese and Japanese L2 learners of English and found that Persians and Arabs produced about as many relative clauses as Americans, and three times as many as Chinese and Japanese learners. But Persians had many more errors due to 'illegal' insertion of a resumptive pronoun than Arabs (81% vs. 39%). Since she assumes obligatory resumptive pronoun insertion is maximal in Persian relative clauses, she concludes that a higher number of errors amongst Persian speakers is to be expected due to the influence of their native language.

Ioup & Kruse (1977) used judgement data from Spanish, Persian, Arabic, and Japanese learners of English. They state that all groups seemed to have problems regarding resumptive pronouns (i.e., they erroneously inserted resumptives in their English sentences). Gass (1979) aimed to see how L1 transfer applied to L2 acquisition of relative clauses. She was particularly concerned with resumptive pronouns and found a possible transfer effect in her judgement data, i.e., those learners with an L1 with resumptive pronouns (Persian, Arabic) were more likely to accept sentences with a pronominal in them than those with an L1 without pronominals (French, Italian, Korean, Portuguese, Japanese and Thai). Tarallo and Myhill (1983) paid attention to factors that affect resumptive usage and found that syntactic function of the relative marker (the Accessibility Hierarchy (Keenan & Comrie 1977)), adjacency of the relative clause and its antecedent, and animacy of the antecedent had some influence on pronominal usage in relative clauses. They believe that resumptive usage is not transferred from learners' L1 but it is a part of the grammar of interlanguage (a natural process in language acquisition). Hyltenstam (1984) used composition, picture identification, imitation and judgement tasks in a study of relative clause acquisition by Persian, Greek, Spanish and Finnish learners of Swedish. He noted that all four groups showed pronominal retention

in a picture task, less of it in compositions, messy results in judgement data and only 11 pronominals in an imitation task. There was nevertheless a difference between the learners according to L1: the Persian and Greek learners supplied more resumptive pronouns than Spanish and Finnish (languages without resumptive pronouns). Schachter (1992) proposes that transfer is not a process at all, it is a constraint on the learners' hypothesis testing process. She presumes that for Persian speakers learning English relative clause structures, the learner hypothesis is to mark a relative clause with a resumptive pronoun, because Persian is a language in which relative clauses are marked by these pronouns. The L1 option is a hypothesis (transfer hypothesis) available to these learners among other hypotheses. Persian learners of second languages use this option (transfer) and mark their relative clauses with more resumptive pronouns than other learners with native languages without resumptive pronouns.

There are studies, however, that have offered universalist arguments for the use of resumptive pronouns in second language speech (Odlin 1989). The evidence for these studies often comes from grammaticality judgement tasks where they find resumptive pronoun insertion errors in the L2 judgments of speakers whose native languages do not prescribe such a phenomenon. All Ioup & Kruse's (1977) subjects (with L1s with or without resumptive pronouns), for instance, had problems regarding resumptive pronouns. In Tarallo & Myhill (1983) native speakers of English studying German, Portuguese and other languages that do not use resumptive pronouns often considered as acceptable ungrammatical sentences that had such pronouns. In another study, native speakers of English accepted sentences with resumptive pronouns in L2 French sentences even though neither French nor English use such pronouns (Birdsong, Johnson, and McMinn 1984).

In sum, various studies of relative clause constructions in the L2 speech of Persian speakers have claimed that:

- Persian relative clauses are marked with resumptive pronouns, optionally for subject and direct objects and obligatorily for the other function of the relative marker.
- Persian learners of other languages use more resumptive pronouns in their L2 speech, and they judge as acceptable more sentences with resumptive pronouns than other learners with languages without resumptive pronouns due to the effects of their L1.

- Transfer from L1 has therefore been supposed as the most important cause of resumptive usage in Persians' second language speech. In addition to L1 transfer effects there might be some universalist arguments for the presence of resumptive pronouns in L2 speech, as well.

One thing that is lacking in these studies is that it was never shown if variation in the L2 learners' errors (as they call it) with respect to the insertion of resumptive pronouns is systematically conditioned by linguistic or non-linguistic factors. Previous studies have not compared patterns⁸⁵ of resumptive usage in L2 to those of learners' native languages to see to what extent the patterns are equivalent. Neither have they dealt with the possibility of resumptive usage in the learners' target languages.

In this study, I make an attempt to find the patterns of resumptive usage in EL2 of the informants and then compare these patterns to those found in learners' native and target languages. In order to achieve this I will perform a variable rule analysis on L2 data from Persian second language speakers of English. The findings will be compared to the results of similar analyses on data from native speakers of English and Persian. If it is found that variable use of resumptive pronouns and its constraints in EL2 mirrors that of Persian and is different from English, my argument will be in favor of transfer as the main source of variability in EL2 in this context. But, if it is found that, in any context, EL2 variability mirrors that of English and is different from Persian, I hypothesize that that feature of the target language is acquired by the learners along with its patterns of variation. Transfer effects hardly exist in these cases. A third possibility, according to the interlanguage hypothesis (Selinker 1972), is to witness contexts where EL2 variation is conditioned by factors, factor weights, and constraint hierarchies independent of and different from Persian or English. In this case, our interpretation will be in favor of an L2-specific pattern of variation.

⁸⁵ Actually many of the studies on L2 resumptives have not looked for any patterns of resumptive marking. They mostly relied on surface presence and frequency counts of these pronominals in L2.

6.2. Analyses

6.2.1. Variable context and coding

In order to study the variable constraints on the use of resumptive pronouns, it is necessary to isolate the linguistic context where variation takes place. Restrictive relative clauses are the major concern of my study here. Any pronoun (and a few tokens of nouns in EL2) in these restrictive relative clauses that had the same reference as the head noun in the matrix clause and could be replaced by a gap was counted as an instance of resumptive pronoun.

I excluded non-restrictive relative clauses from the study because there were not many tokens of them in the three corpora. Moreover, of eight non-restrictive relative clauses in English, none included a resumptive pronoun. Ambiguous tokens (where phonological and syntactic clues were of no help to distinguish if the relative clause is restrictive or non-restrictive), headless relative clauses, and incomplete sentences were also excluded from the study. Some Persian structures that look like relative clauses with a pronominal head (always *in* 'this') and a modifying clause as in 108 (see Samar 1998a) were also excluded since they are categorically not marked with resumptive pronouns.

108. *ruhiyye-e jam?i be in nist ke shoma: un vasile-e ba:zi-ro be yekdigar raddo-badal koni* (NA. Ia.114)

spirit social to this not that you that means play to each-other
interchange do

Social spirit is not in interchanging the thing that you play with.

The variable context, therefore, includes all instances of unambiguous restrictive relative clauses in Persian, English, and EL2, regardless of the presence or absence of resumptive pronouns. As in the previous chapter, data from less proficient learners are not included in the study of the resumptive pronouns (since they did not frequently use relative clauses).

In order to study the effects of different factors conditioning the presence/absence of resumptive pronouns, first, all tokens of restrictive relative clauses were extracted from the data, then the following linguistic factor groups, whose effects may condition pronominal usage, were taken into consideration.

- *Syntactic factor groups*: syntactic function of the head noun in the matrix clause, syntactic function of the relative marker in the relative clause, syntactic category of the subject of the relative clause, type of relative marker, and premodification of the head noun.
- *Formal factor groups*: length of the relative clause, position of the relative clause, and adjacency of the relative clause to the head noun.
- *Semantic factor groups*: humanness and definiteness of the head noun (antecedent).

6.2.1.1. Syntactic factor groups

According to the grammatical accounts, resumptive pronoun usage is a part of Persian underlying grammar; and should, thus, be conditioned by mostly syntactic factors. If resumptive usage in EL2 is also found to be constrained by syntactic factors, as in Persian, there will be grounds to argue for the effects of L1 on L2 (transfer). The lack of influence of syntactic factors on EL2 will indicate that resumptive usage is not (yet) a part of underlying grammar of EL2 speakers and they do not follow the Persian system in this respect. Following syntactic factor groups were incorporated in the analyses.

Syntactic function of the relative marker in the relative clause

Keenan & Comrie (1977) present data from different languages and claim that in all languages it is possible and easier to relativize the subject of the relative clause. Then, they add, come the other positions: direct object, indirect object, etc. They claim that languages do not select from these types at random; an implicational relationship called the Accessibility Hierarchy runs from subjects down to the object of comparison.

Accessibility Hierarchy:

Subject > direct object > indirect object > object of preposition > genitive > object of comparison

If a language has object of comparison, it has all the other types higher in the hierarchy.

If a language has genitive, it also has all the other types higher in the Accessibility Hierarchy, but it may not have object of comparison. No language leaves a particular type

out of the sequence, e.g., no language has indirect object clauses that does not have object and subject clauses.

Resumptive pronoun usage works in the opposite direction in the Accessibility Hierarchy. In Persian, for example, the lower positions in the Accessibility Hierarchy (object of comparison, genitive, object of preposition) are far more likely to have resumptive pronouns (obligatory, according to grammars) than those in the higher levels (subject, direct object). Then one factor whose effects on the presence or absence of resumptive pronouns are studied here is the function of the relativized NP in the relative clause. This factor group includes subjects as in 109, direct objects (110), objects of preposition/indirect objects (111) and other functions. Genitives and objects of comparison were not frequently used in the data. Only one or two tokens of each were used in different data sets; they were then collapsed in the *other* factor. Indirect objects were also not frequent. Persian indirect objects are always governed by prepositional items (then coded as object of preposition); and in English and EL2 indirect objects were collapsed with object of prepositions because of their related syntactic functions (as objects), their analogous behaviour in English relative clauses (Gass, 1979), and infrequent usage of indirect objects⁸⁶.

109. I'm gonna send it to the person that *he's* in States. (EL2 MM. IIa. 20)

110. Yea, I know everybody Ø you name *them*. (OSLA 107. Ia. 24)

111. Yea, well, it's not plane crash that I knew about *it*. (OSLA 107. IIa.284)

Syntactic function of the head noun in the matrix clause

In a pilot study on Persian resumptive pronouns (Samar 1998b), it was found that the syntactic function of the head noun of the relative clause affects the use of these pronominals, with subjects and objects of preposition exerting the most favourable

⁸⁶ Keenan & Comrie (1977) notice that indirect objects do not have a clear typological status in terms of relative clause formation, i.e., they tend to behave as either direct objects or objects of preposition. Pavesi (1986) also asserts that from a more general point of view, indirect objects are widely realized as adverbs of movement. English is a typical example of a language in which indirect objects are not distinguishable from other oblique objects; therefore they are treated as simple objects of preposition even by L2 learners of English (Pavesi 1986).

effects. I coded the function of the head noun as subject (112), direct object (113), object of preposition (114), complement subject (115) and other, exemplified in the following:

112. *The side street* that I was on it joined another street. (OSLA 268: Ia. 226)

113. I know many of *them* by name that they didn't pass at all. (EL2 PM. IIb. 318)

114. He has some problem *with the nerves* which they're sensing the temperature. (EL2 MG. Ib. 84)

115. It's *a name* that you can find it in different languages and different religions. (EL2 MA. Ib. 88)

I also incorporated factor groups of 'type of the subject of the relative clause', 'type of the relative marker' (*THAT/ZERO/WH-*), and 'premodification' of the head noun in the analyses. Type of the subject of the relative clause was not considered in the analyses because nominal subjects did not correlate with resumptive pronouns at all. In the Persian data all the instances of relative clauses with resumptive pronouns were introduced by *ke*, English relative clauses were evenly marked by *THAT/ZERO/WH-* (2%, 3%, and 3% respectively), and EL2 relative clauses were marked by *THAT* 31%, *ZERO* 14%, and *WH_* 16% of the time without significant influence in variable rule analyses. Premodification of the antecedent was not found to have a major role in resumptive marking in any context, either. These factor groups will therefore not be discussed further.

