AMERICAN UNIVERSITY OF ARMENIA

College of Humanities and Social Sciences

An Investigation of the Relationship between Linguistic Features of Spoken English and English Learning Experiences of Armenian Learners of English

A thesis submitted in

partial fulfillment of the requirements for the degree

Master of Arts in Teaching English as a Foreign Language

By

Gayane Tadevosyan

Irshat Madyarov, Ph.D., Adviser

Lilianna Edilian, MA, Reader

Alexan Simonyan, Ph.D., Statistics Consultant

Yerevan, Armenia

September 7, 2012

We hereby approve that this thesis/paper

By

Gayane Tadevosyan

Entitled

An Investigation of the Relationship between Linguistic Features of Spoken English and English learning experiences of Armenian Learners of English

Be accepted in partial fulfillment for the requirements of the degree

Master of Arts in Teaching English as a Foreign Language

Committee on the MA Thesis

……..……
Irshat Madyarov, Ph.D., Adviser

……..
Lilianna Edilian, MA, Reader

……..……
Catherine Buon, Ph.D.

Associate Dean of College of Humanities and Social Sciences

Yerevan, Armenia

September 7, 2012

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Irshat Madyarov for his kind attitude, endless patience, motivation and continuous support throughout all the process of my thesis research. His guidance helped me across all the time of my research and writing of my paper. Besides this, I not only managed to accomplish my research successfully thanks to him, but also got certain knowledge in research and developed so many important techniques and skills for my further academic path. I could not have imagined having a better advisor for my MA thesis. Thus, I really appreciate all the time and effort he devoted to my study in general.

I would like to thank my thesis reader Ms. Lilianna Edilyan for her patience in reading out my paper and giving insightful comments and feedback. Her knowledge helped me to find out solutions to specific linguistic complications in my study.

I am grateful to the statistics consultant Dr. Alexan Simonyan for his support. With his help I was able to implement the right statistical tools of data analysis.

I would also like to thank all the participants of my research for their efforts dedicated to my study. Thanks to their contribution I managed to get the data I needed.

In particular, I am grateful to all the instructors and professors of the College of Humanity and Social Sciences at the American University of Armenia who gave me fundamental knowledge and support across the long way of my study at the master's program of Teaching English as a Foreign Language.

Last but not the least, particular thanks go to my family, especially to my lovely daughter for their love, understanding and support throughout the tough days in my study.

TABLE of CONTENT

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The present study investigates the relationship between the learning experiences of Armenian nonnative speakers of English and the target linguistic features of their spoken English. The number of the participants of the study was 34, of whom 10 were American native speakers of English, and 24 Armenian nonnative speakers of English. Two instruments have been used for data collection: 1) a structured interview to compile a corpus of native speaker and nonnative speaker speech samples and 2) a questionnaire to identify Armenian nonnative speakers' English learning experiences in secondary, undergraduate, graduate and postgraduate schools, as well as private tutoring and self study efforts.

The results of the investigation show that nonnative speakers used the target linguistic features less frequently than native speakers did. Correlation analysis has shown that such kinds of qualities of exposure as interaction with native speakers, listening and summarizing materials, doing debates and role plays and giving oral presentations significantly correlate with native like competence as separate independent variables. However, the results of the backward multiple regressions have identified that only native speaker interaction in private tutoring experience has significant predictability of native like competence of Armenian nonnative speakers of English.

CHAPTER ONE: INTRODUCTION

This chapter states the problem for the investigation of the present corpus-based research. It also describes the purpose and significance of the study, as well as presents the research questions and the definitions of the terms used across the paper.

1.1 Statement of the Problem

As the literature indicates, investigations of spoken-corpora of English learners who are speakers of other languages, such as Swedish, French, Chinese, Thai, Korean and Spanish, have identified differences in the use of linguistic features as compared to native speakers (Phoocharoensil, 2011; McCarthy & Carter, 2006b; Koya, 2003; Nesselhauf, 2003; De Cock, 2004; Aijmer, 2004; Mukherjee, 2009; Conklin & Schmitt, 2008; Jiang & Nekrasova, 2007; Ellis, Simpson-Vlach & Maynard, 2008). The speech of nonnative speakers of English who are speakers of other languages can be correct in general sense, but it may not sound native like.

1.2 Purpose of the Study

The literature provides some findings about the cause of the differences between linguistic features used by native speakers and nonnative speakers of English who are speakers of other languages (Nesselhauf, 2003; Mukherjee, 2009; Phoocharoensil, 2011; Koya, 2003). The potential cause seems to be the influence of L1 (first language) or negative transfer; that is, the learners of English oftentimes literally translate from their mother tongue to the foreign language (Nesselhauf, 2003; Phoocharoensil, 2011). However, there appears to be a scarcity of research on the influence of English learners'

English learning experiences on the difference in the use of specific linguistic features in nonnative speakers' spoken English and that of native speakers.

Thus, the aim of this study is to investigate the relationship between English learning experiences of Armenian learners of English and the target linguistic features of their spoken English. This will help to draw better conclusions about possible causes of English learning experiences on native like competence of nonnative speakers of English. For this purpose, first, the data of the target linguistic features of spoken English of native and nonnative speakers of English have been collected and compared. This has been done to see whether the target linguistic features can serve as distinguishing features of native-like competence. Then, the relationship between the linguistic features of spoken English of nonnative speakers of English and their English learning experiences has been investigated.

1.3 Significance of the Study

Since there seems to be a lack of research on the investigation of possible influence of English learning experiences on the spoken English of learners who are speakers of other languages, the study will provide valuable findings for research and theory of EFL (English as a Foreign Language) and SLA (Second Language Acquisition) in general. However, in particular, based on the results of the data analysis, some conclusions to facilitate English teaching and learning in Armenia can also be made.

1.4 Research Questions

The following two research questions have been raised to address the goal of the present study:

Research Question 1

Is there a difference between the linguistic features in spoken English of Armenian nonnative speakers and American native speakers of English?

Research Question 2

What is the relationship between the English learning experiences of Armenian nonnative speakers of English and the linguistic features of their spoken English?

1.5 Definition of Terms

NNS – Armenian nonnative speakers of English, graduate students at the American University of Armenia (AUA), aged 20 years old and older;

NS – American native speakers of English, aged 20 years old and older;

LF – linguistic features (fillers, modals, adverbs, collocations, phrasal verbs, structure), operational definitions of whose are given in Chapter Three of the paper;

ELE – English learning experiences of the NNS participants including secondary, undergraduate, graduate (AUA and / or other) and postgraduate schools, as well as private-tutoring and self-study experiences (Chapter 3);

CHAPTER TWO: LITERATURE REVIEW

Introduction

Literature review related to the present study addressed two main aspects of the investigation of the English language. First, an overview of the criteria for the English spoken grammar is presented (McCarthy & Carter 2006a; Mukherjee, 2009). Afterwards, the findings of different studies of investigations of native and nonnative speakers' spoken corpora are discussed (Phoocharoensil, 2011; McCarthy & Carter 2006b; Koya, 2003; De Cock, 2004; Aijmer, 2004, Conklin & Schmitt, 2008; Mukherjee, 2009; Jiang & Nekrasova, 2007; Ellis, Simpson-Vlach & Maynard, 2008). Some studies have investigated English written corpora of native and nonnative speakers, too (Nesselhauf, 2003; Mukherjee, 2009). The literature also investigates the possibility of L1 influence on the use of linguistic features in learner corpora (Koya, 2003; Nesselhauf, 2003; Phoocharoensil, 2011). And finally, the summary of literature review with conclusions for the focus of the study is presented.

2.1 Ten criteria of spoken grammar

First of all, to establish the ground for the linguistic features specific to spoken language, it is worth having an in-depth look at the ten criteria of spoken language suggested by McCarthy and Carter (2006a). Those criteria were established based on the evidence in native speaker corpus, more specifically CANCODE spoken corpus established at the Department of English Studies, University of Nottingham, United Kingdom, funded by Cambridge University Press. All the ten criteria of spoken corpus are

suggested by the authors to be included in the grammar of English. Those criteria are listed and discussed below.

1. Establish core units of a spoken grammar

McCarthy and Carter (2006a) believe that sentences in spoken language seem to appear in shorter conversational forms that consist of incomplete clauses, either just phrases or subordinate clauses that do not link to any particular main clause. Although they are not sentences, these clauses may be communicatively complete and intelligible for the speakers. They also might contain words of unclear grammatical class, such as *'yow', 'rm', 'erm'*. And these forms of spoken grammar seem to be considered 'adequately formed' to the specific context rather than 'well formed' which seems to exist in the written language.

2. Phrasal complexity

In spoken language simple noun-adjective combinations are more likely to occur, rather than more sophisticated forms of similar combinations which are specific to written language. For the explanation of the phenomenon of complexity in spoken language, the authors distinguish two types of grammar: *deterministic* and *probabilistic*. Deterministic grammar addresses structural features specific to traditional grammatical rules.Whereas, probabilistic grammar interprets features as they occur in corpus data in specific contexts and considers the likelihood of the use of certain forms in particular contexts. The issue of phrasal and other kinds of complexities seems to be subject to the notion of probabilistic grammar. The authors also use grammatical possibilities to describe, for instance, the

occurrence of *get*-passive verb phrase (e.g. '*He got killed'* instead of '*He was killed'*) in native speaker spoken discourse.

3. Tense, aspect, voice, and interpersonal and textual meaning

A specific feature of spoken language in the choice of tense-aspect-form is tentativeness and indirectness, that is a politeness strategy that weakens the impositions and threat to face and shows detachment. It also appears that the use of progressive with verbs that barely would exist in progressive in written language is common to the spoken mode of the language (e.g. '*I'm wanting to go to Holland next week'*). This fact may be explained by tentativeness and indirectness, which is usually the case in spoken narratives, where there seems to be more freedom in the use of tense-aspect-form. In spoken narratives the mixture of past and historical present is employed purposefully to create a more real picture of the story for the hearers as if they are participating in the drama (e.g. '*He got himself locked out'*, '*The tape seems to have got stuck'* or '*He had the ring stolen'*).

4. Position of clause elements

The position of clause elements in spoken language can mainly refer to the flexibility of using adverbs in various places in a sentence, even sometimes splitting infinitives, which is often unacceptable for prescriptive grammarians. Cases of placing content matter outside the core clausal positions are also frequent in spoken data (e.g. '*His cousin in Beccles, her boyfriend, his parents bought him a Ford Escort for his birthday'*). Also, verbs can be left- or right-dislocated to the beginning or the end of the sentence such as in '*It can leave you feeling very weak, it can, though, apparently, shingles, can't it'*

[when talking about someone who has just had the disease, shingles]. To sum up, the flexibility in word order seems common to spoken English.

5. Clause complexes

The issue of 'subordination' is another phenomenon to be considered as a criterion for spoken language, most specifically '*which*' clauses (e.g. '*I can't angle it to shine on the music stand, and the bulb's gone, which doesn't help')*. Moreover, sometimes main-subordinate clauses may 'blend' together *(e.g. 'Which it's all relative I suppose').*

6. Unpleasant anomalies

Such a deviation as double negatives can be an example of an unpleasing anomaly, e.g.:

Speaker 1: We probably won't see much wildlife.

Speaker 2: Not without binoculars we won't.

Another anomaly that appears in spoken data is conditional clause divergence (e.g. '*If I'd have stopped I probably would have wondered what she was going to say'*, instead of '*if I had stopped...'*).

7. Larger sequences

Some sequences are characteristic both to oral and written corpus such as *used to – would* sequence (e.g. '*They used to, you know ring up early hours of the morning', 'well you would', 'the phone wouldn't ring', 'they'd ring that computer'*). However, there are some sequences that are specific to both spoken and written modes but rare in everyday use in spoken mode, such as initial *be to*-plus-*will* (e.g. '*Five thousand jobs are to be axed by electricity generating firm National Power, it was announced yesterday. Smaller power*

stations will close but bosses pledged no compulsory redundancies over the next five years').

8. The comparative criterion

It seems that spoken and written grammars have notable differences, discussed above. However, it is worth comparing grammatical features occurring in the two modes to find relevant similarities and differences and integrate them into the grammar of English. For instance, conjunctions seem to be used both in spoken and in written English, but it appears that some of them are more specific to written language rather than informal speech. For example, '*on the contrary'* is mostly specific to formal conversations, but in writing it is more common usually in the beginning (e.g. '*He had no private understanding with Mr X. On the contrary he knew very little of him')*. However, '*on the other hand'* is equally common both in written and in spoken contexts, but in spoken corpus '*but then'* seems a more preferable counterpart of the latter. Other conjunctions such as '*moreover', 'furthermore', 'afterwards',* '*as a consequence'* are more used in written contexts; whereas, some others are more specific to spoken language (e.g. '*what's more', 'as I say', 'because of that'* and so on).

9. Metalanguage

The ninth criterion is necessary for linguists to take into account when investigating the grammar of speech in transcribed texts. It mainly addresses the placement of elements or clauses of the sentences to find out cases of displacement or dislocation in spoken context. This is particularly useful to readdress the SVO (subject-verb-object) word order in spoken corpus.

