The Dissertation Committee for Navid Hayeri Certifies that this is the approved version of the following dissertation:

Does gender affect translation?
Analysis of English talks translated to Arabic

Committee:

James W. Pennebaker, Co-Supervisor

Michael C. Hillmann, Co-Supervisor

Elaine K. Horwitz

Rebecca M. Callahan

Peter F. Abboud

Does gender affect translation?

Analysis of English talks translated to Arabic

by

Navid Hayeri, B.S. Math; M.S. Math; M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2014

Dedication

To my beloved wife Maryam

Acknowledgements

Many exceptional people helped me to complete this work. First, I want to thank Dr. James W. Pennebaker who gave me the opportunity to work in his lab and to start such an unforgettable journey. I never knew that counting insignificant words led to very significant findings. He provided me all the resources for such an expensive project.

I am in debt to Dr. Michael C. Hillmann who introduced me to the world of corpus linguistics and lexicography. He taught me lessons of research and life. He was always supportive and provided many useful comments on this work.

Thanks to my dissertation committee members: Dr. James W. Pennebaker, Dr. Michael C. Hillmann, Dr. Peter F. Abboud, Dr. Elaine K. Horwitz, and Dr. Rebecca M. Callahan.

I will never forget working with Dr. Peter F. Abboud on an advanced Arabic textbook project. Through our weekly meetings, he helped me gain a better understanding of grammatical issues in Arabic.

I have a deep respect for Cindy K. Chung, my mentor, colleague and friend who played a major role in forming this dissertation. Also, I am grateful to Dr. Aman Attieh from Swarthmore College, Dr. Ernest McCarus from the University of Michigan, and Dr Raji Rammuny from the University of Michigan who accepted me in their private club working on the Arabic textbook project. Thanks to Maryam Shariati my colleague and a true friend in need.

I wish to acknowledge my parents, Nasrin and Mohammadtaghi, and my brothers, Vahid and Omid, who I never fell short in their love for me. A million thanks to the best wife in the world, Dr. Maryam Mousavi, who tolerated the situation of kids plus a graduate student at home.

Starting the program, I have been blessed with a new born son, Sepehr, and finishing it with another joyful blessing a baby girl, Delsa. Thanks to Delsa for crying all the night before my oral defense.

Does gender affect translation?

Analysis of English talks translated to Arabic

Navid Hayeri, PhD

The University of Texas at Austin, 2014

Supervisors: James W. Pennebaker

Michael C. Hillmann

Abstract:

When a text in a foreign language is translated into English, many of the features of the original language disappear. The tools described in this paper can give people who work with translators and translations an insight into dimensions of a culture that may escape the notice of someone not familiar with the source language or culture. A set of computer programs are described that analyze both English and Arabic texts using each language's function word or closed-class words categories. First, the LIWC (Pennebaker, Booth, & Francis, 2007) text analysis program was translated into Arabic. Then, the grammatical dimensions of Arabic function words was determined that served as a basis for the Arabic LIWC designed for Arabic texts. These same Arabic dimensions were used to fit English words into the same categories. A large corpus of Modern Standard Arabic

and English text files that have been translated in both directions were used to establish the equivalence of the translated word lists.

Then, the uses and applications of the dictionaries for computer-based text analysis within and across cultures are described in the study of influence of gender on translation of TED talks between English and Arabic. Differences were identified in language style between men and women in their English language TED talks, and these features were examined whether they were faithfully maintained in translations to Arabic. The rates of function word use was employed to measure language style. Function words (e.g., pronouns, prepositions, conjunctions) appear at high rates in both English and in Arabic, and they have been shown to provide social, demographic, and psychological information about authors and speakers in English and a variety of other languages. The sample included 328 (196 male and 132 female) TED talks delivered in English from 2004 to 2010 and their translations to Arabic. Rates of function word use in the original and translated texts were examined using the English version of the word counting software. The function word use compared between male and female speakers, male and female translators, and their interaction. The results confirmed gender differences in language style for English texts found in previous studies in English. For example, women used more pronouns, more negatives, and fewer numbers than did men. It was

further found that several of the distinguishing language style features between men and women in English disappeared in Arabic translations. Importantly, there was a significant gender difference in the language style of male and female translators: first person singular pronouns, second person pronouns, conjunctions, and prepositions were used more by female translators, and quantity words were used more by male translators, regardless of the gender of the original speaker.

This study presents one application of computerized text analysis to examine differences in language style that may be lost or gained in translations. Future research and applications within personality, forensic, and literary psychology, linguistics, and foreign language studies are discussed.

Keywords: Arabic, English, LIWC, Translation, Gender

Table of Contents

List of Tables

List of Illustrations

List of Abbreviations

AA-LIWC: Arabic-Arabic Linguistic Inquiry and Word Count

AE-LIWC: Arabic-English Linguistic Inquiry and Word Count

ANOVA: Analysis of Variance

EA-LIWC: English-Arabic Linguistic Inquiry and Word Count

EE-LIWC: English-English Linguistic Inquiry and Word Count

LIWC: Linguistic Inquiry and Word Count (text analysis software)

MANOVA: Multivariate Analysis of Variance

MSA: Modern Standard Arabic

TED: Technology, Entertainment, Design Conference, an open translation project

Chapter One: Introduction

When a text in a foreign language is translated into English, many features of the original language are stripped away. For example, Americans reading Arabic stories may never learn about differences in the social closeness of the characters because the words used to convey this information which are inherent in Arabic do not have clear English equivalents; for example there are eight scenarios for the word "cousin" in Arabic where Arabs distinguish between the children of uncles and aunts, boys and girls, on the father's side and the mother's side. Most English translations, then, force the original stories to be viewed through English eyes. Similarly American stories, interviews, or speeches that are translated into another language are restructured to fit the categorization schemes of the other language and culture.

Translators and publishers mostly choose material appropriate for their own culture and market. Lawrence Venuti examined several translation studies in his book called *The Scandals of Translation* to show that translations in many ways are marginalized, and people are not well-informed about this biased translation selection. He raised the issue of authorship and copyright of a translation. Also, he found that there are fewer translations from foreign language literature to English compared to material translated from English, and most of these literary translations were selective or as he puts it, marginalized. For example, Japanese literature translated into English covers only a few authors who do not necessarily represent Japanese literature. Venuti mentioned a book translated into English in Britain as *Smilla's Sense of Snow* and in the United States

as *Miss Smilla's Feeling for Snow* based on target language culture (Venuti, 1998). Venuti shows that the culture of the target language may influence the direction of translations.

On May 23, 1969, just before the Six Day Arab-Israeli war, Gamal Abdul Nassir the president of Egypt said defiantly " ...نحن مستعدون للحرب..." which literally means: "…we are prepared for war …", but it was translated "we prepared for war" (See Oren, 2002, McNamara 2003, Quotes by Gamal Abdel Nasser). Of course the war did not happen because Nasir's statement was misinterpreted, but a huge difference exists between "being prepared and ready for a war" as opposed to "intentional preparation to initiate a war". The wording of both sentences is almost identical except for the use of auxiliary verb "are". Although, the word "prepared" as an active verb in "we prepared" imply taking time to lay out a plan, and, "prepared" as an adjective in "we are prepared" imply having all the resources ready, there is important difference between the meanings of two phrases. This example shows how misusing of a simple particle, pronoun or auxiliary can affect an accurate translation.

With the advent of computerized text analysis tools, it is known that the function words within a language are at the heart of that culture's way of categorizing the world. Function words (i.e., in English pronouns, prepositions, articles) are used at high rates in all languages, are difficult to master, and are categorized differently across different languages and cultures (Selkirk, 1996). Indeed, the analysis of function words provides important clues about the social relationship between a speaker and the audience, the

psychological state of the speaker, and information about the speaker's social role and status in the community (Chung & Pennebaker, 2007, 2008; Pennebaker, Mehl, & Neiderhoffer, 2003). By developing ways of studying function words using a text analysis approach across languages, the world can be seen through the different lenses of different cultures.

In this study, the development of computerized text analysis tools is described that categorize function words of the Arabic language according to an English grammatical scheme. The computerized text analysis tools can give people who work with Arabic translators and translations an insight into dimensions of a culture that may be invisible to someone not familiar with the language or culture. With any written documents, including transcripts of interviews, conversations, or other language samples, analyses of the social and psychological dynamics of the interaction or the speaker can be provided almost instantaneously.

In this dissertation, I first describe the rationale for and the development of the English and Arabic word lists of function words. Next, I assess their equivalence on a translation corpus comprised of texts from varied sources and genres. Finally, I apply this Arabic computer-based text analysis in the study of talks translated from English to Arabic. I examine whether the gender of the translator may affect the mood of translation to Arabic.