6.2.1.2. Formal factor groups

Based on Jespersen's assertion (1928, see also Tarallo & Myhill 1982), resumptive marking is not an underlying principle of English grammar, but a formal (discourse) strategy to help make the meaning of the utterance transparent to the hearer/reader. If I find that resumptive pronouns fulfill a similar function in EL2, and are conditioned by similar constraints as English, it will be argued that the EL2 system of resumptive marking parallels that of native English. To examine the effects of the formal factor groups the following were incorporated in the analyses.

Adjacency, length and position of the relative clause

Keenan & Comrie suggest that use of resumptive pronoun aids in comprehension of speech by making the structure of the relative clause more transparent. I also quoted Jespersen (ft. 82) who asserted that when the relative clause “is continued irregularly as if the power of the relative were exhausted”, a personal pronoun is substituted for it in the relative clause. The assumption is that when the relative clause does not immediately follow the head noun, or its length is more than the length of a regular sentence, appeal to the use of a pronominal helps make the clause more transparent. Therefore, these clauses are marked more with resumptive pronouns than adjacent or shorter relative clauses. I included the two factor groups of adjacency of the relative clause to its head noun and the length of the relative clause in the analyses. In the former, I coded as adjacent relative clauses that immediately follow their head nouns (116), and non-adjacent when the relative clause is separated from the head noun by one or more words, exemplified in 117. Relative clauses with seven words⁸⁷ or more were coded as long (118); and those with six words or less were coded as regular relative clauses under the factor group of length, as in 119.

116. It's all **the same thing** *we keep doing it.* (OSLA 107. IIb. 77)

117. We had **these huge tanks** under the ground *that we would fill it with water.* (EL2 RF. Ia. 156)

118. I was something Ø *you know, you can go home and watch it on 11 o'clock news.* (OSLA 278. Ia. 83)

119. Yea, I know everybody Ø *you name them.* (OSLA 107. Ia. 24)

Moreover, in order to find the effects of the position of the relative clause, sentence final relative clauses were coded as final and others were coded as sentence-medial.

⁸⁷ I chose to use six words as the average number of words in a sentence. Short term memory is also said to have capacity to memorize 6-7 words/numbers at a time.

6.2.1.3. Semantic factor groups

Definiteness and humanness of the antecedent

Prince (1995), adopting Heim's (1983, cited in Prince) 'filecard' terms, assumes that, semantically speaking, when hearing a relative clause, the hearer is instructed to construct or select the appropriate filecard simply on the basis of the head noun; once the card is constructed or selected, the information in the relative clause is added to it. In restrictive relative clauses, if the head is indefinite, then the relative clause is not required for construction of the filecard (the NP head is hearer-new, so s/he constructs a card) and a resumptive pronoun may occur in the relative clause. When, on the other hand, the head is definite (a filecard exists, the hearer should select the appropriate one), the hearer must wait for the relative clause before selecting the appropriate filecard, thus a resumptive pronoun may not occur in this case. This means that indefinite head nouns may co-occur more with resumptive pronouns than definite ones. I coded all antecedent head nouns of the relative clauses as definite or indefinite to examine the effects of this factor group on resumptive marking, as in 120&121.

120. *The school* Ø I was going to it, it was very famous back then (EL2 RF. Ib. 30)

121. It's *a name* that you can find it in different languages and different religions. (EL2 MA. Ib. 88)

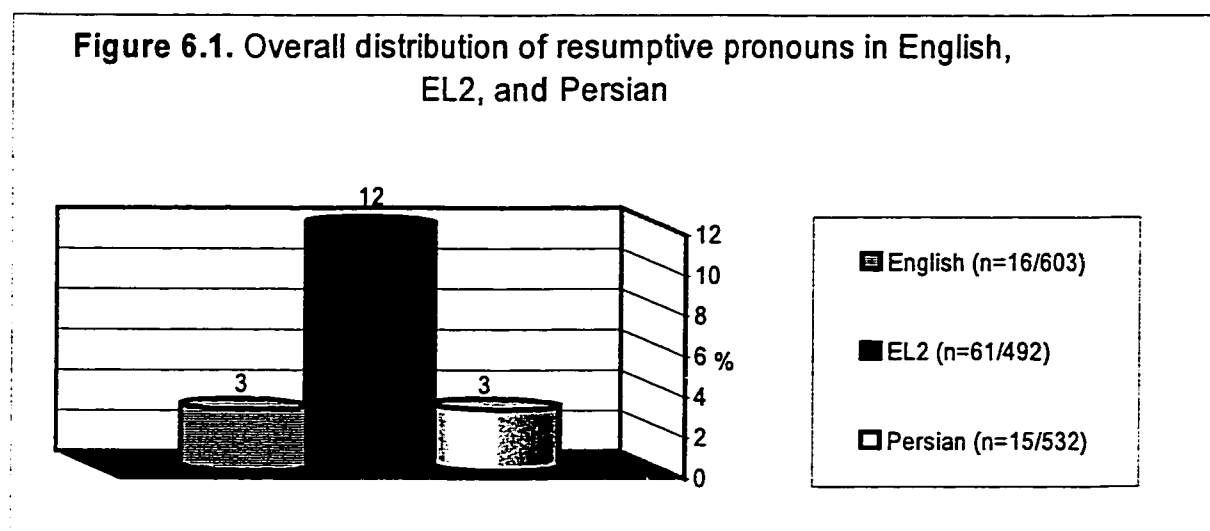
Tarallo & Myhill (1983; also Liceras (1986) for learners of Spanish and Mollica (1981, cited in Liceras) for Portuguese) found that animacy of the antecedent was among the factors that had some influence on the pronominal usage in relative clauses. In order to examine the effects of animacy of the antecedent, I coded head nouns as human (122) and non-human (123).

122. There were *guys* Ø they like to talk to her. (EL2 AK. Ib. 32)

123. They have *these lights* that come out and they shine. (OSLA 241. Ia. 186)

6.2.2. Data

All the instances of restrictive relative clauses produced by the EL2 informants in the three sets of corpora (Persian, English, and EL2) are used in this part of the study. The overall distribution of the data in each data set is presented in Figure 6.1. As shown in this figure, the EL2 corpus includes far more tokens of resumptive pronouns than the other data sets (see 6.3 for a discussion on this).



Surprisingly, the Persian corpus features only 15 (3%) tokens of resumptive pronouns out of a total of 532 restrictive relative clause constructions. The frequency of their usage is vanishingly small in the corpus, contrary to the assumptions made in the literature (e.g., Sadighi 1982; Schachter 1992) with respect to the popularity of these constructions in Persian. There might be two reasons for this sparse usage of resumptive pronouns in Persian. Of the contexts obligatorily requiring resumptives, object of preposition is the one that has been produced the most: only 60 tokens, of which 11 are marked with resumptives. Moreover, direct objects and subjective relative clauses (optional contexts) do not favour resumptives at all (3/197 and 1/226 respectively). It may be argued that since obligatory contexts are not frequently used by Persians, and optional contexts do not favour resumptive usage, there are consequently few resumptives in the data. A second explanation is that since traditional grammarians (this tradition is still followed by many Persian linguists) rely on written Persian in their

description of the language, their account of resumptive pronouns reflects their finding on written Persian where resumptives might have been frequently used. This explanation needs further empirical research on Persian texts.

English, on the other hand, a language without resumptive pronouns, has 16 tokens of them. Is variable usage of resumptive pronouns (although not frequent) conditioned by any linguistic factors in English and Persian? What pattern/s do EL2 learners follow in their second language speech?

In what follows I will attempt to answer these and other related questions. First I will examine the effects of factors incorporated in the study on resumptive marking in each individual context: English, Persian and EL2. I will then compare and contrast the behaviour of resumptive pronouns in EL2 to those of English and Persian.

6.2.3. Results

6.2.3.1. English

The results of the variable rule analysis are shown in Table 6.1. In several initial analyses, some interaction was found between factor groups: e.g., human head nouns favoured relatives with subjective function (127/216) and non-human head nouns favoured relatives as direct object of the relative clause (168/199). But further analyses, with or without interacting factor groups or with interacting factor groups combined, indicated that the results reported in Table 6.1 show the independent effect of the length of the relative clause as the only significant contributing factor to the possibility of marking relative clauses with a resumptive pronoun in English.

Table 6.1. *Contribution of factors selected as significant to resumptive marking in English.*

Input: .022		Significance: .216	
Total: 603		Log Likelihood: -70.571	
Length of relative clause	prob.	%	N
Seven words or more	.74	6	8/129
Six words or less	.43	2	8/474

Factors not selected: Definiteness, premodification, adjacency, humanness, and the syntactic function in matrix or relative clauses.

The finding here confirms that rather less frequent variable use of resumptive pronouns in English is constrained by one factor group. Length of the relative clause as a formal factor is the only significant factor contributing to the presence of these pronouns. Whenever the relative clause continues for seven words or more, speakers insert a resumptive pronoun in the clause to help their interlocutor understand the meaning the clause conveys.

6.2.3.2. Persian

Unlike prescriptive grammars that assume that all antecedent head nouns of Persian restrictive relative clauses are obligatorily followed by a demonstrative *-i* (Windfuhr 1979), I found that this item is used variably in relative clause constructions⁸⁸. In order to test the effects of the presence or absence of demonstrative *-i* on resumptive marking⁸⁹, a new factor group was added to the list of Persian factor groups. When the demonstrative *-i* was present, it was less likely to have a resumptive pronoun in the relative clause (6%), while in its absence the percentage was increased to 9%. As I found that the presence or absence of this item does not significantly affect resumptive marking in Persian. I will not discuss it further.

The results of the variable rule analysis are reported in Table 6.2. Since subjective relative clauses were categorically lacking resumptive pronouns, they were excluded from the analysis⁹⁰. What we understand from these results is that resumptive marking is an established part of the grammatical system of Persian; it is strongly conditioned by only syntactic constraints in the language. This finding is also on par with the

⁸⁸ This discrepancy between grammatical accounts and actual usage of Persian (as for resumptive pronouns) is due to, I suppose, linguists' lack of attention to language use in its natural contexts.

⁸⁹ Although antecedents with *-i* affix were coded as definite in the definiteness factor groups, here I wanted to examine the individual effect of demonstrative *-i*.

⁹⁰ One instance of relative clause with resumptive pronoun with other syntactic function was also excluded. The variable rule runs indicated interaction between adjacency and other factors, such as the syntactic function of the relative marker in the relative clause (direct objects were mostly adjacent (144 vs. 31 non-adjacent tokens)). Since it was not possible to elide this interaction effect, adjacency was excluded from further analysis. However results (both marginals and several analyses) indicate that, on the one hand, adjacency does not have a significant role in resumptive marking. And, on the other hand, only adjacent direct objective relative clauses slightly favor resumptive marking (2 out of 3 tokens), a finding that is not quantitatively generalizable. Not only did definiteness not have an important effect on the presence of resumptive pronouns, but also definite objective antecedents categorically disfavored resumptives. Definiteness was thus excluded from the analyses as well.

assumptions of the Accessibility Hierarchy: it is first the object of preposition that favors resumptive pronoun marking most, then (we will ignore *other* because there are only a few tokens of this factor in the data) comes direct object; and finally subjective relative clauses with no resumptive pronouns at all.

Table 6.2. *Contribution of factors selected as significant to resumptive marking in Persian.*

Input: .031	Significance: .083
Total: 223	Log Likelihood: -41.006
Function of the relative marker in the relative clause	
Object of preposition	.90 23 11/48
Direct object	.35 2 3/175

Factors not selected: premodification, humanness, syntactic function in matrix clause, and length.