10. Native and non-native users

The final criterion refers to the authority of grammatical description with respect to the spoken corpus. It has been recently observed that the authority users and dictators of grammar of conversation shift from highly literate members of population (e.g. writers) towards a wide range of people who use the language on a regular basis in their everyday communication. In the latter case a variety is more important to consider in corpus data of the English language. Since English is a language that nowadays is widely used all over the world, we need to take into account the fact that not only native speakers of the language use it in spoken interaction. Moreover, in fact native speakers seem to be in a minority among the total number of English users. Hence, for the grammar of conversation, a variety is a crucial aspect for consideration. To establish the corpus-representative grammar of conversation, different kinds of data, such as country-based, regionally-based, native-speaker-based, nonnative-speaker-based as well as mixed-based should be cross-compared. Thus, given this condition it seems hard to establish the notion of 'expert' users of spoken English, since it can both include successful nonnative communicators and native speakers of English in the same category.

Some of the ten criteria given by Mc Carthy and Carter (2006a) are characterized as specific features of spontaneous spoken language by Mukharjee (2009), too. Those are: 1) the use of relatively unintegrated non-clausal 'fragmentary' units, 2) the inappropriateness of the sentences to the grammatical and syntactical analyses, 3) the simplicity of phrase structures and 4) the repetitive use of lexico-grammatical units. Mukharjee (2009) also claims that long passives (with an explicit *by*-agent) and left-dislocations (e.g. '*The car', 'where is it?')* need to be included in the grammar of

conversation, as it was also suggested by Mc Carthy and Carter (2006a). Another aspect to be included into the spoken grammar of English is phraseology, including collocations (‘*little baby*’) and lexical bundles, i.e. recurrent multi-word sequences (e.g. ‘*I don’t know*’).

Taking the classification of the ten criteria proposed by McCarthy and Carter (2006a) supported by Mukharjee (2009), as a headstone for the literature investigation, studies of spoken and written communication of nonnative speakers of English are discussed in the second subsection of the current literature review.

2.2 Specific components of learner spoken corpus

In the following part of the literature review, the findings of different features of native speaker and learner corpora of English are presented. The attention has been drawn to the use of multi-word units such as *collocations* (Phoocharoensil, 2011; McCarthy & Carter 2006b; Koya, 2003; Nesselhauf 2003), *discourse markers* (De Cock, 2004; Aijmer, 2004; McCarthy & Carter, 2006a; Mukherjee, 2009), and *idioms* (Conklin & Schmitt, 2008; Jiang & Nekrasova, 2007; Ellis, Simpson-Vlach & Maynard, 2008) in nonnative and native speaker corpora. The studies in literature have investigated the corpora of English learners of different linguistic backgrounds such as Japanese (Koya, 2003), Thai (Phoocharoensil, 2011), German (Nesselhauf 2003), French (De Cock, 2004), Korean (No & Park, 2010), Swedish (Aijmer, 2004) and mix-groups of linguistic backgrounds such as Arabic, Bulgarian, Chinese, Czech, Korean, Polish, Portuguese, Romanian, Russian, Spanish and Turkish (Conklin & Schmitt, 2008; Jiang & Nekrasova, 2007; Ellis, Simpson-Vlach & Maynard, 2008).

2.2.1 Multiword strings that function as discourse markers

Discourse markers and general extender are included into the category of multi-word sequences. As Pichler and Levey (2010) state, discourse markers are interactional maneuvering devices that are strategically used by speakers to signal speakers' attitudes and to structure discourse. They can perform different functions such as gap filling, hesitation, agreement or polite disagreement, statement conclusion, and so forth (Pichler & Levey, 2010; Jabeen, Rai & Arif, 2011). Tagliamonte and Denis (2010) indicate that general extenders also function as discourse markers and are used to display vagueness, approximation, continuity, deflection in speech, face-threatening mitigations, and can be used as delay-devices in spoken language.

Corpora studies (Phoocharoensil, 2011; McCarthy & Carter, 2006b; Koya, 2003; Nesselhauf, 2003, De Cock, 2004; Aijmer, 2004) show that much of the English lexis is used as multi-word ready-made units. McCarthy and Carter (2006b) believe that some multi-word strings maintain functions other than syntactic and semantic. Those features hold interactive functions of pragmatic integrity such as vagueness, hedging, discourse marking, and so on (e.g. '*and that sort of thing', 'you know', 'a couple of'*). De Cock (2004), Aijmer (2004) and McCarthy and Carter (2006b) have identified that markers of purposive vagueness and approximation are of high frequency clusters in native speaker spoken corpora (e.g. '*A couple of', 'And things like that', 'Or something like that', '(And) that sort of thing', '(And) this that and the other', 'All the rest of it', '(And) all this/that sort of thing'*).

The implications of the study conducted by McCarthy and Carter are based on the data of 5-million-word 'Cambridge and Nottingham Corpus of Discourse in English'

(CANCODE) of transcribed conversations of spoken corpus (McCarthy & Carter 2006b). CANCODE was established at University of Nottingham. The corpus recordings were made in a variety of settings such as private homes, shops, offices and other public places, in non-formal settings across the islands of Great Britain and Ireland. The population represented a wide range of demographic groups.

McCarthy and Carter (2006b) have identified the most frequent pragmatically integrated clusters and described their functions. They conclude that multi-word strings are employed more frequently than single words of core vocabulary of English. Moreover, the authors claim that idioms occur in spoken corpora quite frequently (e.g. '*kick the bucket', 'pass the buck'*). *Lexical bundles* are also considered multi-word units that appear to be incomplete but meaningful (e.g. '*to be able to', 'as a result of*', '*on the other hand', 'a lot of the'*). Usually, single word items are believed to be central for learning English basics, while acquisition of units that consist of more than one word (e.g. phrasal verbs, compounds and idioms) is usually considered to belong to a higher level of achievement. However, there are also multi-word expressions that are likely to be learnt and used by low level learners of English, too. Those can be expressions like '*How's it going?', 'See you soon', 'Thanks a lot'*, some specialized functional phrases (e.g., '*Happy birthday', 'Good luck'*), basic prepositional phrases (e.g., '*in the morning', 'at home'*), and compounds (e.g., '*car park', 'check-in'*).

Investigated multi-word sequences in the literature include two-, three-, four-, five- and six-word combinations, with the least frequent being the latter (e.g. '*and all the rest of it'* or '*this that and the other'*), and the most common, two-word sequences (e.g. '*you know', 'I mean', 'I think', 'in the', 'it was'*). Some of the most common clusters maintain

discourse-marking functions (‘*You know’, ‘I mean’, ‘And then’, ‘But I mean’, ‘You know what I mean’, ‘If you see what I mean’).* Aijmer (2004) and McCarthy and Carter (2006b) believe that speakers use indirect forms when performing certain speech acts (directives and requests) to “protect the face” of their receivers and show politeness and non-face-threatening expression of attitude, opinion or position (‘*Do you think’, ‘Do you want (me) (to)’, ‘I don’t know if/whether’, ‘What do you think’, ‘I was going to say’).* The authors consider the use of multi-word units or *formulaic sequences* such as collocations, lexical bundles, idioms and phrases in learner language as core measures for evaluating native like competence.

De Cock (2004), No and Park (2010) and Aijmer (2004) have analysed the data of spoken corpora of French, Korean and Swedish learners of English, respectively, and compared the data with those of native speakers (NS). They revealed different features occurred in the two spoken corpora.

De Cock (2004) conducted corpus-driven qualitative and quantitative analyses of recurrent sequences of more than two words occurred in the spoken corpus of French advanced learners of English and that of NSs. The 50 nonnative speaker participants (NNSs) were third and fourth year students at the Université catholique de Louvain aged from twenty to twenty-six. The 50 native speakers of English were first to fourth year undergraduate and postgraduate students at Lancaster University, Great Britain. The NNS data consisted of the French component of Louvain International Database of Spoken English Interlanguage (LINDSEI), and the data of native speaker corpus was the Louvain Corpus of Native English Conversation (LOCNEC). The NS corpus consisted of 117,417 words and that of the NNSs’ of 90,300 words, with almost the same length of 2000 words

and duration of 15 minutes on average for each interview. The informal interviews were open discussions about the topics such as university life, hobbies, foreign travel or plans for the future and a picture to describe. The interviewees were given some time to get ready before the talks, but for the sake of spontaneity, they were asked not to take notes. Questionnaires, as profiles of biographical data of the learners, were given to the interviewees during the interview sessions. The focus of linguistic units of the investigation was limited to 2 to 6 recurrent word sequences. The quantitative part of the study was drawn upon the recurrence of the word combinations, whereas the qualitative one was devoted to the exploration of the functional aspects of those word sequences.

As findings from the quantitative analysis have shown, the most frequent sequences were two and three word sequences in both corpora. The results were congruent with the findings in the study conducted by McCarthy and Carter (2006b), where the most frequent multi-word strings were two-word sequences. De Cock (2004) identified that one of the main differences between the two corpora was the higher proportion of three-word sequences containing repetitions (e.g. '*the…the', 'I… I'*) and hesitation items (e.g. '*er', 'erm'*) in the native speaker corpus. The top 20 three-word sequences of nonnative speaker speeches were: '*I don't know', 'I I I', 'and it was', 'and er well', 'the the the', 'and er I', 'and er the', 'it was really', 'it was er', 'it was a', 'and er we', 'and so on', 'no no no', 'but I I', 'to to to', 'I I was', 'yes yes yes', 'a lot of', 'I would say', 'I went to'*. The most frequent word sequences (almost 60 %) were initial thematic jumping offs both in native speaker and nonnative speaker data, but with a higher proportion in the latter data. This means that nonnative speakers seemed to need more planning time at the beginning of clauses. Moreover, there seemed to be a higher proportion of phrasal sequences containing

repetitions or hesitation items. Learners also had encoding problems in that they were using words from their mother tongue, namely French. For example '*Enfin',* a French equivalent of English '*well'* or '*I mean'* that actually occurred 74 times (per 100,000 words). But the author did not consider the use of this item as code-switching because of the unconscious nature of its use.

The qualitative data indentified a wide range of structural and functional variety, where within the structural aspect, the distinction between complete ('*It's not too bad'; 'at the moment'*) and incomplete sequences ('*I really enjoy', 'a couple of*') was made. From the functional perspective, three major categories of the word-sequences were classified: 1) *referential sequences* (e.g. markers of time/place: '*at night', 'during the day', 'in front of*'; quantifying sequences: '*loads of', 'one of the', 'an awful lot of*'; topic-dependent sequences: *'a film'*; etc.), 2) *interactional/interpersonal sequences* (e.g. markers of attitudinal stance: *'I really enjoyed', 'which is good', 'it was very', 'I'm hoping to'*; markers of epistemic stance: *'but I think', 'I don't know if', 'I can't remember'*; responses: '*yeah definitely', 'that's it'*; markers of vagueness: '*sort of', 'and things like that'*; etc.), and 3) *discourse-organizing sequences* (e.g. markers of speech/thought reporting: '*so I thought', 'and I was like oh'*; markers of contrast: '*on the other hand'*; makers of cause: '*due to the fact'*; exemplifiers: '*for example', 'for instance',* etc.). The author has mainly addressed the use of interactional/interpersonal sequences, specifically, markers of vagueness. It was revealed that recurrent sequences as vagueness markers in native speaker corpus included response items (e.g. '*yeah', 'oh', 'well'*), discourse items ('*you know', 'I mean', 'like'*), first and second person pronouns ('*I'*, '*you'*), private verbs ('*think', 'know', 'remember'*) used to convey attitudinal or knowledge positions. These vagueness markers were

underused by learners, and were, on the contrary, less interactional and involved in the nature of the talk. Two main sets of vagueness markers were particularly addressed in the study, namely 'vagueness tags' (VT) such as for example, '*or something'*, '*and things'* or '*or anything'* and 'discourse items' (DI) '*sort of*' and '*kind of*'. It was found out that learners employed certain VT with low frequencies ('*and stuff', 'and stuff like that (and)', '(and) that sort of thing', 'sort of thing (but/so)', 'or anything', 'and places like that'* and *'all the rest of it'*). The use of VT was somewhat limited in the learner corpora to '*and so on', 'for example/ instance'* or '*et cetera'*, which appeared to be more detached and formal. Learners also significantly underused the sequences containing '*sort of*' and '*kind of'*. These findings coincide with those in McCarthy and Carter (2006b) where it is said that native speakers use plenty of vagueness markers, specifically those of purposive and approximation functions in their spoken English. Moreover, native speakers used these sequences to show ambiguity of the coming utterances; whereas, learners used them as communication strategies such as 'language switch', since in most cases the words following those sequences were borrowed from French. They also preferred the use of *'kind of*' over *'sort of',* as a result of influence of American movies or songs. It is noteworthy that the sequence *'of course'* was mainly overused and sometimes misused by nonnative speakers both in its longer or shorter forms (e.g. *'yes of course'*, '*well of course'*).

No and Park (2010) have investigated the differences in the interviews of Korean and American speakers of English. The study was conducted on 14 interviewees: 7 American Native Speakers of English (ANSE) and 7 Korean Speakers of English (KSE). The study investigated the topics, the use of word number and class, as well as the use of

discourse markers in the self-introductory interviews of the participants. It is worth mentioning that 'fillers' such as *'haha', 'and(ah)', 'yah', 'uhm (ahm, uh*)', so called echo-type discourse markers (DM), were also considered as discourse markers in the scope of this study, since they were used to make the conversation flow more natural, cohesive and smooth. The findings of the research showed that there were some differences between ANSE and KSE in the use of four components in their self-introduction topics (early years, education, work experience and current situation). As it was revealed, 70 % of ANSE talked about their current state, 86 % gave information about their birth, 71 % included information about their educational background, and 57% talked about their work experience. Whereas, only 28% of KSE included information about their birth, 57% talked about their educational background, and more than 50% did not talk about their work experience at all.