SUMMARY

Translations to and from a foreign language may not fit in the categorization schemes of the source language and culture. Also, the culture of the target language may influence the direction of translations. Sometimes misuse of a simple particle, pronoun or auxiliary can affect an accurate translation.

The main focus of this study will be on function words of English and Arabic through an examination of translations between English and Arabic. Specifically, the gender considers as a factor which may affect translations accuracy.

Chapter Two: Literature Review

LANGUAGE, CULTURE, AND TRANSLATIONS

Although there are countless ways to segment language, two dimensions are of interest here: content words and function words. Content words (also called open class words) are typically nouns, verbs, adjectives, and most adverbs that label objects, events, and actions. Function words (also known as closed class words, particles, or even junk words) reflect linguistic style and are made up of pronouns, prepositions, articles, auxiliary verbs, negative particles, and conjunctions. While the translation of content words is generally straightforward, function words pose a different set of problems. In our native language, we automatically process the meanings of function words. When learning a new language, however, function words tend to be the most difficult to master (Bloom, 1981, 1984). One difficulty stems from the complex ways that function words are used. As described below, function words often convey subtle social and psychological processes that can reflect almost invisible features of the speaker, the audience, the message, and the culture itself.

One example of the different roles of function words across languages comes from a phenomenon called "pronoun-drop", where pronouns can be excluded from an utterance (Kashima & Kashima, 2003). Pronouns exist in pronoun-drop languages but, for example, in Japanese, they are longer in length and arise later in children's language development than in American English speakers (see Rumsey, 2003). Japanese speakers have a wider range of self-referential pronouns to choose from and their uses are designated according to gender, status, and intimacy of speaker and addressee, if used at

all (Ono & Thompson, 2003). Accordingly, pronoun-drop has the effect of de-emphasizing the significance of the actor, forcing speakers and addressees to pay more weight to the content of the utterance. However, when a text is translated into English, the appropriate pronoun is automatically inserted into the English text in order for the English text to be understood coherently.

Why are the subtleties of translations important? The first reason is that translations and difficulties in translation can reveal actual differences in modes of thought. Recent studies in cognitive psychology have demonstrated the effects of language on color discrimination (Winawer, Wittholt, Frank, Wu, Wade, & Boroditsky, 2007), on how people think about time (Boroditsky, 2001), and on memory and similarity assessments (Boroditsky, Schmidt, & Phillips, 2003). Second, translations provide measurable cultural differences. The translation of texts from one language to another can convey the underlying pronouns and roles that they presupposes. The very act of translation reveals intended and implied meanings (Boroditsky, Schmidt, & Phillips, 2003).

Translating the content of a message is the primary goal in translations. However, translators also implicitly aim to convey other qualities of a message, such as emotional tone and formality. While some these qualities may be under conscious control through the selection of appropriate vocabulary, i.e., through the content of what is said, other qualities operate more implicitly, such as through the style of how something is said. The content of a message is typically conveyed through the use of nouns, verbs, some

adjectives, and some adverbs. On the other hand, language style is reflected in the use of function words such as articles, prepositions, and pronouns which are difficult to manipulate consciously or to monitor without sacrificing the coherence or understanding of the content of a message. Accordingly, function words have received little attention in research on translations. Since function words have been shown to be associated with a variety of demographic and psychological states (Chung & Pennebaker, 2007; Pennebaker, 2011), the current study attempted to discover the features of language style that are lost or gained in translations by measuring rates of function word use.

GENDER DIFFERENCE IN THE ENGLISH LANGUAGE

It is widely known that there are gender differences in English language use (for a review, see Mulac, Bradac, & Gibbons, 2001). For example, women use more hedges (e.g. "I guess", "Could you maybe…"; e.g., Hartman, 1976), and question phrases (e.g. "Shall we eat?" instead of "Let's eat." (Mulac, Weimann, Widenmann, and Gibson, 1988).

The term "language differences" and "differences in language style" are not well-defined and scholars often have different methodologies to examine language differences. This has led to diverse and sometimes contradictory findings in regard to studying gender difference in language use. Contradictory findings in gender studies usually occur due to small sample size, diverse methodology and false generalization (Newman et. al. 2008).

Newman and colleagues (2008) analyzed over 14,000 varied text files by men and women, including speeches, recorded conversation, published literature, and e-mails. In

terms of content words, women tended to use more social words, references to home, and fewer swear words. In terms of function words, women tended to use more pronouns, more negatives, fewer articles, fewer numbers, and fewer prepositions overall. These findings parallel previous findings on language style using a corpus linguistic approach that show that women's language style is more "involved" and men's language style is more "informational" (Biber, 1995). Overall, women's language style is more personal, and women tend to focus more on social dynamics than men, who are more concerned with objects.

GENDER DIFFERENCE IN THE ARABIC LANGUAGE

Gender differences have also been found in Arabic language use. Much of such works has focused on the degree to which spoken Arabic more closely conforms to the "high variety" of Arabic used in the Koran and in Modern Standard Arabic (see Walters, 1996). Note that the degree to which formal or literary Arabic is used in everyday speech is positively correlated with the degree of education one has attained (Ibrahim, 1986; Walters, 1996). One might expect, then, that women would tend to use a less formal language style, given their historically lower status and education relative to men in Arabic cultures. However, as educational opportunities and academic participation have increased for women, the opposite has been found. For example, Abu-Haidar (1988, 1989) reported that Baghdad women tend to use a more formal or more prestigious language style made up of Muslim dialect and literary Arabic than Baghdad men, as indicated by the use of passive verb forms as opposed to third person plural verb forms,

the use of calendar month names (i.e., January, March) instead of ordinal numbers (i.e., 1st month, 3rd month), adverbs, adjectives, and fewer loanwords from other languages, and fewer dialect words. One interpretation of these findings is that because of their lower status in Arabic cultures, women are more prestige conscious and show that they cannot afford to slip into a less formal register (Ibrahim, 1986).

Also, Trudgill (1972) reported that men use less formal speaking styles towards other men, with the goal of establishing solidarity rather than fighting for status. These findings confirmed that men tend to style shift more to accommodate their audience when speaking in Arabic, as found in a previous study on the language styles of men and women in Iraq (Abu-Haidar, 1988). In sum, women are generally more prestige conscious than men and thus adopt a more formal linguistic style across contexts.

OPPOSITE GENDER TRANSLATORS

Studying men's translations of women's writing and women's translations of men's writing reveals that men and women not only express meanings differently, but also that they understand and thus interpret meanings differently in target languages. This phenomenon appears in much research about the translation of sacred texts by women and the translation of literary texts by opposite gender translators. However, there is a lack of literature about whether the gender is an obstacle to understand and ultimately deliver meaning to another language and how gender may affect translation.

It is interesting that the English language has two obsolete words "translatress" and "translatrix" for a female translator. Translatress appeared in the first edition of *the*

Oxford English Dictionary in 1884 and in 1913 version of *Webster's New International Dictionary of the English Language*. Virginia Woolf considered Aphra Behn the first novelist, dramatist and translatress (Agorni 1998, Chalmers 2004). Scholars often distinguish between female voices in translation (see Krontiris 1997, Brown 2005). Variety of approaches considered to study different female translators (Palusci 2011).

Historically, it was not acceptable to hear a woman's voice in the translation of sacred texts before the 16th century. Women translators like Mary Sidney Hilbert and Christiane Nord did translate the Bible under male authority; however they left a trace of not only of women's language but also female perspectives towards the sacred text (Long, 2011). Long (2011) argued that several female translators of the Bible and the Koran changed the traditional masculine interpretation of the Koran and the Bible when female translation are coherent/direct in language and context with the feminist perspective. For example, Laleh Bakhtiar published a translation of Koran as *The Sublime Quran*. In her translation of Koran into English, Bakhtiar used the phrase "ones who are ungrateful" for the Arabic word /*kufr*/ which traditionally translated as "disbelievers" or "infidels". Also, her translation of "God" instead of "*Allah*", "Jesus" and "Mary" instead of Islamic renderings "*Isa*" and "*Maryam*" aroused much controversy among Muslim scholars (Ahmed-Ullah, 2007; Macfarquhar, 2007).