The small number of instances of relative clauses with resumptive pronouns in the corpus makes it difficult to generalise the findings reported here. What Persian speakers do instead of using resumptive pronouns is to appeal to the strategy of ellipsis (Windfuhr 1979) by deleting the whole NP or PP (or related pronominal) in the relative clause, a strategy that Sankoff and Tarallo (1987) refer to as PP-chopping. The question is why do so many SLA researchers insist on the popularity of resumptive pronouns and their effect on the acquisition of target languages by native speakers of Persian? One may argue that the frequency of an item or a construction is not important when it comes to the contact of two or more linguistic systems in a learning situation. What matters is the availability of a system of resumptive marking to the Persian learners of English through their native language. Then, it might be concluded, presence of resumptive marking in the EL2 of Persian speakers indicates the transfer of that system of pronoun marking to the new system they are acquiring, no matter what the rate of frequency of that construction is in the original system. In what follows I will make an effort to empirically describe the system of resumptive marking in the EL2 of Persian speakers. If it is found that the two systems are identical, regardless of the rate of frequency, the interpretation will be in favour of the transfer of the system of Persian resumptive marking to EL2. If the systems

are found to be different, I will interpret the new system based on the quantitative findings of the variable rule analyses.

6.2.3.3. EL2

The results of the variable rule analysis in Table 6.3 show that adjacency, humanness and the function of the antecedent in the matrix clause were selected as significant. Non-adjacent relative clauses are more likely to include resumptive pronouns (.75), while adjacent clauses do not show significant preference for resumptive marking (.48). Human head nouns slightly favour resumptives with .61 probability; and among the

Table 6.3. *Contribution of factors selected as significant to resumptive marking in EL2.*

Input: .108	Significance: .066		
Total: 482	Log Likelihood: -171.216		
Adjacency of the head noun and the relative clause	probability	%	N
Non-adjacent	.75	30	12/40
Adjacent	.48	11	49/442
	Range .27		
Humanness of the head noun			
Human	.61	18	24/133
Non-human	.46	11	37/349
	Range .15		
Syntactic function of the antecedent in the matrix clause			
Object of preposition	.62	16	15/91
Other	.56	14	28/195
Direct object	.53	13	15/113
Subject	.22	4	3/83
	Range .40		

Factors not selected: Definiteness, premodification, syntactic function in the relative clauses, and length.

factors incorporated in the syntactic function of the antecedent factor group, only subjects strongly disfavour resumptive pronouns (.22); other factors do not show important preferences. Cross-tabulations showed that human head nouns functioning as the subject of the matrix clause and subjects that were not adjacent to their relative clauses were never marked with resumptive pronouns. If, in order to elide (or reduce) the interaction

effects, subjective head nouns are excluded, only humanness and adjacency remain as significant factor groups conditioning resumptive marking in EL2.

Another interaction effect existed between humanness and syntactic function of the antecedent. Most human antecedents functioned as subjects (107 vs. 89 non-human subjects) and most non-human antecedents functioned as objects (100 vs. 12 for direct objects). In the next phase of the analyses, to remove any interaction between humanness and function of the head noun, these two factor groups were collapsed together and formed a new factor group. Human subjects categorically disfavored resumptive marking, creating a knock out and were thus excluded. The results of variable rule analysis (Table 6.4) indicate that adjacency is selected as the significant factor group. No further important interaction was found among different factor groups.

Table 6.4. *Contribution of factors selected as significant to resumptive marking in EL2: humanness and function in matrix collapsed.*

Input: .142	Significance: .332		
Total: 377	Log Likelihood: -153.908		
Adjacency of the head noun and the relative clause	probability	%	N
Non-adjacent	.72	29	10/34
Adjacent	.48	13	45/343

Factors not selected: Syntactic function of the relative marker, length, and the new factor group of humanness/syntactic function in the matrix (definiteness and premodification showed no effects; they were thus not considered here).

According to the findings reported in Tables 6.3 and 6.4, when a relative clause is separated from its antecedent it is more likely to be marked with a resumptive pronoun in EL2. Adjacent relative clauses, on the other hand, do not have any significant preference for resumptive insertion.

6.3. Discussion

In this section I discuss the results of the variable rule analyses for the three contexts of English, Persian and EL2 together. Table 6.5 presents the results with the probabilities from the first stepping down level of the analyses for the factor groups that were not selected in one context but selected in the other/s. The first observation is that Persian and English contexts behave totally differently. Not only are the selected

significant factors different but also the constraint hierarchies in selected factor groups are in opposite directions in the two languages, providing ideal contrasting sites to discover what patterns EL2 learners follow.

Table 6.5. *Comparison of contribution of factors selected as significant in English, Persian and EL2.*

	English			EL2			Persian		
Significance:	.216			.332			.083		
Input:	.022			.142			.031		
Total:	603			377			234		
	prob.	%	N	prob.	%	N	prob.	%	N
Length of the relative clause									
Seven words or more	.74	6	8/129	[.53]	19	19/101	[.30]*	4	1/25
Six words or less	.43	2	8/474	[.49]	13	36/276	[.53]	7	13/198
Range	.31								
Function of the relative marker in the relative clause									
Object of preposition	[.44]	2	2/88	[.61]	19	16/85	.90	23	11/48
Other functions	[.40]	2	2/88	[NA]			NA		
Direct object	[.61]	3	7/206	[.58]	15	16/107	.35	2	3/175
Subject	[.46]	2	5/221	[.40]	12	23/185	NA		
Range							.55		
Adjacency of the head noun and the relative clause									
Non-adjacent	[.58]	4	2/51	.72	29	10/34	NA**		
Adjacent	[.49]	3	14/552	.48	13	45/343	NA		
Range				.24					

* Probabilities in square brackets were not selected as significant. **See footnote 90

In English, length of the relative clause is selected as significant with long clauses favouring more resumptive marking (.74). In Persian, not only was this factor group not selected, but also short relative clauses favoured more resumptive marking (with a probability of .54). Persian speakers, unlike English, follow the Accessibility Hierarchy (the only selected factor group) and mark objects of preposition with more resumptives (.90). They do not use resumptive pronouns in subjective relative clauses at all.

Let us see how the EL2 speakers behave compared to Persian and English. In the first factor group, length, the constraint hierarchy for EL2 speakers is the same as for

English speakers, and different from Persian. They prefer to mark longer clauses with resumptive pronouns. Adjacency was the only selected factor group in EL2. EL2 speakers marked non-adjacent relative clauses more with resumptive pronouns, just like English. While in Persian, adjacency was never found to show a significant effect on resumptive marking (see footnote 90). However, the marginal results indicated a slight preference for resumptive pronouns in non-adjacent contexts. Considering the effects of adjacency and length together, what seems to be important in English and EL2 contexts is the distance between the antecedent head noun and its reference in the relative clause. For this reason, adjacency and length are found to assert similar effects on resumptive marking in these contexts. For the syntactic function of the relative marker in the relative clause, both Persian and EL2 learners mark objects of preposition with resumptives more than direct objects. Since in Persian resumptive marking is a grammatical feature of the language, syntactic constraints play the most significant roles. Therefore the higher rate of resumptive marking for EL2 objects of prepositions might be an effect of Persian (transfer), although the tendency of learners to follow the Accessibility Hierarchy may be another possibility. EL2 learners, however, distance themselves from Persian and look similar to English in non-selection of this factor group, effects of subjects in resumptive marking (knocked out in Persian because of their categorical behaviour), and similar effects of direct objects in EL2 and English.

Now, do the results show any traces of language transfer from EL2 speakers' native language to their second language speech? It is obvious, that in some cases second language speakers behave similarly to their native language, with respect to the function of the relative marker as object of a preposition. But, neither are the selected factor groups the same in Persian or EL2, nor are the constraint hierarchies in selected factor groups similar. The preference of the EL2 learners to mark the objects of a preposition with resumptives may also be attributed to the effects of the universals captured in the Accessibility Hierarchy. Therefore, regardless of the effects of a non-selected factor group (function of the relative), the system EL2 learners utilise is different from that of L1⁹¹.

⁹¹ Findings of this chapter may not be generalizable though, because of the scarcity of the data in Persian and English.

Then why are there so many claims about transfer of Persian resumptive marking to Persian speakers' EL2? Reliance on the empirical, variationist method in this study and the results I obtained here⁹² indicate that there might be three disadvantages in the previous studies of the same phenomenon in the SLA research. Firstly, they rely on the prescriptive, often traditional grammatical accounts of the system of the native language of their informants. The Persian corpus with more than 500 tokens of restrictive relative clauses from natural speech (with only 15 resumptive pronouns) evidences that one cannot rely only on the prescriptive grammar of a language to do any reliable study on the structure and system of that language or to compare and contrast it to other linguistic systems. We saw that, contrary to what is prescribed in the prescriptive literature, resumptive pronoun usage is vanishingly rare in Persian. The second disadvantage is their reliance on the use of artificial, mostly grammaticality judgement data in SLA studies. When judging data, or even producing written or oral samples of language in unnatural situations and under psychological pressure of data production or testing circumstances, the speakers appeal to every available means and system to produce the required samples as accurately as possible (see Chapter three). The native language is one of the first and easiest available options for the speakers to judge the grammaticality of the sentences in these situations. In judging the grammaticality of relative clauses with resumptive pronouns, L2 learners' appeal to their L1 may have resulted in the frequent acceptance of those sentences as found in SLA studies so far (Schachter 1983, 1992, for instance)⁹³. The final disadvantage of previous studies is their dependence on only the surface features of second language speech. What is found in this study indicates that the presence or absence of resumptive pronouns in one language, either L1 or L2, shows neither the variability in their usage, nor the underlying system of factors that condition the variation. It is only through the variationist method that one can adequately find the

⁹² Although the quantitative data in our study were sufficient enough to enable us to run the variable rule analysis, it is necessary to emphasize that because of the scarcity of resumptive pronouns in Persian and English contexts, the results obtained in this part of the study may not be generalized to other contexts or other languages.

⁹³ See also Pavesi (1986) where Italian learners of English (both prescriptively non-resumptive languages) who have acquired English under formal circumstances (formal English classes) make more use of the "strategy of pronoun retention, whereas the informal group (acquired English through natural exposure to the language) used it to a considerably lesser extent" (P. 48).

actual sources of variation and the system governing the variable presence or absence of a linguistic construction in a language.

One further point in need of more elaboration is the kind of system that operates in second language speech. We have seen, so far, that patterns of EL2 speech are more similar to those of English, the target language of the informants. But at the same time two different factor groups were selected as significant in the variable rule analyses for English and EL2. Does there exist a specific pattern of EL2, let's say, a sort of interlanguage?

Factor groups selected as significant in English and EL2 are both discourse (formal) factors. They have nothing to do with the syntactic systems of the respective languages, i.e., resumptive marking is not conditioned by syntactic factors. Whenever the relative clause is not adjacent (as in EL2) or is too long (as in English), English or EL2 speakers add a resumptive pronoun to the relative clause to make the reference of the relative marker easier to trace and the meaning of the clause more transparent to understand. What matters in both contexts is the distance between the antecedent and the relativized noun in the relative clause. The learners have a greater need for the resumptive pronoun when the relativization site is farther from the head (see also Tarallo & Myhill, 1983 for this).

Why do the EL2 speakers use more resumptive pronouns than native English speakers? A first possible explanation deals with the effect of the behavior of the individual informants on the data. Of my 12 EL2 informants, three of them produced the largest number of resumptive pronouns, an average of 10 tokens each, about half the instances of resumptives in EL2. They had therefore a remarkable effect on the total number of resumptive pronouns in the EL2 data.