It appeared that KSE used slightly more nouns, verbs, and adjectives; whereas, ANSE seemed to use conjunctions and contractions more. ANSE tended to be briefer and not willing to give information about personal matters, while KSE, on the other hand, seemed to talk more about themselves to make sure they were understood.

As to discourse markers, KSE and ANSE mainly employed different DMs. Only five DMs were used both by ANSE and KSE. They were the following: *'almost,' 'and,' 'so,' 'well,' and 'and then.'* Some DMs were only used by KSE (e.g. *'do you understand…?' 'Right' 'I think that's about it,' 'and later on,' and 'you know'*). ANSE appeared to use DMs such as *'or', 'that's it', 'very typically', 'but then', 'and since then', 'OK' and 'let's see'* more frequently. The most frequently used DM in the speeches of KSE was *'now'*, while ANSE employed *'that's about it'* most frequently. The following

DMs were not found in KSE self-introduction at all: *'That's it', 'or', 'very typically', 'but then', 'and since the', 'OK', and 'let's see'*. DMs such as *'do you understand…?' 'right', 'what else', 'I think…', 'and later on', 'then', 'actually'* and *'you know'* were not found in the speeches of ANSE.

Echo-type DMs, *'ah', 'andah', 'oh', 'uhm', 'yah', 'uh', 'and uh'* and *'haha'* were the ones that were used both by ANSE and KSE. Moreover, the echo-type DMs used both by ANSE and KSE were the same in frequency per type. However, ANSE tended to employ more DMs other than 'echo-type' than KSE and it seemed reasonable that KSE might want to have more time to think about what to say further on.

Some limitations of the research done by No and Park (2010) do not seem to provide generalizable conclusions. One of the major limitations was the small number of participants of the study.

Aijmer (2004) investigated linguistic features of spoken English of advanced Swedish learners of English when being interviewed by a native speaker. It was an exploratory study that consisted of 50 interviews (around 100,000 words), each which lasted 15 minutes, but in the scope of the study only 10,000-word corpus was analysed (Aijmer, 2004). The participants were third year students at Göteborg University. The topics of the interviews were a recent trip or a movie they had seen and they were also asked to describe a series of pictures from a comic strip. Like No and Park (2010), Aijmer (2004) also identified that markers co-occurred with pauses as planning devices in the learner speeches. Though the non-native speakers and native speakers used the same markers, they had differences in the frequencies of the use and in the functions of individual markers. '*I think', 'you know', 'sort of', 'I mean', 'well', 'actually', 'really'*

were frequent in both groups, but, only the learners used '*don't know'* and '*yeah';* while only native speakers employed '*you see'*. The author also claimed that learners used vagueness markers to express uncertainty or hesitation, rather than to show face-saving or to signal politeness. Moreover, the learners used markers as strategies when having faced communication problems. For example, they would put the vagueness markers into conversation for the hearer to complete the message. The marker '*I don't know'* was used by the learners more frequently, which made the learners sound more uncertain than native speakers, thus it even occurred in combination with other markers.

The high frequency markers employed by the learners were: '*I think' (40), 'sort of' (38), 'well' (38),'I don't know' (28), 'actually' (26), 'you know' (23), 'like' (14), 'I mean' (13), 'yeah' (13) [not as an answer to a question], 'or something' (11), 'kind of' (8)*. Other markers were used by learners with less frequency (e.g. *'I guess', 'and all that', 'and everything', 'and stuff like that', 'or anything', 'or something like that', 'and things like that', 'and stuff', 'something like that', 'stuff like that'*). *'Well'* was used by the learners inside the turns as a pause-filler or before reformulation, and '*yeah'* was used as a pause-filler where '*well'* would have been expected. '*You know', 'I think'* and '*sort of'* were hard to distinguish functionally, though *'sort of'* was more frequent among learners. '*I think'* was more preferable for both groups than '*you know'*. Like De Cock (2004), Aijmer (2004) believes that the use of some markers by the learners can be explained by the American influence (e.g. '*I guess'* and '*kind of'*).

Learners also used '*I don't know'* more frequently, which might serve an uncertainty device or 'filler'; whereas, the native speakers used it to signal disagreement

and to avoid commitment in addition to being a marker of uncertainty. This findings show replication with the results of the study conducted by DeCock (2004).

2.2.2 Collocations

The notion of *collocations* has changed the emphasis in linguistics from a single word to pairs of words as integrated chunks of meaning so that collocations have become one of the major elements in the field of English teaching. As it is shown by Mukharjee (2009) and Phoocharoensil (2011), collocations are considered *words that appear in context with other words with greater than random possibility, which are fixed and highly predictable from one of the components*. Collocations are classified into lexical and grammatical subcategories (Phoocharoensil 2011). Lexical collocations include two or more words, mainly content words such as nouns, adjective, verbs and adverbs, while grammatical collocations consist of a content word and a function word (usually prepositions).

Examples of lexical collocations are:

adjective + noun: sour milk

verb + noun: conduct research

noun + verb: dust accumulates

adverb + adjective: mentally disabled

verb + adverb: move freely

adverb + verb: proudly present

Examples of grammatical collocations are:

noun + preposition: an increase in

verb + preposition: elaborate on

adjective + preposition: familiar with

preposition + noun: on probation

Koya (2003), Nesselhauf (2003), and Phoocharoensil (2011) investigated the use of collocations in learner corpora. All the three authors found similarities in terms of the

use of collocations in learner corpora regardless the difference in the linguistic backgrounds of the participants.

The participants of the study conducted by Phoocharoensil (2011) were 90 first-year undergraduate students, Thai learners of English at a university in Thailand. The students were of two levels of proficiency (high and low) determined according to the results of the Ordinary National Educational Test (O-NET), a University Entrance Exam under the Ministry of Education of Thailand. The participants' native language was Thai and they learned English as a foreign language (EFL) for at least 12 years (1^{st}-12^{th} grades). Koya (2003), Nesselhauf (2003) and Phoocharoensil (2011) show that most of the errors in the use of collocations seem to be a result of negative transfer (first language influence). In their studies they found that the collocations in the mother tongue of the participants and those in English were congruent.

Koya (2003) investigated the use of collocations by 93 Japanese learners of English, first-year university students, who learnt English as an EFL for at least six years. Three tests were conducted to establish 1) the receptive and 2) productive vocabulary knowledge of the participants, and 3) collocational use. Furthermore, the participants were divided into three level groups by vocabulary knowledge. Then, the author did a correlation analysis (using SPSS program) of the use of collocations, and concluded that the richer general vocabulary the learners had, the more accurately they used collocations. It was also difficult for the participants to describe or paraphrase, provided they did not know the target collocations regardless their level of proficiency.

Nesselhauf (2003) investigated the use of collocations by advanced German learners of English in 32 argumentative essays (each of about 500-word length) in the

German subcorpus of International Corpus of Learner English (Ge ICLE). The author classified word combinations into three categories based on the notion of *restricted sense*:

1) free combinations (in which both a verb and a noun can be freely substituted): '*want a car', 'buy a car', 'drive a car'*;

2) collocations (in which the sense of elements is restricted to certain words): '*take a photo/picture', 'but not take a film/movie';*

3) idioms (in which substitution of the elements is unaccepted): *sweeten the pill*.

Overall 1072 verb-object-noun combinations were identified in the essays, of which 213 were collocations, 846 – free combinations and 13 – idioms. It is noteworthy that 255 of all 1072 combinations contained 'mistakes' ('*make one's homework', 'give a solution to', 'take one's task'*).

Nesselhauf (2003) found out that the wrong collocations had equivalents in German, so the students could possibly transfer the collocations from their first language (L1) into English ('*make*/do homework', 'close lacks*/gaps', 'train*/exercise one's muscles', 'draw a picture from*/of'*). Though there were some occasions of positive L1 transfer into English too, in most of the cases learners had considerable deficiencies.

2.2.3 Discourse markers, collocations, repetitions and contractions

Mukherjee (2009) also examined the grammar of conversation in spoken and written corpora of German advanced learners and compared it with that of native speakers. The author examined three case studies of lexico-grammatical structures used in advanced German learners' spoken corpus of English. Like in the study done by Nesselhauf (2003), Mukharjee (2009) also examined the German component of ICLE for written data

(GeCLE). For spoken data Mukharjee (2009) analysed German component of Louvin International Database of Spoken English Interlanguage (LINDSEI-Ger).

The author investigated verb-noun collocations, the discourse marker '*you know*' and performance phenomena such as repetitions and contractions. As it was also studied by De Cock (2004), Mukharjee (2009) has identified that native speakers tend to frequently use repetitions in their speech.

The author found out that the most frequent collocations in GeCLE (e.g. *'solve a problem', 'have a reason', 'commit a crime'*) did not appear in LINDSEI-Ger at all. Whereas, the most frequent collocations in LINDSEI-Ger (e.g. *'have time', 'have a problem', 'have a chance*') occurred less frequently in LINDSEI-Ger. Moreover, the topmost items in LINDSEI-Ger occurred more frequently than those in GeCLE. The author also points out that German nonnative speakers of English use orally performed written English in their speech. The author believes that spoken language tends to be 1) more formulaic and less varied and 2) more clausal with fewer verb-noun collocations. In the learner corpus there are also deviations from the native speaker spoken corpus ('*has*/makes some sense', 'talk*/tell a story', 'made*/had the experience'*). Learners tend to premodify nouns with adjectives in verb-noun collocations too, which is communicatively unnecessary, syntactically undesirable and semantically incompatible in the context (e.g. '*have a quite big choice', 'doing some little homework'*).

As to the use of the discourse marker '*you know*', it seems that learners underuse it in their speech. Although of all transcriptions of 50 interviewees the half did not use '*you know*' as a discourse marker, in the other half there were cases of quite effective use of it. Learners used '*you know*' discourse marker 1) to hesitate in the flow, 2) at the beginning

of the clause, 3) when thinking about the following statement, 4) to show impreciseness of those words, and 5) to explain the previously expressed idea.

Mukherjee (2009) has examined the occurrence of repetitions and disfluencies in the data and found that learners use repetitions less frequently. Although both native speakers and learners mostly repeat pronouns, in contrast with learners, native speakers use repetitions of '*I*' at the beginning of the clause without further hesitation phenomenon. Whereas, learners have further disfluencies after repetitions, which can mean that speech planning pressure does not decrease with the help of repetitions in nonnative speaker speech. Moreover, learners seem to have more unfilled pauses than native speakers.

Like No and Park (2010), Mukherjee (2009) has also found that contractions in spoken language are used by native speakers of English more often. Moreover, learners appear to underuse subject-verb contractions which seems to be oriented towards written grammar.

2.2.4 Formulaic sequences used as idioms

The researchers have conducted the experiments to identify fluencies of the learners and native speakers of English in processing or recognizing formulaic sequences (Conklin & Schmitt, 2008; Jiang & Nekrasova, 2007; Ellis, Simpson-Vlach & Maynard, 2008). The researchers compared the participants' reaction times for the two types of formulaic sequences in English.

Jiang and Nekrasova (2007) have investigated 20 native speakers and 20 nonnative speakers of English for the reaction times to formulaic and nonformulaic sequences (13 for each category) in a phrase judgment task. The participants were undergraduate and graduate students at Georgia State University (25 females and 15 males, aged from 20 to

40). The nonnative speaker participants were of different linguistic backgrounds (Arabic, Bulgarian, Chinese, Czech, Korean, Polish, Portuguese, Romanian, Russian, Thai, and Turkish). They were relatively proficient speakers of English based on the test scores (about 213 points for the computer-based Test of English as a Foreign Language (TOEFL) or 550 points for the paper test PB TOEFL). For getting nonformulaic sets, one word was changed in the formula (e.g. '*as soon as' – 'as means as'*). The participants were to respond to the sets of word sequences as being 'correct' or 'incorrect'. The authors concluded that both native speakers and nonnative speakers of English responded to formulaic phrases faster and with fewer errors than to nonformulaic controls.

Based on the findings of their research Ellis, Simpson-Vlach and Maynard (2008) have concluded that different aspects of formulaicity seem to affect the accuracy and fluency in the processing of those formulas by native speakers and learners of English. The aspects chosen for the investigation of the study were the length, frequency and mutual information of the word-sequences. Three-, four-, and five-word formulas were extracted from the corpora of 2.1 million words of academic spoken language samples from Michigan Institute Corpus of Academic Spoken English (MICASE), 2.1 academic spoken and written language samples from the British National Corpus (BNC), 2.9 million words of nonacademic speech from the Switchboard corpus, and 1.9 million words of nonacademic writings from the Freiburg Lancaster Oslo/Bergen (FLOB) corpus. There were 11 (7 females and 4 males, aged average 23.4 years old) English native speaker students or staff from the University of Michigan. The learners of English were 11 (6 females and 5 males, aged average 31.3 years old) international students at the University of Michigan who were taking English for Academic Purposes classes. The linguistic

backgrounds of nonnative speakers of English was various; Chinese (5), Thai (4), Korean (1), and Spanish (1). The learners' level of proficiency in English was adequate to graduate study through English as the language of instruction. They had studied English for about 15 years. The judgments of the formulas were to be made according to the length, frequency, and mutual information of the sequences. The participants of the study were asked to read out loud and judge whether visually presented word strings were likely to exist in English or not (e.g. *'on phone the', 'by way the', 'put on shirt his'*). The sequences were shown on a computer screen one at a time. For the native speakers mutual information of the strings (which is the extent to which the elements of the formula cohere) explained the response times; whereas, for the learners it was the frequency of the string which appeared on the screen that determined the time of their responses. The results showed that 1) the longer the formulaic sequence was, the longer time it took for the participants to judge and 2) the stronger the coherence of the formula, the shorter the judgment time.