Peretz (1992) studied translations of the female Polish poet, Anna Świrszczyńska, into English by a male translator, the 1980 Nobel Prize winner Czesław Miłosz, published in a collection called *Happy as a Dog's Tail*. Peretz found that even though the

male translator had native Polish proficiency, he misunderstood and eventually misinterpreted the female author in several pieces. Miłosz rewrote the poems in English with the male dominant perspective and Świrszczyńska's register was lost in translation. Peretz (1992) argued that because the translator was younger and had a more poetic attitude, he changed the more mature female voice. This observation reveals that men and women face difficulty understanding and ultimately translating opposite gender work and more study is needed to discern the impact of gender on translation paradigms.

SCOPE OF THE PRESENT STUDY

This study focus is on the effect of gender in translation from English into Arabic. In other words, it investigates whether the language use of male and female translators are similar across translation. To do that, the rate of function word use is compared in both the original language and the corresponding translation for several translators. To count the rate of function words, a tool is needed to capture and categorize the occurrence of function words in both English and Arabic. Therefore, I introduce the development of a new text analysis tool which can process Arabic language in Chapter Three. Then in Chapter Four, I apply this text analysis tool to study 328 males and females in various talks and their translations into Arabic (1) to examine language differences between males and females, and (2) to measure the extent to which the gender of translators affect language style in translations.

SUMMARY

It is well known that there are gender differences in English language use. For example, women's language style in English is more personal, and it tends to focus more on social dynamics than is the style of men, who are more concerned with objects. Gender differences have also been found in Arabic language use. For example, women use more formal language.

Studies reveal that, translators often have difficulty understanding and ultimately translating opposite gender language. Also, in translating sacred texts, men and women had their own perspective in understanding the meaning of the text.

Function words often convey subtle social and psychological processes that can reflect almost invisible features of languages. Function words have received little attention in research on translations. In present study, I will examine function word used in several translations between English and Arabic language to analyze gender differences in English, Arabic, and across translation.

Chapter Three: Development of Arabic LIWC

Delivering the meaning to another language is the primary goal of any translation; however, maintaining all aspects of original texts often is an impossible task by translators. Studying the impact of the gender of translators on translations from English to Arabic is the main goal of current dissertation. According to chapter two, function words are the heart of a language and embody cultural aspects of a language. Recall Jamal Nasir's statement, misusing a simple auxiliary led to misinterpreting a whole phrase. Therefore, a method of monitoring function words of Arabic and English will be adopted to study possible gender difference in men and women translating into Arabic. For the purpose of studying translations between English and Arabic, a computer text analysis tool has been developed to classify and count Arabic function words according to English and Arabic grammar.

In this chapter, I introduce the development of an Arabic version of the computerized text analysis program called *Linguistic Inquiry and Word Count*, hereafter LIWC (Pennebaker, Booth, & Francis, 2007). Arabic LIWC can be used to demonstrate what is lost, maintained, or gained in translations between English and Arabic. Arabic LIWC provides parallel information about function word use in the original and translated version of English and Arabic texts. This information enables researchers to compare and contrast the degree in which function word use was maintained, dropped or increased in several categories in Arabic-English translation.

First, I introduce the original LIWC program. Then, several Arabic function word categories were introduced and included in LIWC to build the desired Arabic LIWC. Next, a comprehensive parallel corpus of Arabic and English texts was collected in order to examine the validity and reliability of the Arabic LIWC. Finally, several statistical tests were conducted to show relationship between English and Arabic function words. Later in chapter four, the Arabic LIWC dictionaries will be used to address a series of research questions regarding gender differences in English and Arabic and the degree to which the gender of translators affect translations.

LINGUISTIC INQUIRY AND WORD COUNT (LIWC)

Linguistic Inquiry and Word Count is a word counting software widely used in the social sciences. LIWC reports the percentage of words devoted to a word category within a given text file, for over 80 categories. The LIWC dictionaries contains over 4000 words and word stems categorized as linguistic (e.g. articles, conjunctions, pronouns), psychological (e.g. emotions, cognitive mechanisms, social), and/or content categories (e.g. work, home, death). Much care was taken to ensure that the word and word stem entries were valid; at least two of three independent judges had to agree whether or not each word or word stem belonged to a particular category (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007).

Across almost two decades of studies using LIWC, a consistent finding has been that function words are more reliable markers of personality and psychological state than are regular content words (Pennebaker, Mehl, & Niederhoffer, 2003). Function words

and/or particles (e.g. pronouns, auxiliaries, articles and conjunctions) appear more frequently in a language and account for more than half of a language's. Function words act as cement for a language which connect content words together and embody social and psychological meanings. In other words, the language markers of personality and psychological state can be assessed through language style (via function words), even when people are talking about different subjects (as determined by content words). Function words make up approximately 50% of the words that people use, and their uses in natural language are difficult to manipulate, meaning that psychologists can sample naturalistic, non-reactive, and frequently occurring behaviors through examination of function words (Chung & Pennebaker, 2007). With the ability to examine behavior in ecologically valid contexts, the number of studies in the social sciences that have used LIWC has grown considerably since 2000.

The use of LIWC in psychological studies has extended beyond the United States, where it was originally developed. This has been made possible because the software includes a feature for the user to select an external dictionary to reference when analyzing text files. This feature, along with the ability of LIWC2007 to process Unicode text, has enabled the processing of texts in many other languages. Currently, there are validated dictionaries available in Spanish (Ramirez-Esparza, Pennebaker, Garcia, & Suria, 2007), German (Wolf, Horn, Mehl, Haug, Pennebaker, & Kordy, 2008), Dutch, Norwegian, Italian, Chinese, Russian, and Korean. Versions in Hungarian and Turkish are in various phases of development. Note that each of these dictionaries has been developed using the

LIWC2001 (Pennebaker, Francis, & Booth, 2001) or LIWC2007 default dictionary categorization scheme. Words in other languages have been fit into the English language categorization scheme used in the LIWC2001 or LIWC2007 default dictionaries.

In developing LIWC dictionaries for other languages, each team faced obstacles in categorizing some common words in their language when key features of their language did not exist in the English language categorization scheme. The greater the difference from the English language, the more trouble the teams had conforming to the English language categorization scheme. For example, in Spanish, the formal pronoun *usted* (you) and the informal pronoun *tu* (you) ended up in the broader category as Second Person Pronoun. Therefore, dimensions such as these were ignored because it was desired that the computer program to convert Spanish into language categories that fit the English categories. In essence, teams were forcing LIWC to see read foreign language through American eyes. The teams resigned to the fact that some key features of their language would get lost in translation; they all (more or less) conformed to the English language categorization scheme for ease of comparison across languages.

ARABIC LANGUAGE

Arabic is a language very different from English (in characters, reading direction, morphology, and syntax), and therefore extremely difficult to conform to an English language categorization scheme. For example, Arabic nouns have a root form upon which various inflections can be added in order to indicate what grammatical function the noun, adjective or verb serves within a sentence. Other inflections on a root word can signal

number, gender, tense, and definiteness (definite or indefinite) case (for a review, see Elbeheri, Everatt, Reid, & Mannai, 2006; Schulz, Krahl, & Reuschel, 2000).

Arabic also has function words that do not have one-to-one equivalents in English. For example, bilingual learners of Arabic often have difficulty in translating the Arabic connective "فـ" /fa/ into English, since it has no exact English equivalent (Saeed & Fareh, 2006). "Fa" literally means "then, therefore, thus" and used frequently as prefix at the beginning of a sentence, however, "fa" often functions as a punctuation delimiter for emphasis. Similarly, Arabic, but not English, is a language that contains an explicit counterfactual marker, suggesting that the Arabic culture has an inclination to entertain counterfactual premises more than do English speakers (Bloom, 1984; Lardiere, 1992). For example, Lardiere (1992) showed that Arab speakers use significantly more "لو" /lau/ (if unreal, hypothetical condition marker) than they use "إن" /in/ (whenever, a high degree of possibility condition) or "إذا" /idha/ (if) when challenged by interviewer's question.

Since function words tend to be stronger and more reliable correlates of psychological states, the author chose to focus primarily on the inclusion of function words in each of the dictionaries. Several other categories with words that occur at high base rates and therefore likely to be culturally relevant were also included such as *Allâh* (God). However, content words (e.g. nouns and regular verbs) were generally excluded as these would likely appear at much lower rates across texts.

It should be emphasized that the author is quite aware that there are pronunciation differences in dialects in spoken Arabic across the Arab-speaking world. The actual words (especially content words), accents, and even some grammatical features vary from region to region. Despite these differences in spoken Arabic, most formal communication, both written and spoken, is in Standard Arabic. For the purposes of the proposed project, only Modern Standard Arabic (MSA) was used as well as Standard English.