Another argument is that EL2 learners use more resumptive pronouns because, as second language speakers, they appeal to any available option to make themselves as clear as possible to their interlocutors. They always try to avoid "ambiguity" and "obscurity of expression" (Grice, 1975)⁹⁴ and be as informative as possible, so that the interlocutor need not ask them to repeat or clarify themselves. Then they add more

resumptive pronouns in the contexts where there is more distance between the relative marker and its head antecedent. Native speakers also do so when they think there might be a break in the understanding of the interlocutors, as in the case of long relative clauses where they sometimes add resumptive pronouns. But they have enough linguistic knowledge in their L1 to obtain transparency of meaning without often appealing to resumptives. EL2 speakers do not share this knowledge; they therefore insert more resumptives in their relative clauses.

It is also possible to argue for a universalist hypothesis with respect to the insertion of resumptive pronouns in second language speech. There are studies that have found resumptive usage in the speech of children whose L1 does not allow such a phenomenon. Perez-Leroux (1995), for instance, has found that resumptives are present in children's English at the ages of four and five; and Guasti & Shlonsky (1995) have found the same thing for children's French. This needs, however, more systematic studies on data from both native children and other second languages, to empirically demonstrate the effects of universalist assumptions.

Transfer of the Persian system into EL2 is another possible reason for the frequency of resumptive usage in EL2. As mentioned before, it may be claimed that resumptive marking, as a feature regardless of its frequency, is available to Persians, which they apply to their L2. However, the difference between EL2 and Persian in selected factor groups and their constraint hierarchies makes this explanation untenable.

In sum, using variable rule analysis, I was able to determine the constraints on variable use of resumptive marking in English, Persian and EL2. The findings showed that the EL2 spoken by Persian speakers more closely resembled English patterns of variation, constraints on variation, and hierarchy of constraints. We also observed that there does not exist any syntactic system of variable resumptive marking in EL2. These speakers were more sensitive to factors related to discourse than to those of syntax.

⁹⁴ H.P. Grice (1975) proposes four statements as the maxims of the cooperative principle in conversation analysis. They include maxims of Quantity (be as informative as is required), Quality (do not say what you believe to be false), Relation (be relevant), and Manner (avoid obscurity of expression and ambiguity).

6.4. Summary of chapter six

In this chapter we saw that researchers in the field of SLA attributed the existence of resumptive pronouns in the second language speech of Persian speakers to the effect of their native language, and therefore argued for the presence of cross-linguistic influence in this respect. The quantitative study on naturalistic data from native speakers of Persian revealed that resumptive pronouns, although syntactically conditioned, are not frequent in Persian, at least in spoken data. The variable rule analysis on data from English spoken by Persian speakers, on the other hand, demonstrated that resumptive usage in EL2 is conditioned by factors different from those that operate in Persian. The EL2 system was found to be similar to the system of resumptive pronoun usage that is utilized by native speakers of English in their vernacular speech. I concluded, therefore, that variable use of resumptive pronouns in EL2 is conditioned by constraints similar to those that operate in English, i.e., this system of variability has been most likely acquired by L2 learners.

Chapter seven: DISCUSSION OF RESULTS AND CONCLUSIONS

7.1. Findings of the study

To the best of my knowledge, this thesis has been the first to perform a variable rule analysis on the alternate use of contracted forms of auxiliary verbs in samples of speech from learners of English as a second language. It is also the first to detail the variable use of relative and resumptive pronouns in second language speech and apply the comparative method to compare the behavior of L2 learners in L2 with their own behavior in their L1 and that of the native speakers of the target language. The groups of the L2 informants (EL2 & LEL2) are also unique in that, in each group, the informants share the same native language and have similar amounts of exposure to the target language, similar instructional background in L2, and similar educational level. Moreover, this study was the first to investigate variable use of auxiliary verbs, relative and resumptive pronouns in Persian.

The evidence I have presented from L2 learners of the same L1 background, quantitatively establishes that in the case of three different linguistic variables (auxiliary verbs (itself including different forms), relative markers (*THAT*, *ZERO*, *WH-*), and resumptive pronouns) variation in L2 is in fact constrained by linguistic environment (phonological, syntactic, semantic or formal factors). This, by itself, demonstrates that second language speech is systematic (disconfirming hypothesis one (chapter one)). The kind of system we observed is not the categorical system of prescriptive reference grammars but that of systematic variation similar to what has been found in native languages and urban dialects which have been investigated by sociolinguists (see also Young 1991). The findings of the present study therefore do not lend support to the notion of non-systematic variability. We saw that at both the early and later stages of acquisition of English auxiliary verbs by native speakers of Persian, variation was far from being the haphazard product of competing rules, as Ellis (1985 and others, in support of non-systematic variation) has claimed. On the contrary, the variable rule analyses indicated that variable contraction of *IS* was conditioned by both syntactic and phonological constraints in EL2 (preceding subject and phonological segment and following grammatical category) and LEL2 (preceding subject and phonological

segment). Following grammatical category (in EL2) and preceding grammatical and phonological factors (in LEL2) were of the most significance in conditioning variable contraction of *ARE*. In the case of other auxiliary verbs where variable rule analysis was not performed, effects of the linguistic factors were clearly shown by the distributional results (marginals). Therefore, Ellis's claim that variation is not systematic during the early stages of acquisition of a form is not supported by the findings of this study.

The present study has found, however, that factors influencing variation change as acquisition of a form proceeds (at two stages of proficiency). At the early stages of acquisition (LEL2) influence of phonological factors is more prominent than syntactic factors, shown in Table 7.1. As learners become more proficient, the effect of phonological environment recedes such that in some contexts (e.g., *ARE*-contraction) for higher proficiency learners it no longer shows a significant effect⁹⁵. This finding is on par with what has been found by Young (1991) in his study of plural marking patterns in second language speech.

Table 7.1. *Contribution of factors selected as significant to the probability of contraction of IS and ARE: low and high proficiency levels compared.*

Factor groups	<i>IS</i>		<i>ARE</i>	
	LEL2	EL2	LEL2	EL2
Subject type	√	√	√	---
Preceding phonological environment	√	√	√	---
Following grammatical category	---	√	---	√

Also on a par with the findings of other studies (Regan 1996), my results indicate that in cases where second language learners have acquired the patterns of variation from their target language, there is little change in the patterns of variation in L2 as acquisition proceeds. The only change observed in these cases is in the rates of rule application, not in patterns of variation, as was demonstrated by factor groups found to affect contraction of *AM* (and to some extent *IS*) in English, EL2 and LEL2. After acquisition of the target language patterns, whatever the rate of contraction, L2 learners apply the system that they

have acquired to their L2 data, as found for contraction of *ARE* and (to some extent) forms of *HAS* in EL2. Generally speaking, before acquisition of the patterns of variability, syntactic rules of the target language are either applied categorically by L2 learners (as for *HAD* in LEL2) or the native language system of variation affects L2 variability (as for *ARE* in LEL2). Acquisition of the variable patterns occurs first in contexts where native and target languages are either identical or totally different. In cases where native and target language are similar on the surface but differ in constraint hierarchy, L2 learners face difficulty in the acquisition of the target rules. After acquisition of the variable patterns, it is only the target language system of variation that operates in second language speech.

Learners' native language played a trivial role in affecting the system of variation in high proficiency second language speech, while its effects were seen more in the speech of the lower proficiency group. At higher proficiency level, we did observe that in some contexts where an L1 factor group exerted a significant effect on L1 variation, that factor group left a trace on EL2 variation as well (e.g., the effect of the syntactic function of relative marker on *THAT*). L1 influence on LEL2 variation appears in contexts where less proficient L2 learners have acquired the syntactic knowledge of the target language rules (i.e., the rules of grammar) but not their patterns of variation yet, such as contraction of *ARE* and *HAS* (to some extent *HAVE*). In these cases, LEL2 patterns of contraction follow L1 in some contexts (mostly phonological) and fluctuate between L1 and English patterns in some other contexts (mostly grammatical⁹⁶). It is interesting to note that in these contexts LEL2 learners follow the English system for some factors and behave differently in some others. This indicates that L1 influence is not a matter of complete influence or the lack of it. L1 exerts a sort of delaying effect and delays acquisition of the target language patterns as Zobl (1980, for categorical, not variable rules) has found.

⁹⁵ I am dealing only with the selection/non-selection of factor groups, not the behavior of factors or their hierarchy.

⁹⁶ Grammatical factors are of more importance since they indicate if a certain pattern is a part of learners' grammatical repertoire or not, which in our case indicate that LEL2 are yet to completely learn patterns of variation from English.

In sum, as expected (as in Taylor 1975), an effect of native language patterns of variation on second language speech was shown mostly at the lower level of proficiency. What is found is that in some contexts (mostly grammatical) the effect of a certain factor group in L2 parallels that of L1, indicating first language transfer. As a result of this L1 influence, acquisition of the target language patterns is delayed. This kind of transfer, however, appears not in the presence/absence or frequency rate of certain linguistic forms but in the conditioning factors and the hierarchical effect of those factors that constrain variation. In advanced second language speech, the effect of L1 patterns of variation on EL2 was not very significant. Why then, have previous studies (particularly on relative and resumptive pronouns) attributed L2 learners' 'errors' to transfer of the native language system?

Mention was made that variationist studies, unlike traditional studies of SLA (and linguistics), deal not only with the presence and absence of linguistic forms, but their frequencies and the systematic conditioning of these frequencies by co-occurring linguistic (and extra-linguistic) environments (Sankoff 1988a). Throughout my analyses, therefore, I distinguished between rates and patterns of variation. If I paid attention to only the rates of presence/absence of the features studied here, like other studies, I could argue for the crosslinguistic influence of Persian on both EL2 and LEL2 in almost all contexts. It was only through discovering the patterns of variation that I could demonstrate the underlying system of variation in the grammatical repertoire of the language speakers. Since previous research relied mostly on the presence or absence of the linguistic forms in L2 speech or, at most, their frequency rates, these studies have assumed an important effect for L1 features on L2 speech. The focus of the present study, unlike others, was on the patterns of variation.

Therefore, following Labov (1996), I believe that by comparison with the other approaches, contrastive analysis "remains the most promising route for those who want to apply a concrete knowledge of language structure to the problem of language learning" (p. 246). And what I have done in this dissertation is also a kind of contrastive analysis, but with a remarkable difference from traditional contrastive studies. In my approach, I contrast and compare the overwhelming variability and its systematic conditioning in second language speech, while in the traditional contrastive analyses, what was

contrasted and compared was the surface category inventories of the languages in contact. In the traditional studies the second language learners' 'errors' were argued to be due to the influence of the learners' first language (transfer) or an interlanguage system specific to L2 learners. In my approach 'errors' are dealt with not as deviations from the target language norms but as variants of a linguistic feature to express the same function. The origin of this variability was found to be (in many cases, in EL2 in particular) the target language of the learners. This demonstrated that what has been dealt with as 'error', a sign of the lack of acquisition of the target language rules, may be an indication of native-like acquisition. These results quantitatively showed that surface presence of linguistic structures is in no way a reliable criterion of acquisition of those structures, nor their language of origin (native or target).

Another issue that the present study attempted to tackle quantitatively was the issue of the acquisition of the variable patterns. The findings of this study showed that many of my second language learners (EL2 in particular) produced all linguistic features (studied in this thesis) according to the syntactic rules of the target language. But from a native speaker point of view their speech in no way parallels the target language vernacular in some contexts, such as the use of *WOULD*, where there is the least amount of variation in EL2. For English speakers this sort of L2 speech is still a 'foreigner talk', because, in addition to the syntactic correctness of the language structures, native-like use involves application of the patterns of variability that are used by the native language speakers in their everyday speech. I argued, thus, that native-like acquisition involves acquisition of the system of variation that operates in the target language and that I called real L2 acquisition. But can variability be acquired?