Conklin and Schmitt (2008) conducted research on the processing of idioms. 19 English native speakers in their undergraduate study at the University of Nottingham and 20 nonnative speakers (mainly graduates MA-ELT of mixed-linguistic background) from the same university took part in the study. The participants were to read formulaic and nonformulaic sequences (e.g. '*everything but the kitchen sink' – 'everything in the kitchen sink'*) provided in contexts. The participants' reading time was to be measured. The formaulaic sequences held idiomatic and literal meanings. The main content words both in formulaic and nonformulaic word-sequences were kept unchanged (e.g. '*hit the nail on the head' – 'hit his head on the nail'*) as it was done by Jiang and Nekrasova (2007). The

authors concluded that both native and nonnative speakers read the formulaic sequences faster than the equivalent non-formulaic ones. It is interesting that regardless the idiomatic or literal meanings, the formulaic sequences were processed equally faster, as it was also supported by Jiang and Nekrasova (2007). Thus, it can be implied that formulaic sequences with both idiomatic and literal meanings seem to be processed by the users of English better than non-formulaic ones.

2.3 Summary of the Literature Review

Having reviewed the findings of previous studies, conducted on spoken and written corpora of learners and native speakers of English, a solid background for the investigation of the present research study has been established. The following linguistic features have been preliminary selected to be investigated in the current study:

1) *fillers* - It has been examined in the literature studies that the learners of English, who are speakers of other languages, had some deficiencies in the use of certain types of multi-word strings that employed different functions, such as for example, discourse marking and gap-filling. Thus, for the current study, certain multi-word strings that employ the function of fillers have been selected to investigate;

2) *collocations* - It seems that learners and native speaker of English have differences in the use of *collocations,* to express the same ideas. So, the other potential aspect of the spoken corpora that can be investigated in both corpora in the scope of the present research nay be collocations.

CHAPTER THREE: METHODOLOGY

Introduction

The major objective of the study was to investigate the relationship between certain LFs that occurred in spoken English of the Armenian learners of English and their ELEs. To find out the relationship, first the NS and NNS spoken corpora were recorded, transcribed and compared with each other. Then, the relationship between the composite scores of the total LFs was calculated for NSs and NNSs. Afterwards, the relationship between the composite scores and NNSs' ELEs was investigated. The details of the research methodology, including the description of the participants, confidentiality issue, sampling, instruments and procedure, as well as the data analysis are presented in the current chapter.

3.1 Restatement of the Research Questions

Two research questions are raised to address the main goal of the study. They are listed as follows:

Research Question 1

Is there a difference between the linguistic features in spoken English of Armenian nonnative speakers and American native speakers of English?

Research Question 2

What is the relationship between the English learning experiences of Armenian nonnative speakers of English and the linguistic features of their spoken English?

3.2 Research Design

The research is a corpus-based study of two corpora (NS and NNS). Both the first and the second research questions were answered through quantitative analyses.

3.3 Participants

3.3.1 Sampling

The sample of the participants of the study was selected according to the random sampling approach (Fraenkel & Wallen, 2009). Certain criteria were taken into account for the selection of the sample from the population.

3.3.2 The target group of NSs

The target group of native speakers consisted of 10 participants selected according to the following criteria of the random sampling principle:

1) the native language of the participants was American English;

2) the participants were current students or faculty at AUA and Peace Corps Volunteers (PCVs) in Armenia.

3) the participants' age was 20 years old and above.

3.3.3 The target group of NNSs

The target group of nonnative speakers consisted of 24 participants selected according to the following criteria of random sampling approach:

1) the participants were EFL learners of English;

2) the first language of the participants was Armenian;

3) the participants were first and second year graduate students at AUA;

4) the age of the participants was 20 and above;

5) the participants' level of English proficiency was established based on their TOEFL IBT entrance scores with 69 cut-off point and higher;

The scores of the speaking section of the TOEFL IBT scores of the NNS participants were particularly taken into consideration. For the purpose of data stratification, the sample was divided into three groups within the speaking score ranges of 18-21, 22-25 and 26-29. Eight students were selected randomly within each speaking score range, so that overall the sample could consist of 24 students with the scores of TOEFL IBT speaking section of 18-29.

3.4 Confidentiality

The participants were assured about the confidential use of their responses to be collected for the investigation within the scope of the present study. They were orally informed about the confidentiality of the data use.

3.5 Instrumentation and Procedure

Two instruments of data collection, namely the interview and the questionnaire were used to answer both research questions of the study (Appendices 2 and 3). The instruments and procedures were piloted and adjusted beforehand (Appendix 1), and the decisions were made having considered the piloting data as well.

The first instrument was used to collect the data about the number of occurrences of certain LFs of spoken English of the participants and to answer the first research question. The results of the interview along with those of the questionnaire aimed at

investigating the relationship between the ELEs of the NNS participants and the LFs of their spoken English to answer the second research question.

3.5.1 Piloting session

The piloting interview session was conducted on 5 students at AUA. Four of the five participants were NNSs (second year TEFL students at AUA) and 1 NS (a first year TEFL student at AUA) of English. The data of 2 participants (1 NS and 1 NNS) of the piloting session were later on used in the study for the analysis. The piloting session aimed at 1) selecting certain speaking tasks for the interviews that could elicit distinguishing LFs of spoken English and 2) adjusting the questions of the questionnaire (Appendix 1). The results of the piloting session showed that among the experimented speaking tasks, six tasks appeared to elicit more representative LFs. Consequently, those six tasks were selected for the interviews.

3.5.2 Interview

The first instrument was a structured interview with six speaking tasks (Appendix 2). The speaking tasks of the interview were the following:

1. Questions about the participants' foreign language learning experience
2. Picture description (beach volleyball in action)
3. Picture story (an ill men story)
4. Video episode summary (a new car-plane developed by students at Massachusetts Institute of Technology)
5. Preference question (ways of communication)
6. Simulation task (travel agent and customer conversation)

For the first question of the interview the NS participants talked about any foreign language they learnt; whereas, the NNSs talked about their ELEs. The interviews were conducted with each participant (NS and NNS) separately, recorded, transcribed manually, and analyzed. The interview recordings have been transcribed into two separate corpora. In the beginning of the interviews, the native speakers were asked to tell about the period of their stay in Armenia, their place of birth in the United States and the number of years spent in the US. The interview tasks were introduced to the participants beforehand, for them to have better understanding of the nature of the questions. The time of the video episode was not calculated in the data analysis.

3.5.3 Questionnaire

The second instrument of the research was the questionnaire that aimed at eliciting certain details about ELEs of the nonnative speaker participants of the research (Appendix 3). It was given to each NNS at the end of the interview session. The questionnaire consisted of 16 questions about different types of English exposure (Table 1). All the questions were to be answered across all the types of experiences (secondary, undergraduate, graduate and postgraduate schools, as well as self and private study).

For the first three questions the number of years and hours was calculated across the schools separately. Questions 2 and 3 were combined into one cumulative score regarding the quantity of English exposure. Each point for question 8 in the questionnaire was considered a separate question in the data analyses. Likert scale was used to quantify the responses to questions 4-8 of the questionnaire, counting from the highest '5' points for 'Always' and the lowest '0' point for 'Never'.

Table 1

Questions of the questionnaire for the NNSs

	Questions about the quality of exposure
Question 1	living in an English speaking country (in months)
Question 2	quantity of exposure of English (in hours)
	Questions about the quality of exposure (Likert scale)
Question 3	frequency of use of authentic video / audio materials;
Question 4	frequency of interaction with native speakers of English on a regular basis
Question 5	frequency of use of textbooks from international publishers (e.g. Cambridge, McMillan, Oxford, etc.)
Question 6	frequency of use of textbooks published in non-international / local publishers (e.g. Russian or Armenian publications, N. A. Bonk, etc.);
Question 7	reading and summarizing orally in your own words
Question 8	reading and translating texts
Question 9	watching and listening to recordings in English and summarizing them orally in your own words
Question 10	watching and listening to recordings in English and repeating them word for word
Question 11	writing grammar exercises (e.g. sentence completion, multiple-choice, gap-filling, etc.)
Question 12	writing coppositions / essays
Question 13	learning vocabulary lists by heart
Question 14	giving oral presentations
Question 15	discussions / debates
Question 16	role plays

3.5.4 Linguistic features of the investigation and data coding

Although preliminary selection of linguistic features was based on the findings of the literature review and those of the piloting session, the final set of target LFs to

investigate was selected inductively. In other words, the target LFs were drawn from the native speaker corpus data. Some of the linguistic features (collocations and fillers) were supported by the literature findings, too. As a result, the following 6 categories of the target LFs of spoken English have been identified (Table 2):

1. fillers – sounds, words or phrases (such as 'you know', 'like', 'stuff like that') used to fill pauses in speaking, which may not have semantic content of their own but may fulfill important linguistic function; and can be used to "buy time" or keep the floor, rather than to leave a pause, when deciding what to say next (Jabeen, Rai, & Arif, 2011; Fillers, n.d.);

2. modals – auxiliary verbs that express necessity or possibility (Modals, n.d.);

3. adverbs - word or phrases that modify the meaning of an adjective, verb, or other adverb, expressing manner, place, time, or degree (Adverbs, n.d.);

4. collocations - words that appear in context with other words with greater than random possibility, which are fixed and highly predictable from one of the components (Mukharjee, 2009; Phoocharoensil, 2011; McCarthy & Carter 2006b);

5. Phrasal verbs - phrases that contain a verb with a preposition or adverb or both and function as verbs whose meaning is different from the combined meanings of the individual words (Phrasal verbs, n.d.);

6. Grammatical structures - sets of actual or presumed prescriptive notions about correct use of a language (Grammar, n.d.).

Table 2

Target LFs per category (with examples from the NS corpora)

Category 1	Fillers		
Subcategory 1	any/something(s) like this/that, and stuff (like that), kind/sort of, this/that sort/king of thing, (it) seems to me/that/like, (it) looks (to me) that/like, (sounds) like, like that		
Subcategory 2	I mean , You know, Well		
Subcategory 3	I guess, I'm guessing		
Subcategory 4	I wonder, I'm wondering		
Category 2	Modals		
	Would Could	Subcategory 1	Infinitive (Base Verb) *e.g. could continue*
	Might Should	Subcategory 2	Perfect infinitive *e.g. should have listened to*
	May Will	Subcategory 3	Progressive infinitive *e.g. will you be getting*
	Must to be supposed to gonna (to be going to) gotta (have got to)	Subcategory 4	Have to/need *e.g. will have to*
Category 3	Adverbs		
Subcategory 1	probably		
Subcategory 2	Maybe		
Subcategory 3	Actually		
Subcategory 4	Typically		
Subcategory 5	Obviously		
Subcategory 6	Most likely		
Subcategory 7	Mostly		
Subcategory 8	Ultimately		
Subcategory 9	Basically		
Subcategory 10	Definitely		
Category 4	Collocations		
Subcategory 1	The very first		
Subcategory 2	Lets smbd/smth + base verb *e.g. let's see, let me get on the computer*		
Subcategory 3	as adj. as *e.g. as late as*		
Subcategory 4	Pretty much / nice		
Subcategory 5	Plenty / a bunch of		
Subcategory 6	That much / adj. *e.g. that practical*		
Subcategory 7	To me		
Subcategory 8	On/over the phone		

Subcategory 9	3 people on each side/team		
Subcategory 10	In the background, On the foreground		
Subcategory 11	Via/in an ambulance / Ambulance comes		
Subcategory 12	At/on/by the beach		
Subcategory 13	Offer a , Special, Package, Have - deal		
Subcategory 14	Never , Do you , Would you - mind		
Subcategory 15	(Price) range		
Subcategory 16	Go to (the/a) doctor		
Subcategory 17	So that , in that , as if		
Subcategory 18	Twice /…times as adj. as *e.g. twice as fast as*		
Category 5	Phrasal verbs		
Subcategory 1		get	Back to, On, Out of, In(to), By on, across,
Subcategory 2		put	up, up with, up in, aside, on, off, smbd down for
Subcategory 3		come	down to, out of, up with, through, over
Subcategory 4		go	Over, into, ahead, on
Subcategory 5	Pick/ print / send / pay / Keep / Wrap / Pop / lift / Crash / Show / Break / End / Care / Move / Drive / Sit / Hang / Figure / Find / Fly / Stuck / Fit / set / Turn / Pull		up / down / off / on / in / with / through / out / into/ up with / in touch with / into / up into / over/ up / ahead / aside
Category 6	Grammatical Structures		
Subcategory 1	She/I does/did like		
Subcategory 2	causative verbs: Make Get Have		

3.6 Data analysis

To address the main goal of the research, the data were analysed quantitatively via the spreadsheet software Microsoft Excel and the Statistical Package of Social Sciences (SPSS version 16). The descriptive statistics of means and standard deviations was calculated to estimate normal distribution of the interview times and word counts for the NSs and NNSs (Table 3). Also the descriptive statistics for the total numbers of the LFs in

the NS and NNS corpora was estimated (for more information refer to Table 4 in Chapter 4). Also, the average percentages of the LFs employed by NS and NNS participants per person have been calculated (Tables 5-10, Figures 1-6). The composite scores of the LFs used by NNSs served as the dependent variable and the ELEs served independent variables for the analysis of Spearman correlation and multiple regressions (Tables 11-15).