General Procedure Used in the Development of Arabic Dictionaries

I started with the original English LIWC function word classification to build the Arabic function word classification. Hereafter, the term EE-LIWC will be used to address text analysis dictionaries that use the current English-based way of categorizing function words in LIWC2007. Recall, that the original LIWC is more comprehensive and is not limited to function words. Then I will derive the Arabic-reading programs which read and categorize Arabic text called EA-LIWC. Again, EA-LIWC contains Arabic function words classified based on English grammar.

Translation of LIWC in English (EE-LIWC) to Arabic (EA-LIWC)

Recall that the original English LIWC is the default dictionary in LIWC2007. All function word categories in the EE-LIWC were to appear in the EA-LIWC. So the first step was to translate the entries of each function word category in the EE-LIWC into Arabic to compile the EA-LIWC dictionaries. Each translated entry was maintained in the function word categories of the EE-LIWC. For example, the pronoun "I" in English

was translated into "I" in Arabic, "أنا" and categorized in its corresponding categories: Pronouns, Personal Pronouns, and First Person Singular Pronouns.

The second step was to include any forms of the word fitting that category (e.g. "me" in Arabic, "إياي"), any combination of characters representing that word (e.g. "as for me" in Arabic, "ولكنى"), or any other relevant word (e.g. "myself" in Arabic, "نفسي") in those categories. These stages of translation were completed with the help of three major dictionaries such as *A Dictionary of Modern Written Arabic* (Wehr & Cowan, 1976), *Al-Munjad* (Ma'luf, 2000), and *Al-Mawrid: A Modern Arabic-English Dictionary* (Ba'albaki, 2005).

The third step was to inflect commonly occurring verbs, decline nouns and prepositions in all their Arabic forms. Note that Arabic has many more forms for each verb since they can be inflected with 14 different pronominal suffixes as a single unit, by tense (past, present, future), and by mode (positive, negative). In this situation, what meant by inflection is conjugation of all the verbs, and generating all possible forms of words in the list with attached pronouns, and attached particles (prefixes). As mentioned previously, Arabic nouns can also be declined into six forms based on gender (masculine, feminine), number (singular, dual, plural), and as either definite or indefinite. Prepositions also are inflected with attached pronouns, resulting in up to 14 possible forms for each preposition. Again, here, inflection and declining of words means generating all possible forms of a word with all possible attached particles and pronouns.

The fourth step was to have two additional Arabic-speaking judges (one male and one female) agree that the new entries indeed had been categorized in the correct function word category. The judges reviewed each word to assess the spelling, the conjugations, and the appropriateness of the categorization. Although disagreements were rare, any disagreements were extensively discussed until a consensus between the two judges had been reached.

The final version of the EA-LIWC included a total of 19 categories and subcategories and 3614 entries (see Table 3.1). It is important to note that the EA-LIWC is designed so that the categories have the same names as in the EE-LIWC. For example, the category "Pronoun" in the EE-LIWC is also called "Pronoun" in the EA-LIWC in order to facilitate comparison of the two programs in statistical analysis, and especially for research in English. Also, note that this classification of Arabic words is based on English grammar which appeared in the classification of the original LIWC (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007).

LIWC Dictionary Categories	English Examples	Arabic Examples
Function Words		هذا، جميعكم، أنتن
Pronouns		أنكم، إنها، كلكم
Personal Pronouns	I, me, myself	أنا، إني، إياي
I	lets, us, we	أنفسنا، نحن، وإنا
We	you, yours, yourself	إياك، عليكما، أنتن
You	hers, him, she	إليها، هو، فجميعها
She/He	theirs, themselves, they	أنهم، فيهن، منهما
They	it, someone, those	أي، فلان، كله
Impersonal Pronouns	never, no, not	لن، ضد، ليس
Negative Particles	and, but, which	التي، أن، حسبما
Conjunctions	about, onto, so	إلى، عن، مع
Prepositions	ever, now, only	أحياناً، أيضا، حوالي
Adverbs	any, each, lots	الأفراد، الفرق، مقدار
Quantity	billion, dozen, tenth	السابعة، مئة، مليار
Number	are, being, have	أصبح، ما زال، لست
Auxiliary Verbs	and, both, with	فمثلا، كذلك، حتى
Inclusive	but, except, without	عدا، فقط، بل
Exclusive	ok, yeah, yes	فعلا، نعم، بلى
Assent	where, how, why	أين، كيف، لماذا
Question		

Table 3.1. English and Arabic Dictionary Categories according to English Grammar

LINGUISTIC FEATURES AND LIMITATIONS OF ARABIC DICTIONARIES

In the development of EA-LIWC, despite English, there are several features of Arabic that hinder progress in identifying a discrete word. The most important obstacle was random use of diacritics, attached pronouns, prefixes, suffixes and articles. In this section, I explain how these issues are addressed in EA-LIWC.

Diacritics. Note that the Arabic alphabet does not include letters for short vowel sounds. Short vowels are represented by diacritics (*harakât*) placed above or below a consonant letter. Diacritics tend to be arbitrarily used in Arabic texts; some texts don't have them, and in texts that have them, they are not used consistently. Only the most formal texts include all diacritics to leave little room for ambiguity. Because there are no standards for the use of diacritics, even very popular search engines ignore diacritics from the keywords that they search. It is up to the reader to infer from the context what the writer intended and therefore interpret the words with none, some, or all diacritics inserted. As such, I kept only one diacritic ("*hamza*") in our Arabic-LIWC dictionaries, since it tends to change the meaning of a word more than do other diacritics. I also planned to pre-process Arabic texts for LIWC by removing the diacritics that do not appear in our dictionaries (i.e. "*fatha*", "*kasra*", "*damma*", "*fathatân*", "*kasratân*", "*dammatân*", "*shadda*" and "*sukūn*").

By removing diacritics, some of the headwords have the same spelling even when their function, pronunciation and meaning are different. For example, after pre-processing, "إِنْ" (if) and "إِنَّ" (truly, indeed) are spelled "إن". Also "أَنْ" (that, to) and

"أنَّ" (that) are spelled "أن", and both examples could have alternate spellings as "ان". Although the EA-LIWC dictionaries cannot distinguish among them, they are categorized in all relevant categories (repeated times). Second, most of them belong to a superordinate category, for example here "ان" belongs to the particle category, and it works either as a coordinating conjunction, a linking preposition, or a conditional marker.

To capture all occurrences of function words, all possible forms of a word were included in our dictionaries; for some words, more than one correct form exists. By processing a large corpus of Modern Standard Arabic, I realized that, in written texts, many common mistakes occur especially when *hamza* or dots are omitted. For example, consider the word "ألذي" (that, which). Three different simplified forms are found as "الذي", "ألذى" and "الذى", and so I included all four possible forms.

Word Inflections. Recall that in each Arabic category, I included all possible forms of nouns such as definite-indefinite, masculine-feminine, singular-dual-plural. In addition, verbs were inflected with different subject pronominal suffuixes (all 14 pronoun forms), with tense (past, present), and with mode (affirmative, negative).

Pronouns. In Arabic, there are two types of pronouns: *attached* and *detached*. Detached pronouns are called subject pronouns or active particles (in Arabic grammar *Mubtadâ*, *Musnad* or *Damīr Fâ'il*), and are arbitrarily used in spoken and written Arabic. Since verbs are inflected with pronominal suffixes in Arabic, speakers might emphasize the subject or contrast using a detached pronoun. On the other hand, attached pronouns appear as suffixes at the end of the words (noun, verb, particle) to make them either

possessive pronouns, object pronouns, or object of preposition respectively. In order to capture the uses of attached pronouns, all prepositions in the list were inflected with attached pronouns in the EA-LIWC.

Particles. Although each particle carries its own independent meaning in Arabic, some particles act as a prefix to other content words (i.e. they attach to the beginning of a word without any spaces). These prefixes can consist of only one letter (e.g. لـ (for)), two letters (e.g. ال (the)) or even more complex, a combination of existing prefixes (e.g. و + بـ + ال → وبال (and by the)). There are seven one-letter connective particles in Arabic that appear as prefixes (i.e. they are connected to the following word without any space to denote a new clause):

i) "و" (waw), the 27th letter in Arabic alphabet. In most cases, it means "and, also, too, along with, as well as, plus". When و comes after a verb, it becomes the preposition of a verb meaning "with". و can also mark the beginning of sentences when it plays a prefix role for a verb. Also, و might indicate that two actions occurred simultaneously, when it means "while, as, when". In rare cases, و denotes a swear word.

ii) "فـ" (fa), the 20th letter, is typically used as a coordinating conjunction, meaning "then, then, thereupon, so, thus, therefore, consequently".

iii) "سـ" (sin), the 12th letter, is a prefix used by present tense verbs to indicate the near future. سـ is translated as "will, shall".

iv) "لـ" (lam), the 23rd letter, appears before nouns, verbs, and some particles. The general meaning is "for, to, because of, on account of, truly, really, definitely, surely, undoubtedly".

v) "بـ" (ba), the 2nd letter, is a preposition with different meanings such as "in, at, on, with, by, by means of, through, for".

vi) "كـ" (kaf), the 22nd letter, means "just as, like, similar to, the same as, such as".

vii) "أ" or the interrogative *hamza*, appears at the beginning of sentences to make yes/no questions (e.g. do/does/did/would/am/is/are ….?).