For those who do not consider language variation as a part of linguistic inquiry, language acquisition is only acquisition of categorical rules of grammar and they believe that variability cannot be acquired (e.g., Gregg 1989). On the contrary, the findings of this dissertation showed that in many contexts, EL2 learners have acquired not only the grammatical properties of the target language forms (grammatical knowledge) but they have also mastered the patterns of variation that operate in target vernacular speech. Regardless of the differences in the frequency rates, patterns of variation in EL2 mirrored English variability and its conditioning factors in many cases.

Mastery of the syntactic rules of the target language by many of the less proficient informants and, at the same time, problems of the proficient learners in applying English patterns of variation in some contexts and their lower rate of frequency, indicate that it is not the acquisition of rules of grammar (syntax) that poses a problem for L2 learners, but that the most important challenge for them is the acquisition of the patterns of variation. This may be why so many L2 learners sound “foreign” while their knowledge of target language grammatical rules (even their pronunciation) is in no way different from native speakers (e.g., many non-native ESL teachers abroad). Once the patterns of variation are acquired, they are applied to the data, whatever the rate of occurrence of the linguistic features.

Of the target language patterns, those that are totally similar to or totally different from the patterns of variation in learners’ native language are the easier patterns to be picked up. Complex patterns, i.e., patterns that are significant in both native and target languages but whose internal constraint hierarchies are different (or according to Myhill (1982) areas that are similar but not identical), are the most likely areas for learning difficulties. This is where interference problems are more likely to appear and in these cases L1 patterns play a delaying role in L2 acquisition. In other words, surface similarity but internal complexity confuses the language learners; therefore, it takes more time and energy for the L2 learners to distinguish this discrepancy and discover the patterns that operate in the target language. Appeal to L1 features (transfer) was mostly observed in these cases (e.g., for *ARE* in LEL2 or *THAT* in EL2).

A major goal of this dissertation was to find the sources of variability in second language speech. My findings revealed that English spoken by Persians indeed variably employs contracted forms of auxiliary verbs, variably chooses among three relative markers and variably marks relative clauses with resumptive pronouns. I was also able to find the linguistic factors that significantly conditioned learners’ choice processes. In order to find the origin of patterns of variability in L2, I compared these findings with the systems of variability I found in informants’ native and target languages. What the comparative method revealed is that patterns of variability used by the advanced EL2 learners mirror those of the target language in most of the cases studies here, i.e., the system of variability in EL2 is similar to the one that is used by native speakers of

English, with some differences in the rates but not in the patterns (hypothesis two). In only a few contexts there appeared to be some (not significant) influence from the patterns of the informants' native language. In less proficient L2 speech, too, the patterns found were not very much different from English, but in some contexts there were parallels between LEL2 and Persian. In these cases I concluded that L1 patterns of variability affected second language speech both in rates and patterns of variation (hypothesis three).

These findings emphasize that patterns of variation in second language speech have originated either from L1 (at lower proficiency level) or the target language. We did not observe any patterns specific to L2 or created by second language learners (disconfirming hypothesis four). Therefore, in the contexts studied here, acquisition from English and less probably transfer of L1 patterns are the only sources of variation in the informants' L2 speech. Interlanguage defined as a natural language with its own specific internal patterns of variation was not found in this study. But if interlanguage is defined as a linguistic system with systematic variability, findings of this study demonstrated presence of such a linguistic system.

And finally with respect to the method applied in this study, the findings indicate that:

- Application of the variationist methods can empirically demonstrate the systematic nature of second language speech.
- Application of the comparative method to the data in this thesis indicates that surface similarities of structures and their frequency patterns are of little help in discovering the underlying systems of language used by L2 speakers. Systematic patterns of variation and their sources are only found by application of variable rule analysis to spontaneous data extracted from speakers of native, L2 and target languages.
- It is only through the application of this method that one can understand the effects of native and target languages on L2 speech and reveal the underlying system of variation in it.

7.2. Limitations of the study

Notwithstanding the fact that the present study investigated variable use of three linguistic features of second language speech, it has been limited to the L2 speech of

learners from one linguistic background. It is not known how far the findings of this dissertation may be generalized to the second language speech of learners from other first languages and also learners of second languages other than English. Although I have spent much time on the comparison of patterns of variation in L2 with those of native and target languages, the only way to be certain that what has been found here is a general phenomenon in SLA is by carrying out further similar studies and applying the comparative method to other languages in a learning situation.

In some variable contexts, such as *HAS/HAVE* auxiliary verbs and resumptive pronouns, analyses were performed with a limited number of tokens. Although sparse use of some linguistic forms is by itself meaningful, as was the case for resumptive usage in Persian, having a higher quantity of tokens with detailed analysis devoted to the behavior of each of those variables may reveal more about the underlying grammar of the L2 learners in this respect.

There is hardly any variationist study on the overwhelming phonological, syntactic, etc. variation in spoken Persian. Moreover, in some cases there does not exist consensus on some grammatical structures of the language among different syntacticians, modern or traditional. I mostly relied on prescriptive accounts of the grammatical features dealt with in this dissertation, in coding procedures in particular. I tried my best to uncover the true factors constraining variability in Persian; however, the factor groups for Persian were very limited, consisting mostly of those used for English. Further variationist studies are needed to describe the actual behavior of native speakers of Persian, which, in turn, will be of great help in defining the variable contexts and in coding for various linguistic environments in comparative studies of language acquisition.

APPENDIX A

Interview protocols

=====

The following interview protocols have been used as guideline in conducting the interviews. Whenever appropriate, other topics of interest and questions were introduced to make the interview situation as natural as possible.

1. Community/neighborhood

When did you move to this place? Why did you move? How is it? Did you have a rent increase? Why are they not controlling rents anymore? How is the neighborhood? Do you know any of the neighbors?

2. Family

Do you have any news from your parents? How old are they? What about your brothers and sisters? Do you often call them or do they call you? Have your parent been here for a visit? What is your relation with them? How was it when you were a kid? Did they ever punish you? What about you, do you punish your kids (if the informant had any)? What do you think about it? Is it not difficult to raise kids here? How old are they?

3. Socializing

Do you have any relatives in Canada? Is it good to have some? Do you have any friends that you socialize with? Any Canadian friends? Is it easy to find one? Why (why not)? Why is the relation between friends so much different in different cultures? Are there any other Iranians in your department/office? How is your relation with them?

4. School

What school are you in? How come you chose this school? How did you get admission? Did you know your supervisor before? How was the first day of school here? Could you speak English? Who was the first person you saw? How was the first class you attended? Could you understand it? What did you do with your English? How was your English

before coming to Canada? Where did you learn it? Do you remember your first English teacher? Who was it? What did you think about learning another language? What high school did you attend? Where was it? Do you remember your math teacher? Was he tough? Who was the toughest of all? Who was the best that you can remember now? Why? Who was the worst? Was he tough? Real creep? Did he ever blame or punish you? How was your behavior at school? Did you have any groups? Were you a member? Did you have any friends that you can remember now? Are you still in touch with them? Who was the best of your friends? Who is a good friend? Do you remember any friends from primary school? Do you remember the first day you went to school? Who was the grade one teacher? You remember him/her?

5. Fights

Did you have any fights at school? Do you remember any? What was it about? How did it start? Who started it? Then what happened? What was the best (the worst, the most important) of your fights? Did you ever have fights with bigger guys? Did you ever have fights at school? What happened? Was anybody expelled from the school? How did the other kids react?

6. Danger of death/fear

Have you ever been in a situation where you were in serious danger of getting killed, when you said to yourself 'this is it'? What was it? What did the others do? Did you go to war fronts when there was a war? For how long? What was your feeling? Were you ever in danger there? Were you afraid? What is it to be really afraid? Where did it happen to you?

7. Marriage

(If married) When did it happen? Do you remember the first time you saw your wife? Where and how was it? Who was the first to offer marriage? Do you remember your engagement (wedding)? Did you live with your in-laws after marriage? Where did you go

then? Did they know that you might go abroad? How is your relation with your wife/husband? How do you solve your problems (if any)? What is the best way to deal with them?

8. Work

Were you employed when you got married? What was your first job? Did your wife work? Should women stay home or work after marriage? How did you find it? What are you doing for a living now? What are you working on at school? What is it about? Does it have any application?

9. Sports/games

Do you play any sports? How often? With whom? What is the most popular sport in Iran? Why? What do you think about the world cup? How will (did) the Iranian team play? What team do you think will be the world champion? What other sports do you play/like? Where did you learn it? Have you ever had any accident when playing it? What game did you play when you were a kid? How did you play it? Do your kids know these games? Do you play with them? How do you do it?

10. Dreams

Do you remember doing sports in your dreams? Are they in color? What is a true dream? Did you have any scary dreams, waking up in a sweat? Do they mean anything to you? Have you ever seen the future in your dreams? Is it possible?

11. Language

Do you speak any other languages? What language do you use when talking to your kids? What about your wife? Are there many Iranians in your neighborhood, in your department? What language do you use when talking to them? Have you recently read any Persian books? What was the last book you read? Have you recently read a novel? What was it about? What is the story? Did you like it? Why?

APPENDIX B

Relative clauses taught to Iranian students (Krohn, R. 1998)

Lesson 17

(This lesson corresponds to lesson XVI in the third editions of *English Sentence Patterns* and *English Pattern practice*)

- A. Relative clauses
 B. *for, during when. while before, after, until*

A.1 Notice the relative clauses.



John
Mary
Mr. Miller

Previous patterns (Lesson 16):

S John sees Mary.
 You know *who* sees Mary.

O Mary sees Mr. Miller.
 You know *who* Mary sees.

New patterns:

S I'm pointing to **THE STUDENT**. **THE STUDENT** sees Mary.
 I'm pointing to the student that sees Mary.
 (who)

O I'm pointing to **THE MAN**. Mary sees **THE MAN**.
 I'm pointing to the man that Mary sees.
 whom
 who

COMMENTS

In the S pattern, the *wh*-word is the subject of the following verb. In the O pattern, the *wh*-word is an object (or sometimes an adverbial).

S. that sees Mary
SUBJECT
O. that Mary sees.
OBJECT

2. In both patterns the verb (or auxiliary) comes immediately after the subject.

S. that sees Mary
SUBJECT VERB
O. that Mary sees
SUBJECT VERB
S. that can see Mary
SUBJECT AUX.
O. that Mary can see.
SUBJECT AUX.

3. Relative clauses always have the SUBJECT-VERB (or SUBJECT-AUXILIARY) word order, even in questions:

S. Are you pointing to the student that sees Mary?
SUBJECT VERB
O. Are you pointing to the man that Mary sees?
SUBJECT VERB

NOTES:

- (1) *Who* and *whom* are used for persons.
Which is used for things and animals.
That is used for persons, things, and animals
- (2) In Pattern S, *that* is more common for things and animals than *which* is. Both *who* and *that* are very common for persons. However, Sometimes *who* is preferred.
- (3) In Pattern O, the relative pronoun may be omitted entirely.
I'm pointing to the student that Mary sees
I'm pointing to the student Mary sees.