3.6.1 NS and NNS corpora

Means (M) and standard deviations (SD) for the time and the number of words in the NS and NNS corpora were calculated (Tables 3). It appeared that the ranges of the interview duration and the number of words produced in that time for NS and NNS were not so different from each other.

Table 3

Descriptive statistics for the interviews for NSs and NNSs

Corpus	*M (SD)*		Total corpus		N	
	NS	*NNS*	*NS*	*NNS*	*NS*	*NNS*
Word count	1,945(523)	1,272(490)	19,446	30,524	10	24
Length (in minutes)	15.7(2.99)	13(3.49)	157,27	313,11		

3.6.2 Data analysis to address the first research question

To find out to what extent the uses of LFs by NSs and NNSs were different, descriptive statistics for means and standard deviations was estimated (Table 4). Then, the two corpora (NS and NNS) were compared to find out differences in the average percentages of LFs occurred in the NS and NNS speech per person (Tables 5-10). Analysis of the comparison was done by Microsoft Excel software through bar graphs, too (Figures 1-6).

3.6.3 Data analysis to address the second research question

For the second research question composite scores of the LFs per each participant were calculated (Table 12). The composite scores were used as the dependent variables (DV) to run Spearman bivariate two-tailed correlation and multiple regressions with the questions of the questionnaire about the ELEs of the NNSs as the independent variables (IV) (Table 13, 14, 15).

CHAPTER FOUR: RESULTS

Introduction

This chapter presents the results of the data analysis for the two research questions separately. The NS and NNS corpora were compared for the average percentages of the LFs used by NS and NNS participants per person to find out differences. Furthermore, the correlation and multiple regression analyses were calculated to find out the relationship between the NNSs' ELEs and the total number of LFs used by them as a composite score per each participant.

4.1 Analysis of the results for the first research question

4.1.1 Descriptive statistics

The descriptive statistics for NS and NNS groups for the LFs used in their speech was calculated. The calculation did not include the percentages of LFs of Category 2 Subcategory 1 ('modal + base verb') and Category 3 Subcategory 2 ('maybe') because of the similar percentages for NSs and NNSs in their speech. The analysis shows that the LFs chosen for the comparison of the two corpora (NS and NNS) do seem to be distinguishing cornerstones for native-like competence because NSs used the target LFs more than three times as frequently as the NNSs did (Table 4).

Table 4

Descriptive Statistics for NS and NNS corpora

Category	N of participants		Total N of LFs		Mean (SD)	
	NS	NNS	NS	NNS	NS	NNS
Total Fillers	10	24	227.00	143.00	22.7 (15.3)	5.96 (6.9)
Total Modals	10	24	30.00	3.00	3.0 (2.4)	.13 (.34)
Total Adverbs	10	24	90.00	95.00	9.0 (5.8)	3.99 (4.9)
Total Collocations	10	24	136.00	93.00	13.6 (4.4)	3.9 (3.7)
Total Phrasal Verbs	10	24	47.00	31.00	4.7 (2.5)	1.3 (2.4)
Total Structure	10	24	11.00	1.00	1.1 (1.1)	.04 (.2)
Total Categories	10	24	541.00	366.00	54.1 (25.3)	15.3 (13.3)

4.1.2 Average percentage of the use of LFs by NSs and NNSs per person

To answer the first research question, the average percentage of the use of certain LFs across the six categories was calculated for NSs and NNSs. The percentage are presented per Subcategory for each category both in tables (Tables 5-10) and bar graphs (Figures 1-5). Overall across all the 6 categories it can be observed that NSs had higher frequencies of the use of the target LFs. Moreover, in some cases NNSs not only used LFs with low frequency but also used inaccurately or differently.

4.1.2.1 Average percentage of the LFs for Category 1 "Fillers"

As it can be seen from Table 5 and Figure 1, NSs used fillers overall about four times as much as NNSs did. Mostly the NNSs appeared to use LFs such as *'any/something(s) like this/that', 'and stuff (like that)', 'kind/sort of', 'this/that sort/king of thing', '(it) seems to me/that/like', '(it) looks (to me) that/like', '(sounds) like'* and *'like that'* in their speech with low frequency. It is also obvious that the filler *'I wonder or I'm wondering'* was used only by NSs. The use of *'I mean', 'You know'*, and *'Well'* was twice

as frequent in NS corpus as in that of the NNSs. It seems that both NSs and NNSs used the filler *'I guess or I'm guessing'* not so frequently, and the difference was also not so apparent.

Table 5

Average percentage of the use of LFs for Category 1 "Fillers"

Category 1	Examples	NS	NNS same	NNS different
Subcategory 1	any/something(s) like this/that, and stuff (like that), kind/sort of, this/that sort/king of thing, (it) seems to me/that/like, (it) looks (to me) that/like, (sounds) like, like that	13.8	3.125	
Subcategory 2	I mean , You know, Well	7.3	2.46	
Subcategory 3	I guess, I'm guessing	1.8	1.17	
Subcategory 4	I wonder, I'm wondering	0.3	0	
TOTAL		23.2	6.755	

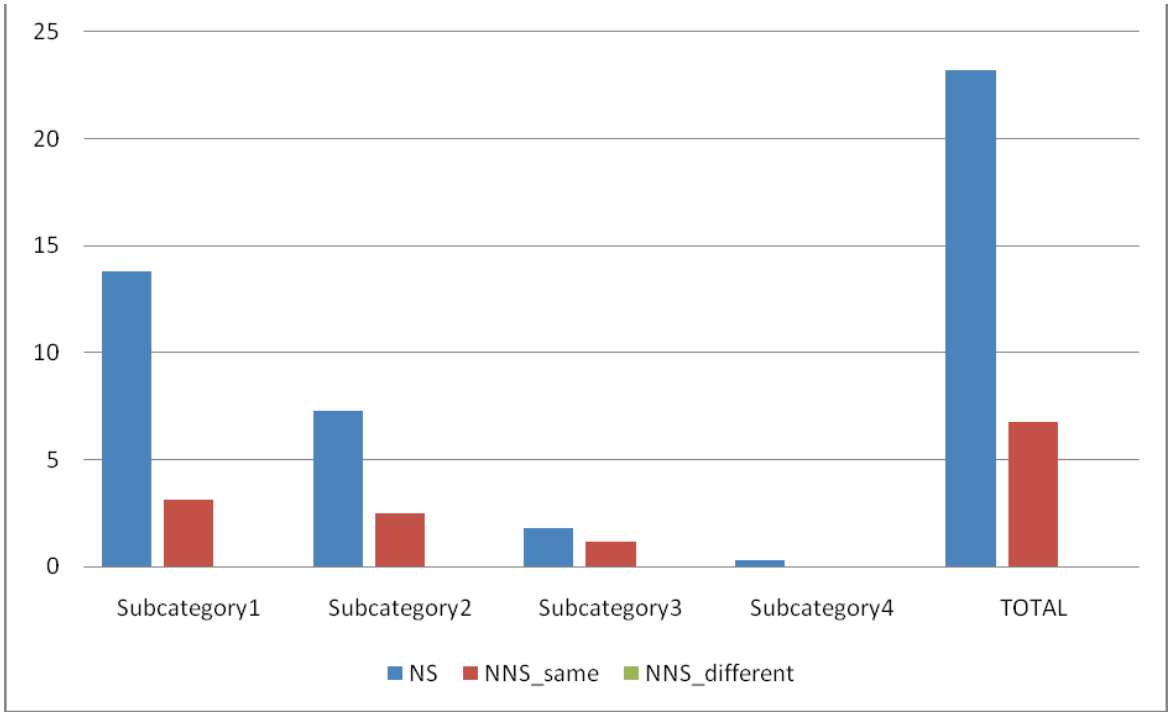

Figure 1. Average percentage of the use of LFs for Category 1 "Fillers"

4.1.2.2 Average percentage of the use of LFs for Category 2 "Modals"

As it can be observed from the data, the difference of the use of modals between NSs and NNSs is not so significant. However, there are differences at the level of the specific subcategories. In general, the most frequently used items both by NSs and NNSs are modals with base verbs. It is also worth mentioning that the difference between the use of item of this subcategory by NSs and NNS is slight. NNSs seemed to use progressive infinitive, perfect infinitive or have to/need with modals, but the percentage of their use by NNSs is not so high (Table 6 and Figure 2).

Table 6

Average percentage of the use of LFs for Category 2 "Modals"

Category 2	examples	NS	NNS same	NNS different
Subcategory 1	Infinitive (Base Verb)	29	22.625	
Subcategory 2	Perfect infinitive	1.5	0.125	
Subcategory 3	Progressive infinitive	0.7	0.42	
Subcategory 4	Have to/need	0.9	0.083	
TOTAL		32.1	22.833	

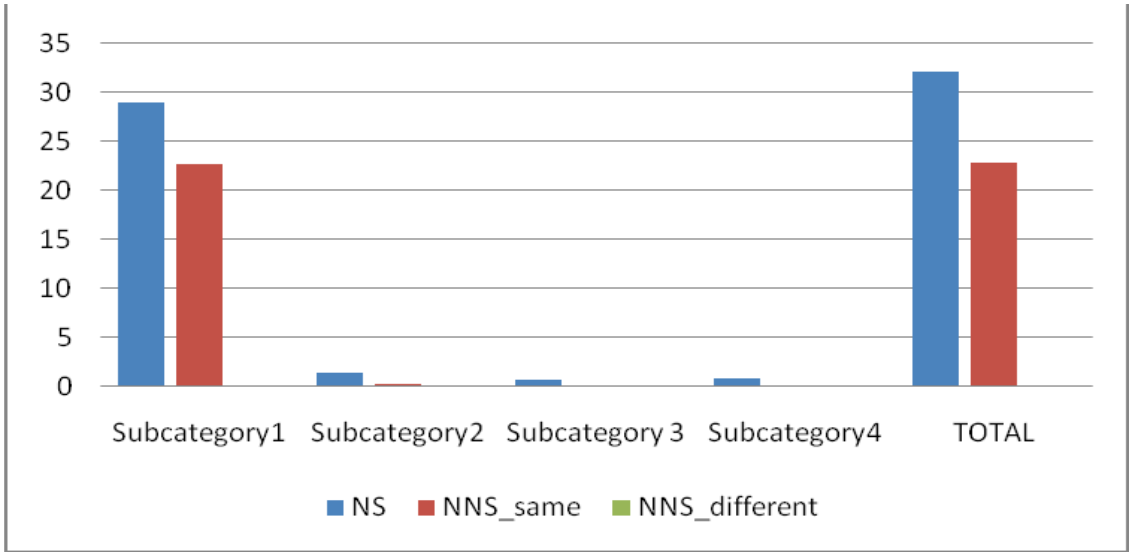

Figure 2. Average percentage of the use of LFs for Category 2 "Modals"

4.1.2.3 Average percentage of the use of LFs for Category 3 "Adverbs"

It has been noticed that the NSs seem to use more adverbs than NNSs in general, but items of some subcategories such as '*probably'* and '*obviously'* are considerably more frequently used by NSs than NNSs. However, there is also one adverb that seems to be more preferably used by NNS, which is '*mostly',* and the adverbs *'typically'* and *'ultimately'* are only used by NSs. The adverb *'maybe'* is use by both NSs and NNSs with similar percentages. The other adverbs do not appear to have obvious differences in the use by NSs or NNSs (Table 7 and Figure 3).