Linguists interpret the use of connecting particles as a punctuation system for Arabic to determine the boundary of sentences or clauses (Akram and Sa'adeddin, 1987). Although the use of a European punctuation system has recently increased, the traditional use of these connecting particles and linking prepositions is still common. Although I did not count the occurrence of these one-letter particles, most of the entries in both the EA-LIWC dictionaries were inflected with some of these particles. Number 1 and 2 are used at extremely high frequencies, and so the size of EA-LIWC dictionaries was tripled by adding them as prefixes to each entry in the EA-LIWC dictionaries. Number 3 was added to the beginning of all verbs in the Arabic dictionaries. Number 4 and 5 were inflected with all attached pronouns. Number 6 and 7 were neglected because of their low frequency rate.

Articles. Besides the seven one-letter particles discussed above, the definite article "ال" (the, pronounced *al*) is the most frequent prefix in Arabic, and consists of two-letters. ال can be combined with other one-letter particles to create six more common forms of the article the in Arabic: وال (and the), كال (like the), فال (then the), بال (by the), لال (for the), and وبال (and by the). Note that in Arabic, in definite adjective clauses, both nouns and adjectives carry the definite article ال, and so the frequency of ال typically appears at least twice the frequency of "the" in English (e.g. "the green house" in English is translated into Arabic as البيت الاخضر "lit. the house the green" in Arabic"). This is because the adjective-noun word combination agrees in all aspects of gender (male, female), number (singular, dual, and plural), and definite or indefinite (as illustrated).

Word count programs (e.g. Wordsmith, Scott, 2008; LIWC; Pennebaker, Booth, & Francis, 2007) typically neglect the attached prefixes, pronouns, and articles mentioned above because the programs only identify words by either a space delimiter or by punctuation. Other Arabic language text analysis software typically neglects these attached prefixes because they are based on word stems. Word stemming refers to the process of removing all of a word's extra characters to produce a stem or root word (Khoja, 2001; Larkey et al., 2002). In the case of Arabic "light stemmers", a set of prefixes and suffixes are simply stemmed off and only the remaining infixes (i.e. patterns or roots) are considered for further processing (Al-Ameed et al., 2005). The three most well-known Arabic word stemmers are: 1) Al-Stem, developed by Larkey at University

of Massachusetts (Darwish, 2002), 2) TREC-2001 stemmer, a modified version of Larkey's stemmer with two additional prefixes (Xu et al., 2001), and 3) TREC-2002, the improved version of TREC-2001(Larkey et al., 2002). Stemmers are used for example, in many Arabic machine translators (e.g. BKYMON; BKYC1; BKYC2; BKYC3; Xu et al., 2002) in order to detect all content words in a sentence for translation.

Removal of prefixes in English is harmful because they can change or reverse the meaning or grammatical function of a word. However, in Arabic, the removal of a prefix does not usually change the meaning of the word (Al-Sughaiyer & Al-Kharashi, 2004). In developing the Arabic dictionaries, it was clear that I could not simply ignore these attached connecting particles because of their widespread use in the Arabic language. Indeed, in the development of the Arabic LIWC dictionaries, I was primarily interested in high frequency words and function words in particular. So, I sought to count the frequencies of these attached particles, pronouns, and articles following a procedure used in Arabic word stemming software.

EQUIVALENCE OF FUNCTION WORDS ACROSS ENGLISH AND ARABIC

In the following section, a series of statistical tests was performed to show the equivalence of categories across languages for the translated dictionaries. First, a comprehensive corpus of texts sampled from diverse genres and sources was introduced. Then, a series of text analyses was conducted to describe the base rate of word use for each of the categories across genres. In addition, statistical tests performed to examine

the equivalence of the categories across translated texts. Discussion about the suitability of the dictionaries for text analyses follows.

Validation corpus. The corpus for testing and validating the LIWC dictionaries included texts that had been faithfully translated between English and Modern Standard Arabic (see Table 3.2). A wide variety of genres was sampled: newspaper articles, television news broadcasts, speeches by public figures, letters, poems, musical lyrics, religious and philosophical texts, and short stories. The corpus had been compiled from various sources, including a professional corpus consortium, books, and online sites for literature and popular music.

The validation corpus included a total of 143 texts in English, and 143 texts in Arabic. As can be seen in Table 3.3, the English texts contained an average of 12,383 words (SD = 28,329), and the Arabic texts contained an average of 4,674 words (SD = 13,052). Note that our Arabic sample contained an average of 51.65% of the total words relative to the English translations, as is typical in translated texts between the two languages.

Corpora	Source	Original Language	Sample
News Media			
Agence France Presse	LDC2007T17	Arabic	5
An Nahar	LDC2007T17	Arabic	5
Xinhua News Service	LDC2007T17	Arabic	5
Human Rights Watch	www.hrw.org	English	5
Published Literature			
Tales of the Arabian Nights		Arabic	6 fiction stories
The Works of Abu Hamid al-Ghazali	www.ghazali.org	Arabic	3 philosophical books
The Koran	http://quranexplorer.com/	Arabic	5 Surats
The Bible	www.netbible.com	English	6 books
TV News Broadcasts			
An Nahar	LDC2004T24	Arabic	5
Dubai TV	LDC2004T24	Arabic	5
Lebanese Broadcast	LDC2004T24	Arabic	5
Nile TV	LDC2004T24	Arabic	5
Voice of America	LDC2004T24	Arabic	5
Speeches			
Hosni Mubarek		Arabic	6
Susanne Mubarek		Arabic	4
Barack Obama		English	3
Kofi Annan, UN Secretary General	www.uno.org	English	1
Sh. Al-Sabah al-Ahmad al-Jaber of Kuwait		Arabic	1
Music Lyrics			
Abdel Halim Hafez		Arabic	19
Hussain Al-Jasmi		Arabic	5
Julia Boutros		Arabic	1
Latifa		Arabic	1
Lena Chamamyan		Arabic	3
Rashid al-Majid		Arabic	1
Ziyad Rahbani		Arabic	1
Poems			
Hassab Al-Baqali		Arabic	1
Gibran Khalil Gibran		Arabic	8
Nazar al Qabbani		Arabic	1
Mahmoud Darwish		Arabic	1
Letters			
Maryam Jameelah	Correspondence between Abi-l-A'la al-Maudoodi and Maryam Jameelah	Arabic	14
Abul Ala Maudoodi	Correspondence between Abi-l-A'la al-Maudoodi and Maryam Jameelah	Arabic	11
Herbert Marcus	Correspondence between Abi-l-A'la al-Maudoodi and Maryam Jameelah	English	1
Total			142 Texts and their translations

Table 3.2. Corpora Genres and Sources

C Corpora	Sample	English Sample			Arabic Sample		
		Total Words	Mean Words Per Text	SD Words Per Text	Total Words	Mean Words Per Text	SD Words Per Text
News Media	20 articles and their translations	29,842	1,492.10	1,168.03	9,330	466.50	290.36
Literature	20 literary sections and their translations	1,250,520	65,526.00	50,758.31	516,355	25,817.75	26,446.26
TV News Broadcasts	20 news broadcasts and their translations	316,140	15,807.00	11,634.17	89,407	4,470.35	3,271.42
Speeches	15 speeches and their translations	46,811	3,120.73	5,837.64	13,454	896.93	1,523.94
Lyrics	31 lyrics and their translations	20,651	666.16	468.95	7,445	240.16	133.54
Poems	11 poems and their translations	9,810	891.82	878.02	4,207	382.45	306.82
Letters	26 letters and their translations	97,043	3,732.42	1,869.72	28,275	1,087.50	509.08
Total	143 Texts and their translations	1,770,817	12,383.34	28,329.31	668,473	4,674.64	13,052.06

Table 3.3. Corpora Word Count

Analytic Strategy. The Arabic texts in the validation corpus were processed using the new Arabic LIWC dictionaries (EA-LIWC). The English texts in the validation corpus were processed using the EE-LIWC dictionaries. The relevant categories (i.e. those that had been translated for EA-LIWC) in the EE-LIWC captured an average of 56.25% (SD = 13.19) of the words in the texts, the EA-LIWC 29.59% (SD = 6.46). In processing the texts, it must be noted that in translating English to Arabic, roughly 30-50% of the volume of texts is reduced. Likewise, in translating Arabic to English, roughly 30-50% of the volume of texts is increased.