Bibliography

- Adamson, H. D. 1988. *Variation theory and Second Language Acquisition*. Washington , D.C.: Georgetown University Press.
- Adamson, H. D. 1992. Social and processing constraints on relative clauses. *American Speech* 67: 123-133.
- Adamson, H. D., B. Fonseca-Greber, K. Kataoka, V. Scardino, & S. Takano. 1996. Tense marking in the English of Spanish-speaking adolescents. In Bayley & Preston (eds.): 121-134.
- Adamson, H. D. and Elliott, P. 1997. Sources of variation in interlanguage. *International Review Of Applied Linguistics* 35: 87-45.
- Adamson, H. D. and C. Kovac. 1981. Variation theory and second language acquisition: An analysis of Schumann's data. In Sankoff & Cedergren 1981: 285-292.
- Adamson, H. D. & V. Regan. 1991. The acquisition of community speech norms by Asian immigrants learning English as a second language: A preliminary study. *Studies in Second language Acquisition* 13, 1: 1-22
- Adjemian, Christian. 1976. On the nature of interlanguage systems. *Language Learning* 26: 297-320.
- Andersen, Roger W. 1983. Transfer to somewhere. In Gass & Selinker 1983: 177-201.
- Andersen, Roger W. (ed.). 1984. *Second Languages: A crosslinguistic perspective*. Rowley, Mass.: Newbury House.
- Andersen, R. 1989. The theoretical status of variation in interlanguage development. In Gass, S. Et al. (eds.). 1989.
- Asher, R.E. 1994. *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press.
- Bailey, C-J. N. 1973. *Variation and Linguistic Theory*. Washington, DC: Center for Applied Linguistics.
- Bailey, N., C. Madden, and S. Krashen. 1974. Is there a "natural sequence" in adult second language learning? *Language Learning* 21: 235-243.
- Baker, C. L. 1978. *Introduction to Generative-Transformational Syntax*. New York: Prentice-Hall.
- Ball, C. 1996. A diachronic study of relative markers in spoken and written English. *Language Variation and Change* 8, 2: 227-258.
- Barasch, Ronald M. and C. V. James (eds.). 1994. *Beyond the Monitor Model: comments on current theory and practice in second language acquisition*. Boston, Mass.: Heinle & Heinle Publishers.
- Baugh, J. 1980. A reexamination of the Black English copula. In Labov (ed.). *Locating Language in Time and Space*. New York: Academic.
- Bayley, Robert. 1994. Interlanguage variation and the quantitative paradigm: past tense marking in Chinese-English. In Tarone et al. (eds.). 1994: 157-181.
- Bayley, Robert. 1996. Competing constraints on variation in the speech of adult Chinese learners of English. In Bayley & Preston (eds.): 97-120.
- Bayley, Robert and Dennis R. Preston (eds.). 1996. *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins.

- Beebe, L. M. 1977. The Influence of the listener on code-switching. *Language Learning* 27: 331-339.
- Beebe, L. M. (ed.). 1988. *Issues in Second language Acquisition: Multiple perspectives*. New York: Newbury House.
- Beebe, L. M. and J. Zuengler. 1983. Accommodation theory: An explanation for style shifting in second language dialects. In Wolfson and Judd. 1983: 195-213.
- Bialystok, E. 1982. On the relationship between knowing and using linguistic forms. *Applied Linguistics* 3, 3: 181-206.
- Biesenbach-Lucas, S. 1987. The use of relative markers in modern American English. In K. Denning, S. Inkelas, F.C. McNair-Knox & J. Rickford (eds.). *Variation in Language: NWAVE XV. at Stanford*. Stanford, CA: Stanford U, Dept. of Linguistics: 13-21.
- Birdsong, David. 1989. Lux et Verisimilitudo: Judgment data in second language acquisition theory. Paper presented at the Fourteenth Annual Boston University Conference on Language Development. Boston, MA.
- Birdsong, David. 1991. On the notion of "Critical Period" in UG/L2 Theory: A response to Flynn and Manuel. In Eubank 1991: 147-166.
- Blackshire-Belay, Carol A. (ed.). 1994. *Current Issues in Second Language Acquisition and Development*. Lanham, Maryland: University Press of America.
- Blake, R. 1997. Defining the envelope of linguistic variation: The case of "don't count" forms in the copula analysis of African American Vernacular English. *Language Variation and Change* 9, 1; 57-79.
- Brown, Gillian, K. Malmkjaer, and J Williams (eds.). 1996. *Performance and Competence in Second Language Acquisition*. Cambridge: Cambridge university Press.
- Brown R. 1973. *A First Language: The early stages*. Cambridge, MA: Harvard University Press.
- Butterworth, G. & E. Hatch. 1978. A Spanish-speaking adolescent's acquisition of English syntax. In Hatch (ed.): 231-245.
- Cancino, H., E. J. Rosansky & J. H. Schumann. 1974. Testing hypothesis about second language acquisition: the copula and negative in three subjects. *Working Papers on Bilingualism* 3: 80-96.
- Cancino, H., E. J. Rosansky & J. H. Schumann. 1975. The acquisition of the English auxiliary by native Spanish speakers. *TESOL Quarterly* 9,4: 421-430.
- Chaudron, C. 1983. Research on metalinguistic judgments: a review of theory, methods, and results. *Language Learning* 33, 3: 343-377.
- Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT.
- Christie, K. & Lantolf, J. 1991. The ontological status of learner grammaticality judgments in UG approaches to L2 acquisition. Paper presented at Second Language Research Forum, Los Angeles.
- Cook V. J. 1988. *Chomsky's Universal Grammar: An introduction*. Oxford: Blackwell.
- Cook, V.J. 1993. *Linguistics and Second Language Acquisition*. London: Macmillan Press.
- Cook, V. J. 1996. Competence and multi-competence. In Brown et al. 1996: 54-69.
- Corder, S. Pit. 1967. The significance of learners' errors. *International review of Applied Linguistics* 5: 161-9.
- Cowan, R. & Y. Abe Hatasa. 1994. Investigating the validity and reliability of native speaker and second-language learner judgments about sentences. In Tarone et al: 287-302.
- Crume, G. O. 1931. *A Grammar of the English Language II: Syntax*. Essex: Verbatim.

- Davies, A., C. Crier and A. Howatt (Eds.). 1984. *Interlanguage*. Edinburgh: Edinburgh University Press.
- Dechert, H. W., Bruggemeier, M & D. Futterer. 1984. *Transfer and Interference: A selected bibliography*. Amsterdam: Benjamins.
- Dechert, Hans W. and Manfred Raupach (Eds.). 1989. *Transfer in Language Production*. Norwood, NJ: Ablex Publishing Corporation.
- Dickerson, L. J. 1975a. *Internal and External Patterning of Phonological variability in the Speech of Japanese Learners of English: Toward a theory of second language acquisition*. Ph. D. dissertation. University of Illinois at Urbana-Champaign.
- Dickerson, L. J. 1975b. The Learner's interlanguage as a system of variable rules. *TESOL Quarterly* 9: 401-407.
- Dickerson, L. J. and W. B. Dickerson. 1977. Interlanguage phonology: current research and future directions. In Corder and Roulet. 1977. *Actes du 5eme Colloque de Linguistique Appliquee de Neuchatel: The notions of simplification, interlanguages and pidgins and their relation to second language acquisition*. Geneva: Droz.
- Dulay, Heidi and Marina K. Burt. 1972. Goofing: an indicator of children's second language learning strategies. *Language Learning* 22: 234-52 (also in Gass & Selinker 1983: 54-68).
- Dulay, Heidi and Marina K. Burt. 1973. Should we teach children syntax? *Language Learning* 23: 245-258.
- Dulay, H., M. Burt, and S. Krashen. 1982. *Language Two*. New York: Oxford University Press.
- Eckman, F. 1977. Markedness and the Contrastive Analysis Hypothesis. *Language Learning* 27: 315-330.
- Eckman, F. 1994. The competence-performance issue in second-language acquisition theory: A debate. In Tarone et al. (eds.). 1994: 3-15.
- Eckman, F., L. H. Bell, and D. Nelson (eds.). 1984. *Universals of second language acquisition*. Rowley, MA: Newbury House.
- Eckman, F., D. Highland, P.W. Lee, J. Mileham, and R.R. Weber (eds.). 1995. *Second Language Acquisition Theory and Pedagogy*. Mahwah, NJ: Lawrence Erlbaum.
- Edwards, W. F. and D. Winford (eds.). 1991. *Verb Phrase Patterns in Black English and Creoles*. Detroit: Wayne State University Press.
- Ellis, Rod. 1985. Sources of variability in interlanguage. *Applied Linguistics* 6: 118-131.
- Ellis, Rod. 1987a. Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition* 9: 1-20.
- Ellis, Rod (ed.). 1987b. *Second Language Acquisition in Context*. London: Prentice Hall International.
- Ellis, Rod. 1988. The effects of linguistic environment on the second language acquisition of grammatical rules. *Applied Linguistics* 9: 257-274.
- Ellis, Rod. 1990. Grammaticality judgments and learner variability. In Burmeister & Rounds (eds.). *Variability in Second Language Acquisition: Proceedings of the Tenth Meeting of the Second Language Research Forum*. Eugene, OR: University of Oregon: 25-60.
- Ellis, Rod. 1991. Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition* 13,2: 161-186.
- Ellis, Rod. 1994a. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, Rod. 1994b. Variability and the natural order hypothesis. In Barasch & James. 1994: 139-158.

- Eubank, Lynn. 1991. *Point Counterpoint: Universal Grammar in the Second Language*. Amsterdam: John Benjamins.
- Fairbanks, K. 1982. Variability in interlanguage. Unpublished manuscript, University of Minnesota.
- Fasold, R. 1984. Variation theory and language learning. In Trudgill. 1984.
- Fasold, R. & Y. Nakano. 1996. Contraction and deletion in vernacular Black English: Creole history and relationship to Euro-American English. In G.R. Guy, C. Feagian, D. Schiffrin & J. Baugh (eds.). *Towards a Social Science of Language, Vol. 1: Variation and Change in Language and Society*. Amsterdam: Benjamins: 373-395.
- Ferguson, C. A. 1968. Absence of copula and the notion of simplicity: A study of normal speech, baby talk, foreigner talk and pidgins. Microfiche No. ED030844.
- Fisiak, J. (ed.). 1981. *Contrastive Linguistics and the Language Teacher*. Oxford: Pergamon.
- Flanigan, B. O. 1995. Anaphora and relativization in child second language acquisition. In *Studies in Second Language Acquisition* 17,3: 331-351.
- Flanigan, B. O. & E. Inal. 1996. Objective relative pronoun use in native and non-native English: a variable rule analysis. *Language Variation & Change* 8, 2: 203-226.
- Flynn, Suzanne. 1989. The role of the head-initial/head-final parameter in the acquisition of English relative clauses by adult Spanish and Japanese speakers. In Gass & Schachter 1989: 89-108.
- Flynn, Suzanne & Manuel, S. 1991. Age-dependent effects in language acquisition: An evaluation of "Critical Period" hypothesis. In Eubank: 117-146.
- Fox, B. A. & Thompson S.A. 1990. A discourse explanation of the grammar of relative clauses in English conversation. *Language* 66, 2: 297-316.
- Gass, Susan M. 1979. Language transfer and universal grammatical relations. *Language Learning* 29: 327-344. (also in Gass & Selinker 1983)
- Gass, Susan M. 1983a. The development of L2 intuitions. *TESOL Quarterly* 17: 273-291.
- Gass, Susan M. 1996. Second language acquisition and linguistic theory: the role of language transfer. In Richie & Bahatia: 317-348.
- Gass, Susan M. and Josh Ard. 1980. L2 Data: their relevance for language universals. *TESOL Quarterly* 14: 443-452.
- Gass, Susan and Larry Selinker (eds). 1983. *Language Transfer in Language Learning*. Rowley, Mass: Newbury House Publishers.
- Gass, Susan and Larry Selinker (eds). 1992. *Language Transfer in Language Learning*. Amsterdam: John Benjamins.
- Gass, Susan and Larry Selinker (eds). 1994. *Second Language Acquisition: An introductory course*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gass, S., C. Madden, D. Preston, and L. Selinker (eds.). 1989. *Variation in Second Language Acquisition: Psycholinguistic Issues*. Clevedon: Multilingual Matters LTD.
- Gass, Susan M. and J. Schachter (eds.). 1989. *Linguistic Perspectives on Second Language Acquisition*. Cambridge: Cambridge University Press.
- Gatbonton, E. 1978. Patterned phonetic variability in second language speech: A gradual diffusion model. *Canadian Modern Language Review* 34: 335-347.
- Goss, N., Z. Ying-Hua, & J. P. Lantolf. 1994. Two heads may be better than one: Mental activity in second-language grammaticality judgments. In Tarone et al: 263-286.
- Gregg, Kevin R. 1989. Second language acquisition theory: The case for generative perspective. In Gass and Schachter 1989: 15-40.