Table 7

Average percentage of the use of LFs for Category 3 "Adverbs"

Category 3	Examples	NS	NNS same	NNS different
Subcategory 1	probably	3.3	0.33	
Subcategory 2	Maybe	3.5	2.92	
Subcategory 3	actually	2.9	2.08	
Subcategory 4	Typically	0.3	0	
Subcategory 5	Obviously	1.1	0.17	
Subcategory 6	Most likely	0.2	0.042	
Subcategory 7	Mostly	0	1.125	
Subcategory 8	ultimately	0.1	0	
Subcategory 9	Basically	0.3	0.292	
Subcategory 10	Definitely	0.3	0.208	
TOTAL		12	7.167	

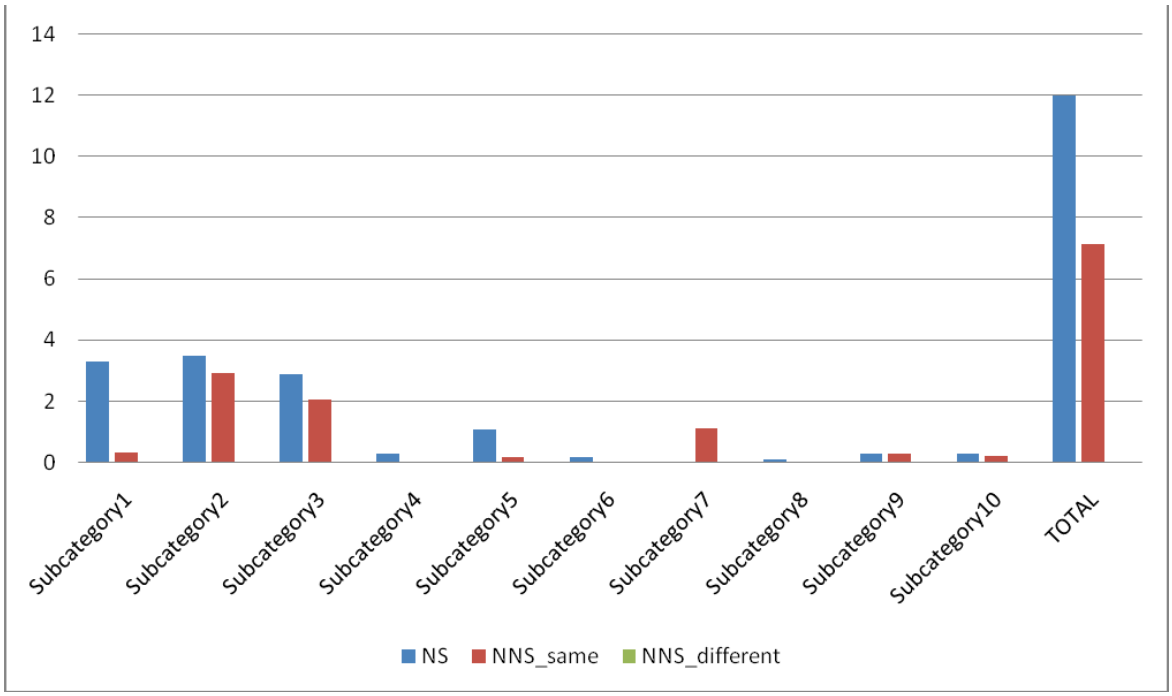

Figure 3. Average percentage of the use of LFs for Category 3 "Adverbs"

4.1.2.4 Average percentage of the use of LFs for Category 4 "Collocations"

It can be seen from the data that NSs appear to use collocations more frequently with the difference of almost 3 times as much. However, it is also evident that NNSs have differences in the correct use of the LFs of this category. More precisely, NNSs may use the same collocations either inaccurately (*'see, visit or meet a doctor'* instead of *'go to the doctor'*) or differently (*'hang around'* instead of *'hang out'* or *'each group has 3 members'* instead of *'3 people on each team'*). The percentage of the accurate use of the same LFs seems to be twice as frequent as the inaccurate or different use. Moreover, some participants used the same LF both correctly and incorrectly. For example *'see a doctor', 'visit a doctor'* and *'go to the doctor'* could have been used by the same participant. It seems that the LFs '*the very first'*, *'plenty/a bunch of'* and '*never/ do you/ would you mind'* are not used by the NNSs at all. In the uses of *'On/over the phone', '3 people on*

each side/team', 'In the background, On the foreground', 'Via/in an ambulance or Ambulance comes', 'At/on/by the beach', '(Price) range', 'Go to (the) doctor', 'twice/…times as adjective as' NNSs have differences or inaccuracies. It is observed that the NSs use '*let'* and '*as adj. as'* significantly more frequently than NNSs (Table 8 and Figure 4).

Table 8

Average percentage of the use of LFs for Category 4 "Collocations"

Category 4		NS	NNS same	NNS different
Subcategory 1	The very first	0.4	0	
Subcategory 2	Let smb/smth BaseVerb	3.2	0.83	
Subcategory 3	as-long,much,adj. as	2.1	0.083	
Subcategory 4	Pretty - much,do,nice	1.1	0.625	
Subcategory 5	Plenty / A bunch of	0.5	0	
Subcategory 6	That much , That adj.	1.1	0.17	
Subcategory 7	To me	0.6	0.125	
Subcategory 8	On/over the phone	0.5	0.042	0.042
Subcategory 9	3 people on each side/team	0.2	0.042	0.208
Subcategory 10	In the background, On the foreground	0.7	0.29	0.083
Subcategory 11	Via/in an ambulance, Ambulance comes	0.5	0.42	0.292
Subcategory 12	At/on/by the beach	0.5	0.46	0.17
Subcategory 13	Deal-Offer a , Special, Package, Have	0.7	0.042	
Subcategory 14	Mind - Never , Do you , Would you	0.3	0	
Subcategory 15	(Price) range	0.5	0.083	0.042
Subcategory 16	Go to the doctor	1.3	1.125	0.83
Subcategory 17	so that, in that, as if	0.6	0.29	
Subcategory 18	twice/…times as adj.as	0.1	0.042	0.29
TOTAL		14.2	4.337	1.667

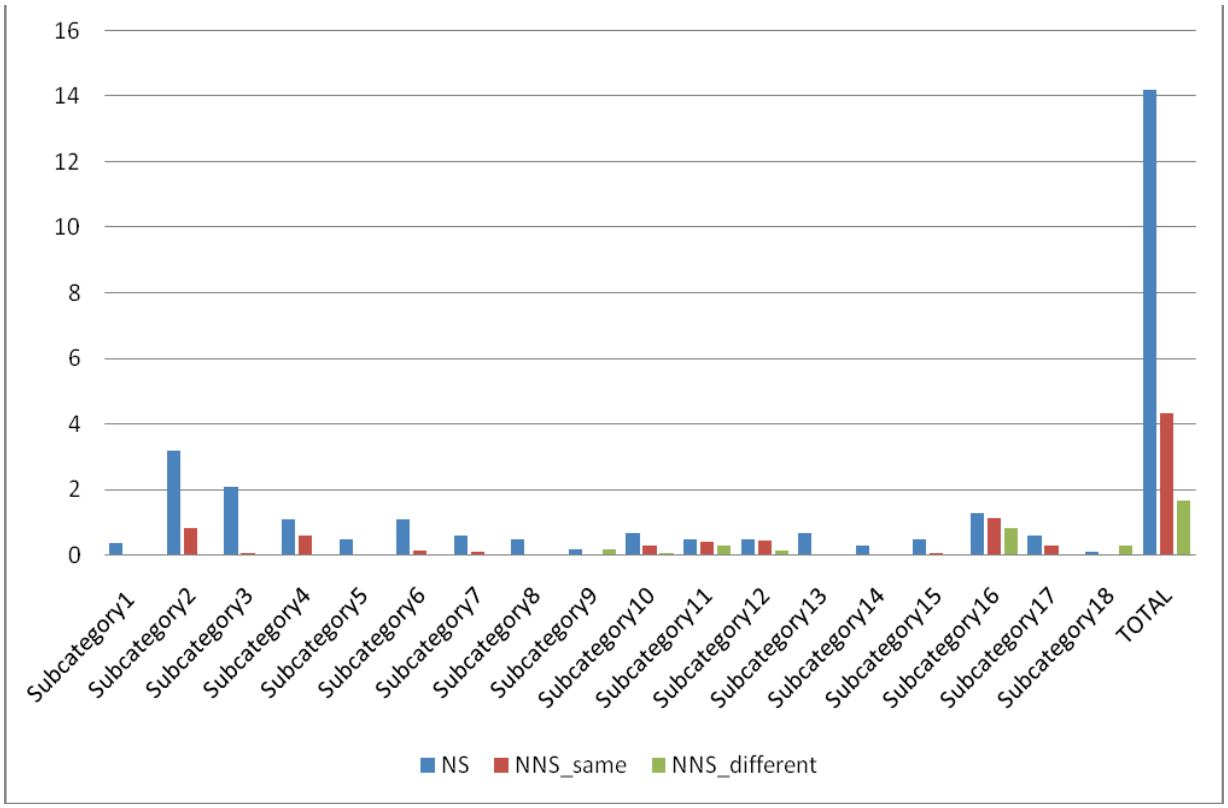

Figure 4. Average percentage of the use of LFs for Category 4 "Collocations"

4.1.2.5 Average percentage of the use of LFs for Category 5 "Phrasal Verbs"

Generally, phrasal verbs are also more frequently used by NSs than NNSs. Moreover, it seems noteworthy that the verbs *'put'* and *'get'* are not favorably used by NNSs. In the use of different verbs with different prepositions NSs show more frequency almost twice as much as the NNSs. In some of the cases in the use of the latter LF NSs seem to have inaccuracies, too (Table 9 and Figure 5).

Table 9

Average percentage of the use of LFs for Category 5 "Phrasal Verbs"

Category 5		NS	NNS same	NNS different
Subcategory 1	get-Back to , On, Out of , In(to) , By on, across	1.2	0.083	
Subcategory 2	put-up, up with, smbd in, aside, on, off, smbd down for	0.4	0.083	
Subcategory 3	come-down to, out of, up with, through, over	0.5	0.083	
Subcategory 4	go-Over, into, ahead, -ing on	0.8	0.33	
Subcategory 5	print/send/pay/Keep/Wrap/Pop/lift/ Crash/Show/Break/End/Care/Move/ Drive/Sit/Hang/Figure/Find/Fly/Stu ck//set/Turn+up / down / off / on / with / through / out / up with / in touch with / into / up into / over	2.2	1.042	0.042
TOTAL		5.1	1.621	0.042

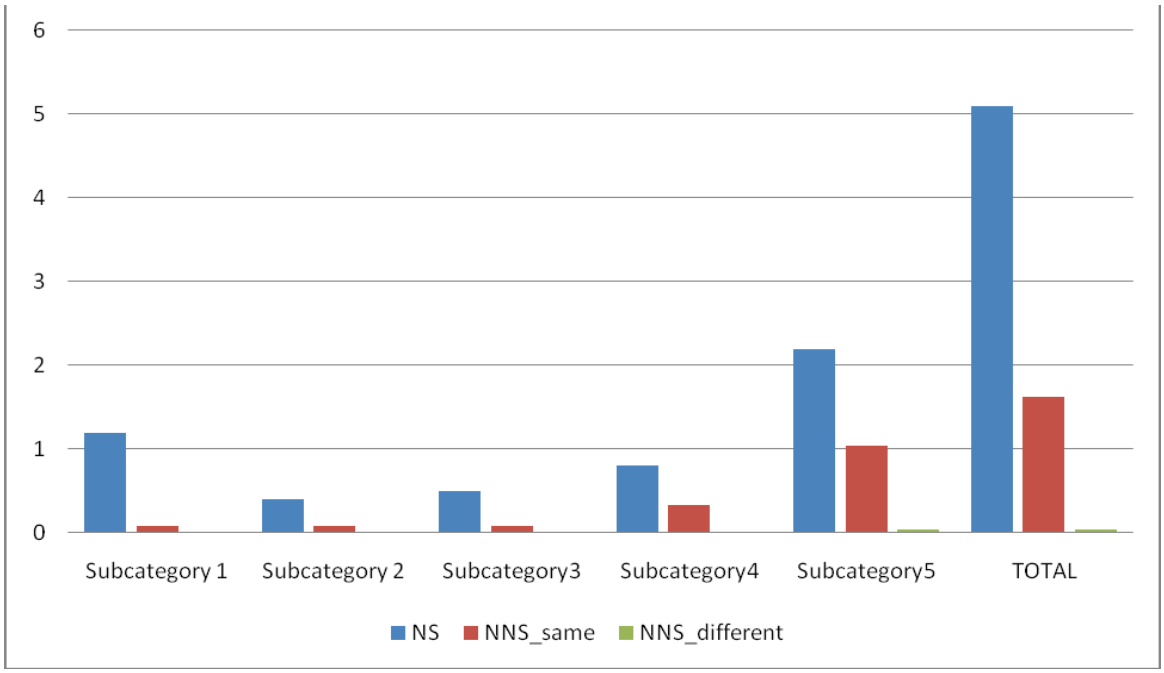

Figure 5. Average percentage of the use of LFs for Category 5 "Phrasal Verbs"

4.1.2.6 Average percentage of the use of LFs for Category 6 "Structure"

Overall, out of the two selected grammatical structures, NNSs appear not to use one at all (auxiliary emphasis). In the use of the other (causative verbs) they seem to have equal number of accurate and inaccurate uses (Table 10 and Figure 6).

Table 10

Average percentage of the use of LFs for Category 6 "Structure"

Category 6		NS	NNS same	NNS different
Subcategory 1	auxiliary emphasis (I do like smth)	0.2	0	
Subcategory 2	causative verbs (make, get, have)	0.7	0.042	0.042
TOTAL		0.9	0.042	0.042

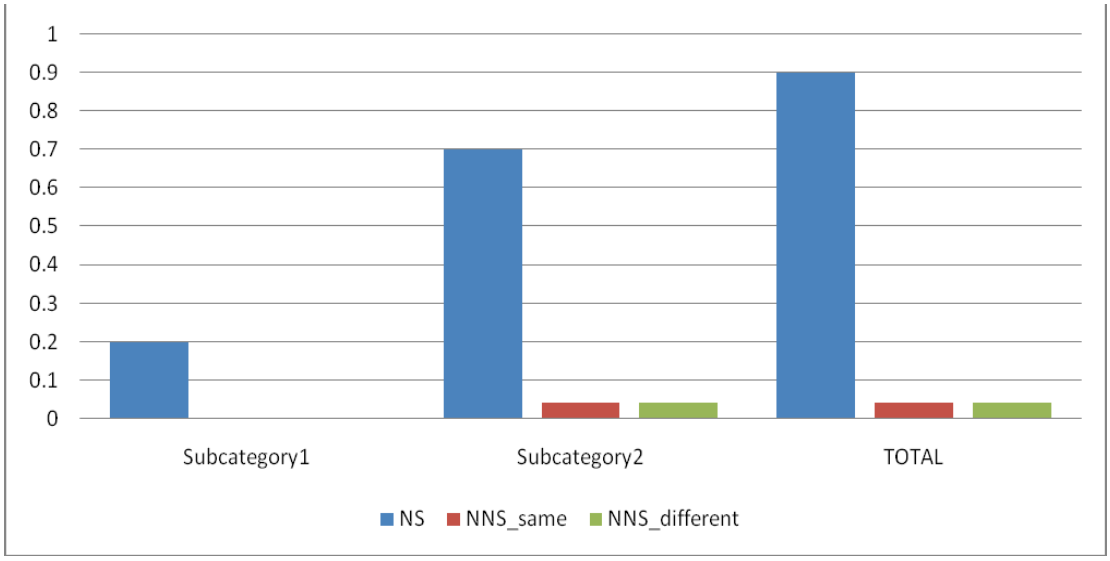

Figure 6. Average percentage of the use of LFs for Category 6 "Structure"

4.2 Analysis of the results for the second research question

To address the second research question, the Spearman bivariate 2-tailed correlations were calculated. Then, multiple regressions were performed with the selected IVs and the composite score of LFs as the DV.