To test for the equivalence of dictionary categories across languages, several sets of comparisons between category means (EE-LIWC categories vs. EA-LIWC categories) were conducted using paired t-tests and correlations.

Results. Table 3.4 (comparing EE-LIWC and EA-LIWC categories) shows the mean percentage of words per text that fell into each category for each dictionary. The first column in each of the tables shows the correlation for each category between a given dictionary and its translation. As can be seen, most category correlations were high, indicating that those categories were equivalent. More specifically, higher correlations indicated that the category entries in one dictionary appeared in their translated from in both the text and in the translated dictionary at higher rates. As can be seen in the correlation column for each table, approximately 14 of the 19 categories were strongly correlated ($p < .001$) in the EE-LIWC vs. EA-LIWC, while two categories (Impersonal Pronouns and Auxiliary Verbs) were moderately correlated ($p < .01$), and three

categories such as Inclusive, Assent and Question were unrelated. Non-correlated category was predictable for the Inclusive category because the second most frequent word in English "and" belongs to the Inclusive category. As discussed in previous section, the Arabic "and" /*waw*/ (و) is always attached to the next word as a prefix. The EA-LIWC cannot count every occurrence of *waw* in a given Arabic text especially when *waw* is attached to a content word. However, EA-LIWC can count certain occurrence of *waw* when it is attached to any other function word in the EA-LIWC dictionaries. Despite this limitation, the Inclusive category still can be useful in research when comparing function word use of different groups of texts. It seems that because words in the Assent and Question categories were used in a very small rate in the validation corpus, not significant non-zero correlation was found between English and Arabic in these categories. Note that all statistical analyses in this study were done by R software, a language and environment for statistical computing (R, 2009).

		EE-LIWC		EA-LIWC			
LIWC Dictionary Categories	R	Mean	SD	Mean	SD	t	p
Function Words							
Pronouns	.37**	14.14	7.49	12.05	4.12	-3.52	.001
Personal Pronouns	.51**	9.62	6.94	7.49	3.91	-12.11	<.001
I	.39**	3.05	4.22	3.52	2.93	1.36	.176
We	.49**	1.20	1.94	.26	.42	-6.33	<.001
You	.45**	2.57	3.19	1.04	1.77	-6.37	<.001
SheHe	.61**	1.70	1.96	2.17	1.49	3.58	<.001
They	.65**	1.10	1.35	.41	.59	-7.67	<.001
Impersonal Pronouns	.25*	4.52	1.92	6.35	2.53	7.92	<.001
Negative Particles	.38**	1.24	1.22	1.67	2.42	2.23	.03
Conjunctions	.40**	6.75	2.31	7.71	3.60	2.23	.03
Prepositions	.44**	13.94	2.83	12.59	3.88	-3.34	.001
Adverbs	.33**	2.53	1.45	2.36	1.43	-4.38	<.001
Quantity	.41**	1.97	1.11	1.50	1.34	-1.28	.20
Number	.66**	.94	.88	1.05	1.28	4.12	<.001
Auxiliary Verbs	.19*	6.75	2.80	3.41	1.95	-1.38	.17
Inclusive	.08	5.89	2.39	.46	.53	-12.61	<.001
Exclusive	.56**	1.85	1.55	3.27	2.96	26.99	<.001
Assent	-.09	.22	.63	.10	.22	-6.93	<.001
Question	.05	.37	.90	3.42	1.63	2.13	.04

Note. R = Pearson-product moment correlations, *$p < .01$, **p < .001. Means and standard deviations represent percentages of words per text. t-tests were paired, *df* (142).

Table 3.4. Equivalence of English and Arabic Dictionary Categories according to English Grammar

APPLICATION OF EA-LIWC

Recall that function words of Arabic included in EA-LIWC captured more than 51% of the comprehensive corpus. Also a strong correlation existed between categories of function words of English and their corresponding translation in EA-LIWC. It is worth mentioning that in some categories like first person singular pronoun, no pattern of similarity was found because of the fact that the pronoun "I" is mostly embedded in Arabic verbs and EA-LIWC is unable to count it accurately. Overall, EA-LIWC could be used to count function words of Modern Standard Arabic texts with respect to certain constraints such as not detecting articles and some attached first person pronouns.

The psychometrics and analyses reported here show how cross-language investigations might be conducted using the set of Arabic LIWC dictionaries, or how other language dictionaries might be created and validated for cross-language investigations. Validation of the use of the various EA-LIWC across languages has many applications in cross-cultural psychology, computational linguistics, Arabic language education and forensic psychology.

In the case of cross-cultural psychology, for example, demographic or psychological characteristics could be assessed (with further validation beyond the gender differences presented in chapter four) in translations or in documents for which the original language is unknown. With further validation work for markers of language style for other psychological features such as depression, deception, or adaptive coping across translations, researchers who are familiar only with the English language could conduct analyses of foreign language texts using translations into English. News articles

from a given Middle Eastern region could be assessed for demographic or psychological characteristics, regardless of knowing the original language of the article.

In the case of forensic psychology, if a translated text is presented to a researcher, it might be difficult to assess the demographic or psychological characteristics of the author if the original text or language skills of the research are unavailable or inaccessible. As another example, consider the case where some documents are available for a given subject in Arabic, while other documents are available only as English language translations. By using the set of Arabic LIWC dictionaries presented here, it would be possible to assess certain features of language style in both texts and treat them as equivalent or be cautious in doing so in order to maximize the use of all available documents without translations. In forensic investigations, this may be the case: captured or overheard communications may be available in a given language, but more public communications might be more readily available in English. Having the set of dictionaries to combine the language samples could maximize the degree to which the results are reliable and representative of the communications from a particular individual or group.

Clearly, more validation work is required to assess the use of the Arabic LIWC dictionaries for cross-language investigations. However, the results presented here show that indices of language style found to be markers of gender differences in English were found in Arabic language texts and their translations. Formal language style was also maintained across translations as measured by the Arabic LIWC. These are promising

findings for future research on language style markers for other demographic and psychological characteristics.

SUMMARY

Linguistic Inquiry and Word Count (LIWC) is a word counting software which reports the percentage of words devoted to a word category within a given text file. LIWC originally developed to analyze English texts and later translated to several languages. The steps taken to translate LIWC to Arabic were discussed in this chapter.

First, all function word categories of original LIWC have translated to Arabic and validated by three independent judges. Then, all Arabic attached pronouns and attached particles were conjugated with all function words in the Arabic LIWC dictionaries. Finally, all verbs and nouns were conjugated and included in Arabic LIWC dictionaries.

In order to validate the equivalence of categories of Arabic LIWC across languages, a parallel corpus was compiled consisting of texts sampled from diverse genres and sources in Arabic and English. Then, a series of text analyses was conducted to describe the base rate of word use for each of the categories. The results show that EA-LIWC capture more than 51% of Arabic corpus, and there is a strong correlation for most of (14 out of 19) categories of function word of English and Arabic.

To show one application of the Arabic LIWC, in next chapter, the impact of gender on translation of several talks from English to Arabic will be studied focusing on function words of both languages.

Chapter Four: Study of Gender Difference across Translation

How men and women could misunderstand opposite gender writings is under study. Many previous studies focused on a case study of literary texts translation or used a small sample qualitative approach to determine how men and women translate differently. As discussed earlier, studying function words in a text may reveal aspects of language invisible to readers. In the present chapter, a quantitative method of counting function words will be adapted to examine how male and female Arabic translators interpret English talks.

More importantly, one application of the developed Arabic text analysis tool, the Arabic LIWC, will be examined to study gendered language across translation between English and Arabic. First, 328 talks from TED conference were collected from both male and female speakers which were translated by male and female translators from English into Arabic. The rate of function words in all talks and corresponding translations were calculated. A series of statistical tests explored similarity and differences in the language style of male and female speakers and translators. Also, excerpts from translations were provided to highlight differences between male and female translators.