- Gregg, Kevin R. 1990. The variable competence model of second language acquisition and why it isn't. *Applied Linguistics* 11: 363-383.
- Grice, H. P. 1975. Logic and conversation. In Cole, P. and Morgan J.L. (eds.). *Syntax and Semantics Vol. 3: Speech Acts*. New York: Academic.
- Guasti, M. T. & Shlonsky U. 1995. The acquisition of French relative clauses reconsidered. *Language Acquisition* 4, 4: 257-276.
- Guy, G. R. & Bayley, R. 1995. On the choice of relative pronouns in English. *American Speech* 70: 148-162.
- Harvie, D. 1998. *Tracking Down Old Relatives: Zero relatives in subject and non-subject function in Early African American English*. M.A. thesis. University of Ottawa.
- Hashemipour, P. 1989. *Pronominalization And Control In Modern Persian*. Ph. D dissertation, UCSD.
- Hatch, E. (ed.). 1978. *Second Language Acquisition: A book of readings*. Rowley, Mass.: Newbury House.
- Heckler, E. E. 1983. The acquisition of auxiliaries by ESL learners. Texas: Microfiche no. ED240878.
- Honar, A. M. 1992. SURATHAYE QADIMIE BUDAN (The old forms of to be). *Iranian Journal of Linguistics* 2: 21-25.
- Huddleston, R. 1971. *The Sentence in Written English*. Cambridge: Cambridge University Press.
- Huebner, T. 1991. Second language acquisition: litmus test for linguistic theory?. In Huebner & Ferguson (eds.).
- Huebner, T. and C. Ferguson (eds.). 1991. *Crosscurrents in Second Language Acquisition and Linguistic Theories*. Amsterdam: John Benjamins.
- Hulstijn, J. H. and W. Hulstijn. 1984. Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning* 34: 23-43.
- Hyltenstam, K. 1977. Implicational patterns in interlanguage syntax variation. *Language Learning* 27: 383-411.
- Hyltenstam, K. 1978. Variability in interlanguage syntax. *Phonetic Laboratory Working Papers No. 18*. Dept. of General Linguistics. Lund University. Sweden: 1-79.
- Hyltenstam, K. 1983. Data types and second language variability. In Ringbom, H. (ed.). 1983. *Psycholinguistics and Foreign Language Learning*. Abo: Abo Akademia.
- Hyltenstam, K. 1984. The use of typological markedness conditions as predictors in second language acquisition: the case of pronominal copies in relative clauses. In Andersen 1984: 39-58.
- Ioup, G. & A. Kruse. 1977. Interference and structural complexity as a predictor of second language relative clause acquisition. In C. Henning (ed.). *Proceedings of the Los Angeles Second Language Acquisition Research forum*. Los Angeles: Dep. Of English, university of California at Los Angeles.
- Jespersen, O. 1928. *A Modern English Grammar on Historical Principles, Part II*. London: George Allen & Unwin.
- Jordens, P. 1980. Interlanguage research: interpretation and explanation. *Language Learning* 30: 195-207.
- Jordens, P. 1991. Linguistic knowledge in second language acquisition. In Eubank; 201-229.
- Jordens, P. 1992. The cognitive function of case marking in German as a native and a foreign language. In Gass & Selinker (eds.) 1992: 138-175.

- Karimi, S. 1989. *Aspects of Persian Syntax: Specificity, and the Theory of Grammar*. Ph.D. dissertation, University of Washington.
- Karstadt, A. 1996. Relative markers in Swedish-American English: Evidence for a contact language phenomenon? *American Speech* 71: 27-48.
- Keenan, E. L. 1985. Relative clauses. In Shopen: 141-170.
- Keenan, E. L. and B. Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8: 63-99.
- Kellerman, Eric. 1977. Toward a characterization of the strategies of transfer in second language learning. *Interlanguage Studies Bulletin* 2: 58-145.
- Kellerman, Eric. 1978. Giving learners a break: native language intuitions as a source of predictions about transferability. *Working Papers on Bilingualism* 15: 59-92.
- Kellerman, Eric. 1979. Transfer and non-transfer: where are we now? *Studies in Second Language Acquisition* 2: 37-57.
- Kellerman, Eric. 1983. Now you see it, now you don't. In Gass & Selinker (eds.). 1983: 112-134.
- Kellerman, E. 1984. The empirical evidence for the influence of the L1 in interlanguage. In Davies et al: 98-122.;
- Kellerman, Eric. 1985. Dative alternation and the analysis of data: A reply to Mazurkewich. *Language Learning* 35: 91-101.
- Kellerman, E. & M. Sharwood-Smith (eds.). 1986. *Crosslinguistic Influence in Second Language Acquisition*. New York: Pergamon Press.
- Kikai, A., M. Schleppegrell, and S. Tagliamonte. 1987. The influence of syntactic position on relativization strategies. In K. Denning, S. Inkelas, F. McNair-Knox, and J. Rickford (eds.). *Variation in Language: NWAVE XV at Stanford*. Stanford, CA: Stanford university.
- Kohn, K. 1986. The analysis of transfer. In Kellerman & Sahrwood-Smith (eds.).
- Krashen, Stephen D. 1977. Some issues relating to Monitor Model. In H. Brown, C. Yorio, & R. Crymes (eds.). *On TESOL '77*. Washington D.C.: TESOL.
- Krashen, Stephen D. 1981. *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.
- Kroch, A. S. 1981. On the role of resumptive pronouns in amnestying island constraint violations. *CLS 17*. Chicago: University of Chicago: 125-135.
- Krohn, R. 1998. *English Sentence Structure*. Teheran: Mina.
- Labov, W. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. 1969. Contraction, deletion and inherent variability of the English copula. *Language* 45, 4.
- Labov, W. 1971. Some principles of Linguistic Methodology. *Language in Society* 1: 97-120.
- Labov, W. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. 1975. The quantitative study of linguistic structure. *Pennsylvania Working Papers on Linguistic Change and Variation* 1,3: 1-66.
- Labov, W. 1984. Field methods of the project on linguistic change. In J. baugh & J. Sherzer (eds.). *Language in Use: Readings in sociolinguistics*. Englewood Cliffs, NJ: Prentice-Hall: 26-66.
- Labov, W. 1996. Some notes on the role of misperception in language learning. In Bayley and Preston. 1996: 245-252.

- Labov, W., P. Cohen, & J. Lewis. 1968. *A Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York city*. 2 vols. USOR final Report, Research Project No. 3288. Philadelphia: US regional survey.
- Lado, Robert. 1957. *Linguistics Across Cultures*. Ann Arbor. University of Michigan Press.
- Larsen-Freeman, D. & M. H. Long. 1991. *An Introduction to Second Language Acquisition Research*. London: Longman.
- Lazard, G. 1992. *A Grammar of Contemporary Persian*. Costa Mesa, CA: Mazda Publishers.
- Liceras, Juana M. 1986. *Linguistic Theory and Second Language Acquisition: The Spanish nonnative grammar of English speakers*. Tübingen: Narr.
- Lightbown, P. 1984. The relationship between theory and method in second language acquisition research. In Davies et al. 1984: 241-252.
- Lightbown, P. 1987. Classroom language as input to second language acquisition. In Pfaff 1987: 169-187.
- Long, M. 1985. Input and second language acquisition theory. In Gass & Madden (eds.). 1985. *Input in second language acquisition*. Rowley, MA: Newbury House. 119-131.
- McElhinny, B. 1993. Copula and auxiliary contraction in the speech of white Americans. *American Speech* 68: 371-399.
- Meechan, M. 1996. In search of missing link: copula contraction in Canadian English. Ms., University of Ottawa.
- Meechan, M. & Poplack, S. 1996. Orphan Categories in Bilingual Discourse: Adjectival strategies in Wolof-French and Fongbe-French. *Language Variation and Change* 7, 2:
- Meisel, J. 1982. The role of transfer as a strategy in natural second language acquisition/processing, mimeo. Hamburg university.
- Munich, E., Flynn, S., & G. Martohardiono. 1994. Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In Tarone et al: 227-244.
- Myhill, J. 1982. The Acquisition of complex sentences: a cross-linguistic study. *Studies in Second Language Acquisition* 4,2: 193-200.
- Nemser, W. 1971. Approximative systems of foreign language learners. *International Review of Applied Linguistics* 9: 115-123.
- Newmeyer, F. & S. Weinberger. 1988. The ontogenesis of the field of second language learning research. In Flynn & O'Neil (eds.). *Linguistic Theory in Second Language Acquisition*. Dordrecht: Kluwer: 34-45.
- Nobahari, M. 1993. *DASTUR KARBORDY ZABAN FARSI (An Applied Grammar of Persian)*. Teheran: Rahnama.
- Odlin, T. 1989. *Language Transfer: Cross-Linguistic Influence In Language Learning*. Cambridge: Cambridge University Press.
- Odlin, T. 1994. Transferability and lexical restructuring (or: what gets lost in the translation-- and why?). In Blackshire-Belay, C.A. (ed.). 1994: 29-46.
- Olofsson, A. 1981. *Relative Junctions in Written American English*. Goteborg: Acta Universitatis Gothoburgensis.
- Pavesi, M. 1986. Markedness, discoursal modes, and relative clause formation in a formal and an informal context. *Studies in Second Language Acquisition* 8: 38-55.
- Perez-Leroux, A. T. 1995. Resumptives in the acquisition of relative clauses. *Language Acquisition*, 4, 182: 105-138.