4.2.1 Calculation of composite scores for NNSs

The computation of the composite scores was based on LFs of the Subcategories selected only using NS corpus. The number of the LFs included into the computation of the composite scores was 22 out of 43 (Table 11). The Subcategories were selected based on the following criteria:

1) the difference between the average percentage of the use of LFs per person for subcategories for NS and NNS should be approximately twice as many (Tables 5-10);

2) the number of occurrences of the LFs in either NS or NNS group should be at least 5.

Table 11

List of Subcategories included in the composite score

	Category	Example
1	Fillers Subcategory 1	kind of, stuff/things like that
2	Fillers Subcategory 2	I mean, you know, well
3	Fillers Subcategory 3	I guess
4	Modals Subcategory 2	Modal+perfect infinitive
5	Modals Subcategory 3	Modal+progressive infinitive
6	Modals Subcategory 4	Modal + have to/need
7	Adverbs Subcategory 1	probably
8	Adverbs Subcategory 5	obviously
9	Collocations Subcategory 2	Let+smbd/smth+base verb
10	Collocations Subcategory 3	As adj.as
11	Collocations Subcategory 4	Pretty much/adj
12	Collocations Subcategory 6	That adj.
13	Collocations Subcategory 8	On the phone
14	Collocations Subcategory 10	In the background
15	Collocations Subcategory 13	Deal-offer
16	Collocations Subcategory 15	Price range
17	Phrasal Verbs Subcategory 1	get+ preposition
18	Phrasal Verbs Subcategory 3	Come +prepositions
19	Phrasal Verbs Subcategory 4	Go+ preposition
20	Phrasal Verbs Subcategory 5	Verbs+ preposition
21	Structure Subcategory 1	One does like
22	Structure Subcategory 2	Causative verbs

The subcategories excluded from the computation of the composite scores did not have obvious differences between NS and NNS groups. The numbers of LFs used inaccurately were not included into the computation of the composite scores (Table 12). The computation of the average score per person for NS and NNS groups show that the average number of LFs for NSs is around three times as big as that of NNSs (Table 12, Figure 6). The total number of the composite scores for NSs is almost twice as higher as that of NNSs (Table 12).

Table 12

Frequency table of composite scores for NNS participants

Composite score (categories 1-6)	
NS	NNS
23	0
24	1
30	1
31	1
42	1
44	1
48	1
52	1
55	2
96	4
	7
	7
	8
	9
	9
	9
	10
	12
	16
	17
	17
	27
	37
	48

Total	
445	246
Average	
44.5	10.25

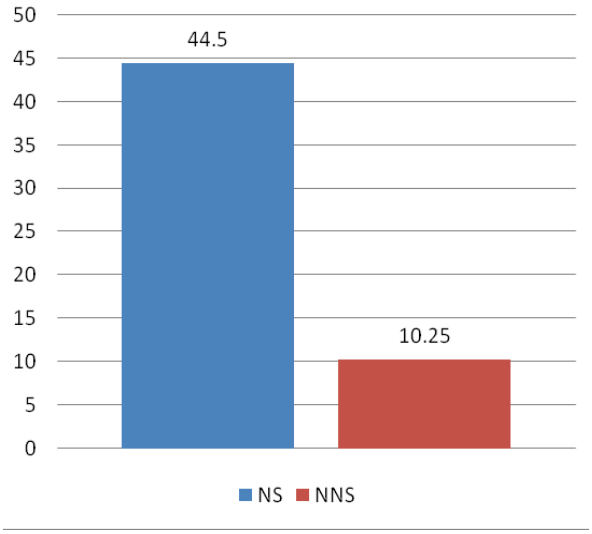

Figure 7. Average composite score per person for NS and NNS

4.2.2 Correlation Analysis

Firstly, the Spearman bivariate 2-tailed correlation analysis was run between the composite scores and the data of the 16 questions of the questionnaire about the NNSs ELEs to identify questions with significant correlations. The latter variables were supposed to significantly contribute to the models of the multiple regressions, too. The questions with significant correlations across all kinds of ELEs (Table 13) were:

- question 4 – interaction with native speakers
- question 9 – listening and summarizing;
- question 14 – oral presentations
- question 15 – debates and discussions;
- question 16 – role plays.

From the results of the correlation analysis it was also identified that the IBT speaking scores significantly correlated with the composite scores of the NNSs.

Table 13

Spearman bivariate 2-tailed correlation

Composite score	Correlation coefficient	Significance level
IBT speaking score	.579**	.003
Total Q2 exposure quantity	.013	.951
Total Q3 CDs for NSs	.363	.082
Total Q4 NS interaction	.509*	.011
Total Q5 international textbooks	.364	.080
Total Q6 non-international textbooks	-.193	.365
Total Q7 read/summarize	.332	.113
Total Q8 read/translate	.138	.521
Total Q9 listen / summarize	.426*	.038
Total Q10 listen / repeat	.336	.108
Total Q11 grammar exercises	.036	.867
Total Q12 essays	.368	.077
Total Q13 vocabulary list	-.120	.575
Total Q14 oral presentation	.465*	.022
Total Q15 debates	.448*	.028
Total Q16 role plays	.439*	.032
Staying abroad	.199	.351

*Notes:*p≤ 0.05 (2-tailed), ** p≤ 0.01 (2-tailed)*

4.2.3 Multiple Regressions Analysis

The multiple regressions were run with the model of significantly correlated questions (4 – interaction with native speakers, 9 – listening and summarizing, 14 – oral presentations, 15 – debates and discussions, 16 – role plays) and the composite scores. Table 14 and Table 15 show the results of the multiple regressions with questions 4, 9, 14, 15 and 16 as IVs and composite score as the DV. Since the number of independent

variables in the model was more than 2, the adjusted R square indexes should be considered to identify the level of predictability of the models.

It can be concluded from the multiple regressions analysis that for the model of these 5 independent variables (4–interaction with native speakers, 9–listening and summarizing, 14–oral presentations, 15–debates and discussions, 16–role plays) chosen based on the correlation, F test did not show significant results. However, significant F results were observed for two models:

Model 1

Dependent or criterion variable – composite score

Independent or predictor variables – questions: 4, 9, 14, 15, 16 (in private tutoring experience)

Model 2

Dependent or criterion variable – composite score

Independent or predictor variables – questions 4-NS interaction (across all types of experiences, excluded PhD experience)

Results for Model 1

The results of the backward multiple regressions for Model 1 show that 40 % of the variance can be explained by NS interaction in private tutoring experience (Table 14).

Table 14

Multiple regressions (backward) for Model 1

Model	*F*	Adjusted R^2	*β*
Step 1	4.43 (5,18) **	.43	
(Constant)			10.7**
NS interaction			5.18**
Listen/summarize			-3.512*
Oral presentation			1.56
Debate			.62
Role play			1.42
Step 2	5.71 (4,19)***	.45	
(Constant)			11.3***
NS interaction			5.11***
Listen/summarize			-3.34*
Oral presentation			1.76
Role play			1.52
Step 3	7.07 (3,20)***	.44	
(Constant)			11.6***
NS interaction			5.45***
Listen/summarize			-3.20*
Oral presentation			2.26
Step 4	8.58 (2,21)***	.40	
(Constant)			12.2***
NS interaction			5.70***
Listen/summarize			-2.07

*Notes: *p≤ 0.05, ** p≤ 0.01, *** p≤ 0.005*

The equation of multiple regressions for Model 1 can help to predict the possible composite scores for NNSs if they are exposed to a certain amount of exposure of NS interaction in private tutoring experience. The equation for Model 1 is the following:

Composite score (Model 1) = 12.2 + 5.7 x NS interaction (in private school)

According to the results of backward multiple regressions for Model 1, if the amount of NS interaction is the highest in private tutoring experience, for example 'always' (6) on a Likert scale format of the questionnaire, the composite score will likely be around the average score of NSs (for comparison refer to Table 12 and Figure 7):

Composite score (Model 1) = 12.2 + 5.7 x 5 = 40.7 (NS interaction in private tutoring)

Results for Model 2

The results of the backward multiple regressions for Model 2 show that 32 % of the variance can be explained by NS interaction in private tutoring experience (Table 15). It can be seen that the results of the multiple regressions for Model 1 and Model 2 match with each other. That is, in both models according to the backward regression analysis NS interaction in private school tutoring has been identified to be the only significant predictor variable.

Table 15

Multiple regressions (backward) for Model 2

Model	*F*	Adjusted R^2	*β*
Step 1	3.76 (6,17) *	.42	
(Constant)			13.9
Secondary			1.02
Undergraduate			-4,66
Private			5.49***
Self			3.90
AUA			-2.73
Non-AUA			7.58
Step 2	4.65 (5,18)**	.44	
(Constant)			13.7
Undergraduate			-5.16*
Private			5.41***
Self			4.69*
AUA			-2.71
Non-AUA			8.13
Step 3	5.08 (4,19)**	.42	
(Constant)			5.20
Undergraduate			-4.93
Private			5.22***
Self			4.09*
Non-AUA			6.20
Step 4	5.51 (3,20) **	.37	
(Constant)			6.10*
Undergraduate			-2.93
Private			5.00*
Self			2.84
Step 5	6.70 (2,21) **	.33	
(Constant)			5.21

Private			4.98***
Self			1.77
Step 6	11.9 (1,22) ***	.32	
(Constant)			7.28***
Private			5.1***

*Notes: *p≤ 0.05, ** p≤ 0.01, *** p≤ 0.005*

The equation of multiple regressions for Model 2 can help to predict the possible composite scores for NNSs if they are exposed to a certain amount of exposure of NS interaction in private tutoring experience. The equation for Model 2 is the following:

Composite score (Model 2) = 7.28 + 5.1 x NS interaction (in private school)

According to the results of backward multiple regressions for Model 2, if the amount of NS interaction is the highest in private tutoring experience for example always (6) on a Likert scale format of the questionnaire, the composite score will be close to the average score of NSs (for comparison refer to Table 12 and Figure 7):

Composite score (Model 2) = 7.28 + 5.1 x 5 = 32.78 (NS interaction in private tutoring)

CHAPTER FIVE: DISCUSSION AND CONCLUSION

Introduction

This section summarizes the findings of the research to answer the first and second research questions. The findings may provide insightful conclusions and pedagogical implications for English teaching and learning methodology in Armenia and in SLA and EFL in general. The chapter also discusses the limitations and delimitations of the study and gives different suggestions for further study.

5.1 Discussion of the findings of the first research question

The results of the data analysis for the first research question have shown that NS and NNS corpora did have differences in the frequencies of the use of the certain LFs selected to be distinguishing (Tables 5-10 and Figures 1-5). The descriptive statistics also provided the evidence for the differences and consequently for the right choice of the distinctive features (Table 4).

The findings of the first research question shed light on the ways of teaching and learning English as a foreign language in Armenia. It appeared that the Armenian EFL learners indeed did not sound native like in terms of the use of certain LFs in their speech.

Generally, the frequency of the use of the specific LFs by NNSs was quite low across almost all the categories. However, there seem to be some LFs that were more favorably used by NNSs (adverb '*mostly'*), or the ones which did not show obvious differences ('*modals with base verbs'* and the adverb *'maybe'*). The fact that the differences between the frequencies of the use of target LFs in NS and NNS corpora were

obvious, could be a headstone for the further analysis to answer the second research question.

5.2 Discussion of the findings of the second research question

First of all, the analysis of the Spearman bivariate 2-tailed correlation between the composite scores of LFs and ELEs (based on 16 questions) have shown significant correlations only for 5 questions (interaction with native speakers, listening and summarizing, giving oral presentations, debates and role plays).

However, the result of the multiple regressions for the model of those 5 questions as independent variables and composite scores as dependent variables, has shown insignificant predictability. According to the multiple regressions analysis those questions separately may serve as predictors for native-like competence but not together in a model.

The results of the multiple regressions have identified that the most significant predictability was seen in 2 models: 1) dependent variable – composite score and independent variables – questions: 4, 9, 14, 15 and 16 (in private tutoring experience); and 2) dependent variable – composite score and independent variables – questions 4 (across all types of experiences, excluded PhD experience). Since the results of the backward multiple regression analysis for both models show that the only significant independent variable in both models is NS interaction in private tutoring experience, it can be concluded that if Armenian EFL learners learn English being exposed to frequent native speaker interaction during their private tutoring experience, they seem to develop near native like competence. However, the conclusions may be tentative because of the limitations of the study.