RESEARCH QUESTIONS

Translation is a process that is potentially affected by many factors such as the nature of the languages themselves (e.g. pedagogical, grammatical, and vocabulary), translator ability (e.g. skills in the source and target languages), and translator personality (attitudes, belief system, mood, gender, age, social background). The question is "does

the gender of the translator affect the mode of translation?". One application of Arabic LIWC was used to examine how gender could influence translations. More specifically, do gender differences in English appear in Arabic translations? To answer these questions, the following steps have taken.

SAMPLE

The corpus consisted of speeches from Technology, Entertainment, Design conference available at TED.com, a publicly accessible site that archives videotaped and transcribed professional talks delivered by experts on a variety of topics to a live audience. The TED Open Translation Project allows translators to post translations of TED talks in order to make the talks available to foreign language audiences around the world. Since the inception of the TED project, more than 4,365 volunteers have translated over 13,000 talks into 80 different languages.

For purposes of this study, 328 talks (196 male and 132 female) were downloaded from TED.com which were translated into Arabic from 2004 to 2010. The contributions of male and female translators varied (241 translations by males and 87 translations by females). Although complete demographic information about translators did not exist, it was assumed that they were volunteers who were mostly native speakers in the target language (Arabic). Table 4.1 shows the number of talks included in our sample corpus, along with the mean and standard deviations of word count for each speaker/translator gender combination. Note that there were some male and female translators who translated several talks from both male and female speakers.

The themes of TED talks are limited to technology, entertainment, design, business, science, culture, arts, and global issues. Speakers at TED conferences are typically at the apex of their fields, being well-known or established researchers in their fields. The length of talks varied from 5 to 20 minutes and 2,100 words per talk in average in the selected sample.

Having no assumption about the length of speeches and corresponding translations, in average, the word count of male and female speech is virtually identical (*average male word count =2137 vs average female word count =2124*, $t = 0.11$, $p = 0.92$). In Table 4.2, all scenarios are considered for word count comparison between male-female speakers and translators in both English and Arabic. Anywhere similarities or differences in length of speech existed in English, the same happened in corresponding Arabic. Also, it is interesting that the male talks translated by female translators were the wordiest pieces on average in our sample. This finding supports the Maluc and Lundell's finding (1994) when they studied several written discourses of adult women and men and found out that women discourses were longer than men.

	N	English WC Mean	WC SD	Arabic WC Mean	WC SD
Male Speakers	196	2137	1052	1791	879
Female Speakers	132	2124	1153	1744	933
Male Translations	241	2017	1112	1684	919
Female Translations	87	2447	974	2014	802
Males translated by males	143	1975	1044	1671	890
Males translated by females	53	2573	951	2115	768
Females translated by males	98	2079	1206	1704	964
Females translated by females	34	2251	992	1858	842
Total	328	2131	1092	1772	900

Table 4.1. Descriptive Corpus of TED Talks Translated to Arabic

Word Count Comparison						
Comparison Group 1	Comparison Group 2	Mean Group 1	Mean Group 2	t	df	p
Male Speakers English	Female Speakers English	2137	2124	0.101	326	0.916
Male Speakers Arabic	Female Speakers Arabic	1791	1744	0.46	326	0.646
Male Translations English	Female Translations English	2017	2447	-3.396	326	0.001*
Male Translations Arabic	Female Translations Arabic	1685	2014	-3.157	326	0.002*
Males Translated by Males English.	Males Translated by Females English	1975	2573	-3.81	194	0.000*
Males Translated by Males Arabic	Males Translated by Females Arabic	1671	2115	-3.436	194	0.001*
Females Translated by Males English	Females Translated by Females English	2079	2251	-0.821	130	0.414
Females Translated by Males Arabic	Females Translated by Females Arabic	1704	1858	-0.884	130	0.380
Males Translated by Males English	Females Translated by Males English	1975	2079	-0.697	239	0.487
Males Translated by Males Arabic	Females Translated by Males Arabic	1671	1704	-0.269	239	0.788
Males Translated by Females English	Females Translated by Females English	2573	2251	1.503	85	0.137
Males Translated by Females Arabic	Females Translated by Females Arabic	2115	1858	1.435	85	0.156
Males Translated by Males English	Females Translated by Females English	1975	2251	-1.445	175	0.155
Males Translated by Males Arabic	Females Translated by Females Arabic	1671	1858	-1.151	175	0.255
Males Translated by Females English	Females Translated by Males English	2573	2079	2.767	146	0.006*
Males Translated by Females Arabic	Females Translated by Males Arabic	2115	1704	2.86	146	0.005*

Table 4.2. Length of Speech Comparison between Male-Female Speakers and Translators

The transcription of each talk (out of 328) and the corresponding translation were saved in separate text files, and processed using EE-LIWC, EA-LIWC. The LIWC output was examined for the percentages of the various function word categories. In other words, the LIWC output contained the percentage of function words used in each category relative to number of words in each text. For example, consider a text with 100 words, where five words from the Conjunction category existed in the text. In the LIWC results, the Conjunction category would receive a 5% score. For every talk in English, EE-LIWC was employed to record the rate of function word use in 18 categories. Similarly, for every translation piece, EA-LIWC was employed to report function word use in Arabic. Later, a series of statistical analyses were conducted on the LIWC results to address whether the rates of function word use in speeches by males and females were accurately maintained in translations by males and females.

MAIN EFFECTS OF GENDER ON LANGUAGE IN ENGLISH

A 2 by 2 MANOVA (Multivariate Analysis of Variance) was conducted to identify the effects speaker gender and translator gender on function word use in English. The Speakers' Gender (SG) and Translators' Gender (TG) were the two independent variables; and the 18 EE-LIWC function word categories were dependent variables.

From Table 4.3, it can be seen that the interaction term between SG and TG was not significant (*Wilks' Lambda=0.970, F=0.577, df=(16, 309), p>0.05*). This enables us to interpret any existing significant single effect of any independent variable. The overall effect of SG was significant (*Wilks' Lambda=0.861, F=3.10, df = (16, 309), p <0.0001*).

It means for the 18 function word categories as a group, there is a significant multivariate effect in relation to the gender of the speakers. In other words, differences might exist between males and females in the use of at least one function word category in English. There was no significant effect of TG on function word use in English (*Wilks' Lambda=0.957, F=0.859, df = (16, 309), p> 0.05*). This is important because it shows that the gender of the translator did not drive selection of talks with different function word patterns. In other words, the collection of talks selected by male translators and the collection of talks selected by female translators were similar in terms of function word use.

	df	Wilks	approx F	num Df	den Df	Pr(>F)
SG	1	0.861	3.099	16	309	0.000***
TG	1	0.957	0.859	16	309	0.616
SG:TG	1	0.970	0.577	16	309	0.899
Residuals	324					

Note: *** *p < 0.001, ** p< 0.01, * p< 0.05*

Table 4.3. MANOVA of Function Word Use in English Talks as a function of SG and TG

To identify the sources of gender differences in English, a series of pair-wise t-tests was conducted on all the function words categories of EE-LIWC (Table 4.4). Consistent with Newman et al. (2008), it was found that women used more pronouns (especially personal pronouns, specifically first person singular and third person singular pronouns), more negatives, and fewer numbers. No effects were found for articles as they were not included in current analysis, nor did the analysis find an effect for prepositions.

Although the main effect of TG was not significant, a series of t-tests was conducted to look at the direction of mean differences between talks selected by male and female translators (Table 4.5). As expected, there were no significant differences between the collection of talks selected by male and female translators in the use of the function word categories. Later, this information is useful in interpreting translator's effect when there are differences in Arabic translations.