- Pfaff, C. (ed.). 1987. *First and Second Language acquisition Processes*. New York: Newbury House.
- Pintzuk, S. 1987. *VARBRUL programs* [computer program]. Philadelphia: U. of Pennsylvania Dep. Of Linguistics.
- Poplack, S. 1989. The care and handling of a mega-corpus: the Ottawa-Hull French project. In Fasold, R. & Schiffrin, D. (eds.). *Language Change and Variation*. Amsterdam: John Benjamins: 411-51.
- Poplack, S. 1993. Variation theory and language contact. In Preston, D. (ed.). *American Dialect Research*. Amsterdam: John Benjamins: 251-286.
- Poplack, S. (ed.). 1999. *The English History of African American English*. Oxford: Blackwell.
- Poplack, S. & Meechan, M. 1995. Patterns of language mixture: Nominal structure in Wolof-French and Fongbe-French bilingual discourse. In Milroy, L. and Muysken, P. (eds.). 1995. *One Speaker. Two Languages*. Cambridge: Cambridge University Press.
- Poplack, S. & M. Meechan (eds.). 1998. Instant Loans, Easy Conditions: the productivity of bilingual borrowing. *The International Journal of Bilingualism* (special issue) 2,2.
- Poplack, S. & Sankoff, D. 1987. The Philadelphia story in the Spanish Caribbean. *American Speech* 62: 291-314.
- Poplack, S. & S. Tagliamonte. 1991. African American English in diaspora: evidence from old-line Nova Scotians. *Language Variation and Change* 3: 301-339.
- Preston, D. 1989. *Sociolinguistics and Second Language Acquisition*. Oxford: Blackwell.
- Preston, D. 1996. Variationist perspective on Second Language Acquisition. In Bayley and Preston (eds). 1996: 1-46.
- Prince, E. F. 1995. On Kind-sentences, resumptive pronouns, and relative clauses. In G.R. Guy, C. Feagian, D. Schiffrin & J. Baugh (eds.). *Towards a Social Science of Language, Vol. 2: Social interaction and discourse structures*. Amsterdam: Benjamins: 223-236.
- Quirk, R. 1957. Relative clauses in educated spoken English. *English Studies* 38; 97-109.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Rand, D. and D. Sankoff. 1990. *GoldVarb: A variable rule application for the Macintosh* (ver. 2.0). Montreal: centre de recherches mathematiques. Universite de Montreal.
- Regan, V. 1996. Variation in French interlanguage: a longitudinal study of sociolinguistic competence. In Bayley & Preston (eds.): 177-202.
- Regan, V. (ed.). 1998. *Contemporary Approaches to Second Language Acquisition in Social Context*. Dublin: University college Dublin Press.
- Richards, J. B. 1990. *Language Development and Individual Differences: A Study Of Auxiliary Verb Learning*. Cambridge: Cambridge University Press.
- Ritchie, William C. & Bahatia, T.K. (eds.). 1996. *Handbook of Second Language Acquisition*. San Diego: academic Press.
- Sadighi, F. 1982. *The acquisition of English restrictive relative clauses by native speakers of Persian*. Ph. D. dissertation. University of Illinois at Urbana-Champaign.
- Sajavaara, K. 1981. Contrastive linguistics past and present and a communicative approach. In Fisiak: 33-56.
- Sajavaara, Kari and Jaakko Lehtonen. 1989. Aspects of transfer in foreign language speaker's reaction to acceptability. In Dechert & Raupach (eds.). 1989: 35-52.
- Samar, R. G. 1998a. Clausal complements in Persian. Paper presented at CLA, 1998. Ottawa.

- Samar, R.G. 1998b. Variability in relative clause usage in Persian: Evidence for transfer? In Paradis, C., D. Vincent, D. Deshaies, & M. Laforest (eds.). *Papers in Sociolinguistics: NWAVE-26*. Quebec: Edition Nota bene: 281-290.
- Samareh, Y. 1985. *AVA SHENASI ZABANE FARSI* (Persian Phonology). Teheran: Iran University Press.
- Sankoff, D. 1982. Sociolinguistic method and linguistic theory. In L.J. Cohen, et al. (eds.). *Logic, Methodology, Philosophy of Science VI*. Amsterdam: North Holland and Warsaw: Polish Scientific: 677-689.
- Sankoff, D. 1988. Variable rules. In U. Ammon, et al. (eds.). *Sociolinguistics: An international handbook of the science of language and society*. Berlin: Walter de Gruyter: 140-161.
- Sankoff, D. 1988. Sociolinguistics and syntactic variation. In Newmeyer, F. (ed.). *Linguistics: The Cambridge survey*. Cambridge: Cambridge University Press: 140-161.
- Sankoff, D. & Cedergren, H. J. (eds.) 1981. *Variation Omnibus*. Edmonton: Linguistic Research Incorporated.
- Sankoff, G. & F. Tarallo. 1987. Relativization and anaphora in spoken language. In *D.E.L.T.A.* 3, 2: 197-214.
- Schachter, Jacquelyn. 1974. An error in error analysis. *Language Learning* 24: 205-214.
- Schachter, Jacquelyn. 1983. A new account of language transfer. In Gass & Selinker 1983: 98-111.
- Schachter, Jacquelyn. 1991. Issues in the accessibility debate: A reply to Felix. In Eubank 1991: 105-116.
- Schachter, Jacquelyn. 1992. A new account of language transfer. In Gass & Selinker 1992: 32-46.
- Schachter, J., Tyson A. and F. Diffley. 1976. Learner intuitions of grammaticality. *Language Learning* 26: 67-76.
- Schmidt, R. W. 1977. Sociolinguistic variation and language transfer in phonology. *Working Papers in Bilingualism* 12. Toronto: DISE: 79-95.
- Schumann, J. H. 1978. *The pidginization process: A model for second language acquisition*. Rowley, MA: Newbury House.
- Selinker, Larry. 1969. Language transfer. *General Linguistics* 9: 67-92.
- Selinker, Larry. 1972. Interlanguage. *IRAL*. 10: 209-231.
- Selinker, Larry. 1984. The current state of interlanguage studies: an attempted critical summary. In Davies et al. (eds.).
- Selinker, L. and D. Douglas. 1985. Wrestling with 'context' in interlanguage theory. *Applied Linguistics* 6: 190-204.
- Sells, P. 1984. *Syntax and Semantics of Resumptive Pronouns*. Ph. D. dissertation. University of Massachusetts.
- Shapira, R. 1978. The non-learning of English: case study of an adult. In Hatch (ed.): 245-255.
- Sharwood Smith, M. & Kellerman, E. 1986. Crosslinguistic influence in second language acquisition: an introduction. In Kellerman & Sharwood-Smith. 1986: 1-9.
- Shohamy, E. 1983. The stability of oral proficiency assessment on the oral interview testing procedures. In *Language Learning* 33, 4: 527-540.
- Shopen, T. (ed.). 1985. *Language Typology and Syntactic Description: Complex Constructions*. Cambridge: Cambridge University Press.

- Smith, J. 1989. Topic and variation in ITA oral proficiency: SPEAK and field-specific tests. In R. Young (ed.). 1989. *The training of international teaching assistants* (special issue). *English for Specific Purposes* 8, 3.
- Sorace, A. 1996. The use of acceptability judgments in second language acquisition research. In Ritchie & Bhatia: 375-409.
- Sridhar, S. N. 1981. Contrastive analysis, error analysis and interlanguage: three phases of one goal. In Fisiak: 207-241.
- Swan, M. 1987. Non-systematic variability: a self-inflicted conundrum. In Ellis (ed.) 1987b.
- Tagliamonte, S. 1991. *A Matter of Time: Past Temporal reference Verbal Structure in Samana English and The Ex-Slave Recordings*. Ph. D. dissertation. University of Ottawa.
- Taglicht, J. 1973. The choice of relative pronouns in written English. *Scripta Hierosolymitana* 25: 327-336.
- Tarallo, F. and L. Myhill. 1983. Interference and natural language in second language acquisition. *Language Learning* 33: 55-76.
- Tarone, Elaine. 1983. On the variability of interlanguage systems. *Applied Linguistics* 4: 142-163.
- Tarone, Elaine. 1985. Variability in interlanguage use: A study of style shifting in morphology and syntax. *Language Learning* 35: 373-404.
- Tarone, Elaine. 1988. *Variation in Interlanguage*. London: Edward Arnold.
- Tarone, Elaine. 1995. A variationist framework for SLA research: examples and pedagogical insights. In Eckman et al. (eds.). 255-272.
- Tarone, E., S. Gass, and A. Cohen (eds.). 1994. *Research Methodology in Second-Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
- Thakerar, J. N., H. Giles, and J. Cheshire. 1982. Psychological and linguistic parameters of speech accommodation theory. In Fraser and Scherer (eds.). 1982. *Advances In The Social Psychology Of Language*. New York: Cambridge University Press: 205-255.
- Tottie, G. 1995. The man Ø I love: an analysis of factors favoring zero relatives in written British and American English. In G. Melchers & B. Warren (eds.). *Studies in Anglistics*. Stockholm: Almqvist and Wiksell. 201-215.
- Tottie, G. & D. Harvie. 1999. It's all relative: relativization strategies in Early African American English. In Poplack (ed.): 198-230.
- Tottie, G. & Rey, M. 1997. Relativization strategies in Earlier African American Vernacular English. *Language Variation and Change* 9: 219-247.
- Towell, R., R. Hawkins, and N. Bazergui. 1993. Systematic and nonsystematic variability in advanced language learning. *Studies in Second Language Learning* 15, 4: 439-460.
- Trudgill, P. (ed.). 1984. *Applied Sociolinguistics*. New York: Academic Press.
- Tushyeh, H. Y. 1985. Transfer and related strategies in the acquisition of English relativization by adult Arab ESL learners. Microfiche no. ED276290...
- Vahedi, M. Mehdi. 1992. *Clausal and Verbal Structures in Farsi*. M.A. thesis. University of Ottawa.
- Veronique, D. 1987. Reference to past events and actions in narratives in second language: insights from North African workers' French. In Pfaff (ed.). 1987.
- Walker, J. 1999. Rephrasing the copula: contraction and zero in Early African American English. To appear in Poplack (ed.).
- Weinreich, U. 1953. *Languages in Contact*. New York: Linguistic Circle of New York.

- White, Lydia. 1989. The adjacency condition on case assignment: Do L2 learners observe the subset principle? In Gass and Schachter 1989: 134-158.
- Windfur, Gernot L. 1979. *Persian Grammar: History And State of Its Study*. The Hague: Mouton.
- Winford, D. 1998. On the origin of African American Vernacular English- a Creolist perspective. In *Diachronica* 15,1: 99-154.
- Wode, H. 1981. *Learning a Second Language 1: an integrated view of language acquisition*. Tübingen: Gunter Narr.
- Wode, H. 1982. Language transfer, language learning, and the function of natural languages. *Arbeitspapiere zum Spracherwerb (APSE)* 28, English Department, Keil University.
- Wode, H. 1984. Some theoretical implications of L2 acquisition research. In Davies et al. (eds.).
- Wolfram, W. 1985. Variability in tense marking: A case for the obvious. *Language Learning* 35: 229-254.
- Wolfram, W. & R. Fasold. 1974. *The Study of Social Dialects in American English*. Englewood Cliffs, N.J.: Prentice-Hall.
- Wolfram, W. and D. Hatfield. 1984. *Tense Marking In Second Language Learning: Patterns of spoken and written English in a Vietnamese community*. Washington, DC: Center for Applied Linguistics.
- Wolfson, N. 1989. *Perspectives: Sociolinguistics and TESOL*. Cambridge, MA: Newbury House.
- Wolfson, N. and E. Judd (eds.). 1983. *Sociolinguistics And Language Acquisition*. Rowley, MA: Newbury House.
- Yip, Virginia. 1995. *Interlanguage and Learnability: From Chinese to English*. Amsterdam: John Benjamins.
- Yorozo, M. 1995. A psycholinguistic study of relative pronoun use by native speakers and non-native speakers of English. *Texas Papers in Foreign Language Education* 2,1.
- Young, Richard. 1986. The acquisition of verbal repertoire in a second language. *Penn Working Papers in Educational Linguistics* 2,1: 85-119.
- Young, Richard. 1991. *Variation in Interlanguage Morphology*. New York: Peter Lang.
- Young, Richard. 1996. Form-Function relations in articles in English interlanguage. In Bayley and Preston 1996: 135-176.
- Young, R. & R. Bayley. 1996. VARBRUL analysis for second language acquisition research. In Bayley & Preston (eds.): 253-306.
- Zobl, Helmut. 1980. The formal and developmental selectivity of L1 influence on L2 acquisition. *Language Learning* 30: 43-57.
- Zobl, Helmut. 1982. A direction for contrastive analysis: the comparative study of developmental sequences. *TESOL Quarterly* 16,2: 167-183.
- Zobl, Helmut. 1983. L1 Acquisition, age of L2 acquisition, and the learning of word order. In Gass & Selinker 1983: 205-221.
- Zobl, Helmut. 1988. Markedness Constraints on acquisition and prior linguistic experience. In Dechert & Raupach 1988: 141-158.
- Zobl, Helmut. 1992. Prior Linguistic Knowledge and the conversation of the learning procedure: Grammaticality judgments of unilingual and multilingual learners. In Gass & Selinker 1992: 176-196.