5.3 Limitations and Delimitations

Like many studies, the current study also has limitations and delimitations. The limitations might have skewing effect on the findings of the study. Thus, all the findings have to be interpreted cautiously and conclusions have to be made tentatively.

5.3.1 Limitations

- The small sample size (the number of NNS participants was 24 and that of NS participants 10) was one of the major limitations along with the number of too many independent variables.
- The interviews were conducted by an Armenian nonnative speaker of English which would reflect on the production of Task 6 of the interview (simulation task).
- Even though the target LFs identified in the native and nonnative speech could be the same, the nonnative speaker participants might use them in different contexts. And the study did not consider the differences due to context.
- Composite scores assumed that all target LFs had equal weighing while it might not be the case in real NS corpus.

5.3.2 Delimitations

- One of the delimitations of the study is that the nonnative participants were selected from a specific setting, namely the American University of Armenia. The participants all were first or second year graduate students at the university where the medium of instruction was English.
- The second delimitation is that the participants of the target group of native speakers were American NSs of English.

- Finally, the number of target LFs is limited to particular categories, that is, fillers, modals, adverbs, collocations, phrasal verbs and structure items.

5.4 Pedagogical Implications

The findings of the first research question showed that NNSs did use the target LFs less frequently. This means that some changes towards English teaching and learning experiences in Armenia should be taken into consideration in order for the learners to get the opportunities to develop near native like competence in spoken English.

Unfortunately, the findings did not appear to have obvious support for English teaching and learning experiences in secondary, undergraduate, graduate and postgraduate schools and self study. However, the types of exposure such as the interaction with native speakers, listening and summarizing material, giving oral presentations, doing debates and role-plays separately could be potentially enriched in all kinds of experiences. Moreover, as a general implication, it can be assumed that the amount of the exposure of native speaker interaction would rather be enriched in private tutoring experiences for nonnative speakers of English to develop near native like competence in the target language.

5.5 Suggestions for further research

- Since obvious differences in the use of certain target LFs have been found out between the NS and NNS corpora, it can be suggested to investigate other linguistic aspects of spoken English of two groups as well.
- It could be suggested to conduct a similar investigation on a larger sample of participants to have more rigor in statistical analysis.

- It would also be interesting to get a certain number of NSs to listen to the recordings of the NNS participants and rate them on the scale of similarity of their speech to that of the NSs. Then, the ratings could first be compared with the results of the quantitative analysis, too.

REFERENCES

Adverbs. (n.d.). In Oxford Dictionaries online. Retrieved August 27, 2012, from http://oxforddictionaries.com/definition/english/adverb

Aijmer, K. (2004). Pragmatic Markers in Spoken Interlanguage. *Nordic Journal of English Studies*, 173-190.

Retrieved from: http://130.241.16.45/ojs/index.php/njes/article/viewFile/241/238

Conklin, K. & Schmitt, N. (2008). Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Oxford University Press, Applied Linguistics 29* (1), 72–89.

Retrieved from: http://sla.sjtu.edu.cn/thesis/formulaic%20sequences--are%20they%20processed%20more%20quickly%20than%20nonformulaic%20language%20by%20native%20and%20nonnative%20speakers.pdf

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English language and Literature (BELL), 2,* 225-246.

Retrieved from: http://scholar.google.co.uk/scholar?hl=en&q=De+Cock+2004+preferred+sequences&btnG=Search&as_sdt=0%2C5&as_ylo=&as_vis=0

Ellis, N., Simpson-Vlach, R. & Maynard, C. (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly, 42* (3), 375-396.

Retrieved from: http://web.mac.com/ncellis/Nick_Ellis/Publications_files/TQ.pdf

Fillers. (n.d.). In Encyclopedia Britannica online. Retrieved August 27, 2012, from http://www.merriam-webster.com/dictionary/filler

Fraenkel, J. & Wallen, N. (2009). *How to design and evaluate research in education. Seventh Edition.* Location of publication: McGraw-Hill Higher Education. Retrieved from: http://amazon.com

Grammar (n.d.). In Oxford Dictionaries online. Retrieved August 27, 2012, from http://oxforddictionaries.com/definition/english/grammar

Jabeen, F. Rai, M. & Arif, S. (2011). A corpus based study of discourse markers in British and Pakistani speech. *International Journal of Language Studies (IJLS), 5*(4), 69-86. Retrieved from: http://www.ijls.net/volumes/volume5issue4/asim1.pdf

Jiang, N. & Nekrasova, T. (2007). The Processing of Formulaic Sequences by Second Language Speakers. *The Modern Language Journal, 91*(3), 433–445. Retrieved from: http://555seminar.pbworks.com/f/The%2BProcessing%2Bof%2BFormulaic%2BBY%2BSLS.pdf

Koya, T. (2003). A Study of Collocation in English and Japanese Noun-Verb Combinations. *Intercultural Communication Studies, 12* (1), 125-147. Retrieved from: http://www.uri.edu/iaics/content/2003v12n1/08%20Taeko%20Koya.pdf

McCarthy, M. & Carter, R. (2006 a). Ten criteria for a spoken grammar. *Explorations in Corpus Linguistics* (pp.27-52). New York: Cambridge University Press. Retrieved from: http://www.cup.es/other_files/downloads/esl/booklets/McCarthy-Corpus-Linguistics.pdf

McCarthy, M. & Carter, R. (2006 b). This that and the other: Multi-word clusters in spoken English as visible patterns of interaction. *Explorations in Corpus Linguistics* (pp.7-26). New York: Cambridge University Press.
Retrieved from:
http://www.cup.es/other_files/downloads/esl/booklets/McCarthy-Corpus-Linguistics.pdf

Modal verbs. (n.d.). In Oxford Dictionaries online. Retrieved August 27, 2012, from http://oxforddictionaries.com/definition/english/modal%2Bverb

Mukharjee, J. (2009). The Grammar of Conversation in Advanced spoken learner English. *Studies in Corpus Linguistics, 33*, 203-230.
Retrieved from:
http://books.google.co.uk/books?hl=en&lr=&id=3yfOJSMTRhoC&oi=fnd&pg=PA203&dq=Mukherjee+2009+LINDSEI&ots=wrdOVjS3vy&sig=5jBVYzqueRSuTdpOTxhbFgt0oLk#v=onepage&q&f=false

Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics, 24* (2), 223-242.
Retrieved from:
http://www.corpus4u.org/forum/upload/forum/2005062000435073.pdf

No, K. & Park, K. (2010). Self-introduction in Spoken Form: What Makes KSE and ANSE different? *Pan-Pacific Association of Applied Linguistics, 14* (2), 85-104.
Retrieved from: http://eric.ed.gov/PDFS/EJ920522.pdf

Phoocharoensil, S. (2011). Collocational Errors in EFLLearners' Interlanguage. *Journal of Education and Practice , 2* (3), 103-120.

Retrieved from: http://scholar.google.co.uk/scholar?hl=en&q=Phoocharoensil+2011&btnG=Search&as_sdt=0%2C5&as_ylo=&as_vis=0

Phrasal verbs. (n.d.). In Encyclopedia Britannica online. Retrieved August 27, 2012, from http://www.merriam-webster.com/dictionary/phrasal%20verb

Pichler, H. & Levey, S. (2010). Variability in the Co-occurrence of Discourse Features. *University of Reading Language Studies Working Papers, 2,* 17-27. Retrieved from: http://wwwcgi.rdg.ac.uk/web/FILES/english-language-and-literature/ell_language_Pichler_vol_2.pdf

Tagliamonte, S. A. & Denis, D. (2010). The Stuff of Change: General Extenders in Toronto, Canada. *Journal of English Linguistics, 38* (4), 335–368. Retrieved from: http://eng.sagepub.com/content/38/4/335.full.pdf

APPENDICES

Appendix 1

SPEAKING TASKS for the Piloting Session (8 Tasks)

Task 1 Personal Experience

Where did you learn English (any foreign language for NSs)?

What did you use to do most often in your English classes (any foreign language for NSs)?

What did you like and what you didn't in your English classes (any foreign language for NSs).

Task 2 Picture Description

Describe the picture. You have some 30 seconds to prepare. Try to talk for 1-2 minutes.

http://studentweb.usq.edu.au/home/D1222041/html/practice_2.htm

Task 3 Picture Story

Look at the pictures below and create a story. You have some 30 seconds to prepare. Try to talk for 1-2 minutes.

Task 4 Telling a well known story

Please tell the story of Romeo and Juliette by Shakespeare? You have some 30 seconds to prepare. Try to talk for 2-3 minutes.

Task 5 Argumentative Speech

The story of Romeo and Juliette raises an issue that could be contradictory these days. should interracial / intercultural marriages be acceptable in society (all over the world). What is your opinion about this? Please give reasons to support your answer. You have some 30 seconds to prepare. Try to talk for 1-2 minutes.

Task 6 Video summary

Watch a video episode.

After you finish watching, summarize the video episode and give your opinion about the usability of this kind of device for everyday needs.

Try to talk for 1-2 minutes.

Source: http://www.youtube.com/watch?v=smGmrpn2Vrk&feature=player_embedded

Task 7 Preference

Look at the pictures and talk about your most preferable way of personal communication. Please support your response by giving advantages and disadvantages to your choice. Try to talk for 1-2 minutes.

Task 8 Role play

Imagine you are a travel agent. Simulate a situation responding to the request of the customer (the interviewer). You can use the following prompts:

1. Different class (business, economy, first) prices
2. Available flight for that day/period
3. The cheapest flight
4. Make Reservation
5. Name
6. Payment (cash/bank transfer)

Appendix 2

Instrument 1. SPEAKING TASKS for INTERVIEW

(for NSs only:

Please provide with the information about your stay in the US.
Where were you born?
How long have you lived / have you been living in the US?
How long have you been living in Armenia?)

Task 1 Personal Experience

Where did you learn English (any foreign language for NSs)?

What did you use to do most often in your English classes (any foreign language for NSs)?

What did you like and what you didn't in your English classes (any foreign language for NSs).

Task 2 Picture Description

Describe the picture. You have some 30 seconds to prepare. Try to talk for 1-2 minutes.

http://studentweb.usq.edu.au/home/D1222041/html/practice_2.htm

Task 3 Picture Story

Look at the pictures below and create a story. You have some 30 seconds to prepare. Try to talk for 1-2 minutes.

Task 4 Video summary

Watch a video episode.

After you finish watching, summarize the video episode and give your opinion about the usability of this kind of device for everyday needs.

Try to talk for 1-2 minutes.

Source: http://www.youtube.com/watch?v=smGmrpn2Vrk&feature=player_embedded

Task 5 Preference

Look at the pictures and talk about your most preferable way of personal communication. Please support your response by giving advantages and disadvantages to your choice. Try to talk for 1-2 minutes.

Task 6 Simulation

Imagine you are a travel agent. Simulate a situation responding to the request of the customer (the interviewer).

Appendix 3

Instrument 2. Questionnaire for Armenian nonnative speakers of English

Please fill in the boxes that most appropriately describe your English learning experience.

1. How long did you live in an English speaking country (or a country where you spoke only in English)? Please write in a number for months and/or years	month	years

QUESTIONS	Secondary school	Undergraduate school	Graduate School (nonAUA)	Private classes	Self-study effort	Graduate School (AUA)	PHD
2. How long have you been learning English (number of months and/or years)?	m. y.	m. y.	m. y.	m. y.	m. y.	m. y.	m. y.
3. Give an accurate estimate of the intensity of your English classes per week in hours (including classes in English at AUA)?							

QUESTIONS (please tick in the boxes)	SECONDARY SCHOOL							UNDERGRADUATE SCHOOL						
	Always	Very often	Often	Some-times	Very rarely	Never		Always	Very often	Often	Some-times	Very rarely	Never	
4. How often did you use audio/video materials produced for or by native speakers of English (movies / videos, audio CDs, songs) in your English learning experience?														
5. Overall how often have you interacted with native speakers of English on a regular basis?														
6. How often have you used textbooks from international publishers in your English learning experience (e.g. Cambridge, McMillan, Oxford, etc.)?														
7. How often have you used textbooks published in non-international / local publishers in your English language learning experience (e.g. Russian or Armenian publications, N. A. Bonk, etc.)?														
8. How often did you use these kinds of activities in your English learning experience?														
• Reading and summarizing orally in your own words														
• Reading and translating texts														
• Watching and listening to recordings in English and summarizing them orally in your own words														
• Watching and listening to recordings in English and repeating them word for word														

QUESTIONS (please tick in the boxes)	Y	Always	Very often	Often	Some-times	Very rarely	Never	DUAT E	Always	Very often	Often	Some-times	Very rarely	Never
• Writing grammar exercises (e.g. sentence completion, multiple-choice, gap-filling, etc.)														
• Writing compositions / essays														
• Learning vocabulary lists by heart														
• Giving oral presentations														
• Discussions / Debates														
• Role plays														

GRADUATE NON-AUA	Always	Very often	Often	Some-times	Very rarely	Never	PRIVATE CLASSES	Always	Very often	Often	Some-times	Very rarely	Never	SELF-STUDY	Always	Very often	Often	Some-times	Very rarely	Never	GRADUATE SCHOOL AUA	Always	Very often	Often	Some-times	Very rarely	Never	PHD	Always	Very often	Often	Some-times	Very rarely	Never