	Mean female	SD female	Mean male	SD male	t	df	p.value
Pronoun	16.86	3.24	15.93	2.87	2.66	326	0.008**
Personal pronoun	9.55	3.16	8.34	2.46	3.72	326	0.000**
I	3.60	2.86	2.49	2.06	3.87	326	0.000**
We	1.83	1.15	2.03	1.15	-1.56	326	0.119
You	1.97	1.55	2.06	1.20	-0.52	326	0.606
She/He	1.00	1.09	0.66	0.88	3.01	326	0.003**
They	1.14	0.79	1.10	0.73	0.43	326	0.665
Impersonal pronoun	7.31	1.86	7.60	1.58	-1.47	326	0.143
Negatives	1.42	1.03	1.21	0.57	2.03	326	0.044*
Conjunction	7.65	1.41	7.20	1.26	3.01	326	0.003**
Preposition	12.90	2.22	13.03	1.55	-0.61	326	0.545
Adverb	5.62	1.41	5.86	1.27	-1.62	326	0.107
Quantity	2.82	0.85	2.82	0.80	-0.02	326	0.981
Number	1.67	0.79	1.99	0.82	-3.57	326	0.000**
Auxiliary Verb	9.42	1.81	9.37	1.43	0.28	326	0.779
Inclusive Particle	6.31	1.45	6.04	1.39	1.69	326	0.092
Exclusive Particle	2.43	0.82	2.40	0.72	0.39	326	0.697
Assent	0.31	0.64	0.21	0.25	1.71	326	0.089

Table 4.4. Rate of function words used by males and females in English

	Mean female	SD female	Mean male	SD male	t	df	p.value
Pronoun	16.47	3.12	16.25	3.03	0.58	326	0.560
Personal pronoun	9.17	2.91	8.70	2.78	1.31	326	0.193
I	3.28	2.30	2.81	2.52	1.60	326	0.111
We	1.98	1.08	1.94	1.18	0.29	326	0.772
You	1.95	1.23	2.05	1.39	-0.59	326	0.557
Shehe	0.84	0.98	0.78	0.99	0.45	326	0.651
They	1.11	0.59	1.12	0.80	-0.06	326	0.950
Impersonal pronoun	7.30	1.57	7.55	1.74	-1.21	326	0.228
Negatives	1.37	0.57	1.27	0.86	1.29	326	0.199
Conjunction	7.19	1.20	7.45	1.38	-1.68	326	0.095
Prep	12.91	1.60	13.00	1.93	-0.41	326	0.683
Adverb	5.56	1.34	5.84	1.32	-1.64	326	0.103
Quantity	2.76	0.80	2.84	0.83	-0.86	326	0.392
Number	1.92	0.74	1.84	0.85	0.81	326	0.421
Auxiliary verb	9.40	1.32	9.39	1.69	0.04		0.967
Inclusive	6.13	1.08	6.16	1.52	-0.18	326	0.854
Exclusive	2.34	0.60	2.44	0.81	-1.18	326	0.238
Assent	0.19	0.23	0.27	0.51	-1.88	326	0.060

Table 4.5. Rate of English function words used in talks selected by male and female translators

MAIN EFFECT OF TRANSLATORS' GENDER ON ARABIC TRANSLATIONS

A 2 by 2 MANOVA was performed to identify the effects of SG and TG on function word use in Arabic. SG and TG were the two independent variables; the 18 EA-LIWC function word categories were the dependent variables in this analysis.

The MANOVA (Table 4.6) yielded a significant multivariate effect on TG (*Wilks Lambda=0.769, F=5.788, df= (16, 309), p < 0.0001*). It means for the 18 function word categories as a group, there is a significant multivariate effect in relation to gender of translators. In other words, differences might exist between male and female translators in the use of at least one function word category in Arabic. However, after translating talks to Arabic, the interaction term was not significant (*Wilks Lambda=0.933, F=1.367, df=(16, 309), p >.05*). This is important because a non-significant interaction term allows us to interpret the single effect of the independent variable. Also, the main effect of SG was not significant (*Wilks Lambda=0.928, F=1.488, df=(16, 309), p > .05*), meaning there was no difference, after translation, in function word use in Arabic between the talks that had been originally delivered by males and those delivered by females. This means that the gender of the original English speaker is not detectable in language style after translations to Arabic.

Recall that the MANOVA in English (Table 4.3) indicated that the gender of the translators did not drive their selection of talks with different function word patterns. This means that there was no difference between male and female translators in the English pattern of function words that they translated. However, in the EA-analysis, the main effect of TG was significant, meaning translators apply a gender bias to their translations.

In other words, male and female translators systematically employ function words in translation. This further suggests that given a translated text from English to Arabic, the EA-LIWC may be used to identify the gender of translator, but not the gender of the original speaker.

	Df	Wilks	approx F	num Df	den Df	Pr(>F)
SG	1	0.928	1.488	16	309	0.102
TG	1	0.769	5.788	16	309	0.000***
SG:TG	1	0.933	1.367	16	309	0.156
Residuals	324					

Note: *** p < 0.001, ** p< 0.01, * p< 0.05

Table 4.6. MANOVA of function word use in Arabic translations as a function of SG and TG

To identify the sources of gender differences in Arabic translation, a series of pair-wise t-tests was conducted on all the function words categories of EA-LIWC (Table 4.7). Recall that in English, it was found that women used more pronouns (especially personal pronouns, specifically first person singular and third person singular pronouns), more negatives, and fewer numbers. In the EA-LIWC analysis, no significant effect was found for any of these categories except a marginally significant effect for more first person singular pronouns and negatives by females. This suggests that the gender language style of the original speaker is, to a large degree, lost in translation to Arabic.

Following up on the strong main effect of TG, a series of post hoc t-tests confirmed that there were several EA-LIWC function word categories that were used differently by male and female translators (Table 4.8). For example, female translators were more likely to use more first person pronouns, conjunctions, and prepositions, and

fewer second person pronouns and quantity words than male translators. The effects for first person singular pronouns, second person pronouns and conjunctions resembled the language style of female English speakers. However, the reverse effect was found for prepositions and the effect of greater use of quantity words by male speakers/translators appeared. Overall, the translators expressed their own gender more than the gender style of the original speakers in their translations.

	Mean female	SD female	Mean male	SD male	t	df	p.value
Pronoun	15.93	2.54	16.38	2.21	-1.64	326	0.103
Personal pronoun	8.75	1.62	8.61	1.41	0.78	326	0.433
I	3.89	1.23	3.65	1.04	1.92	326	0.056*
We	0.60	0.44	0.61	0.40	-0.19	326	0.846
You	0.99	0.59	0.98	0.51	0.17	326	0.865
Shehe	2.83	0.97	2.99	0.99	-1.51	326	0.131
They	0.45	0.35	0.40	0.35	1.20	326	0.232
Impersonal pronoun	8.17	1.87	8.48	1.63	-1.58	326	0.115
Negatives	1.58	1.02	1.39	0.73	1.79	326	0.075
Conjunction	10.31	2.39	10.06	2.26	0.92	326	0.356
Prepositions	14.07	2.28	14.45	1.69	-1.63	326	0.104
Adverb	4.15	1.05	4.17	0.90	-0.16	326	0.875
Quantity	2.54	0.77	2.67	0.81	-1.48	326	0.140
Number	1.10	0.61	1.18	0.56	-1.25	326	0.211
Auxiliary verb	4.78	1.41	4.51	1.43	1.70	326	0.091
Inclusive	0.67	0.37	0.72	0.40	-1.06	326	0.290
Exclusive	3.17	1.20	3.16	0.98	0.11	326	0.912
Assent	0.16	0.19	0.14	0.17	0.82	326	0.411

Table 4.7. Rate of Arabic function words used in translation of male talks vs translation of female talks

	Mean female	SD female	Mean male	SD male	t	df	p.value
Pronoun	16.14	2.29	16.22	2.38	-0.28	326	0.783
Personal pronoun	8.83	1.56	8.61	1.47	1.14	326	0.254
I	4.04	0.98	3.64	1.15	3.10	326	0.002**
We	0.62	0.36	0.60	0.44	0.40	326	0.691
You	0.86	0.42	1.03	0.57	-2.79	326	0.006**
Shehe	2.91	0.91	2.93	1.01	-0.14	326	0.888
They	0.40	0.31	0.43	0.37	-0.57	326	0.572
Impersonal pronoun	8.17	1.60	8.42	1.78	-1.25	326	0.213
Negatives	1.53	0.62	1.44	0.93	0.96	326	0.340
Conjunction	11.10	2.51	9.82	2.14	4.21	326	0.000**
Prepositions	14.62	1.70	14.18	2.03	1.98	326	0.050*
Adverb	4.08	0.96	4.19	0.96	-0.90	326	0.370
Quantity	2.42	0.61	2.69	0.85	-3.25	326	0.001**
Number	1.11	0.53	1.16	0.60	-0.72	326	0.471
Auxiliary verb	4.46	1.27	4.67	1.48	-1.27	326	0.204
Inclusive	0.69	0.34	0.71	0.41	-0.45	326	0.651
Exclusive	3.19	0.78	3.16	1.16	0.26	326	0.797
Assent	0.12	0.12	0.15	0.20	-1.86	326	0.064

Table 4.8. Rate of function word use by male and female translators in Arabic

TRANSLATION AND GENDER OF SPEAKERS

In order to examine the relationship between function word use of English and Arabic for male and female speakers, a series of correlation comparison tests was designed to see how different translators interpret male and female speech. In other words, it is important to find the extent to which male speech is translated similarly or differently than female speech.

Illustration in Model 1 showed the correlations between function word use in the original English talks and the function word use in corresponding Arabic translations across gender of speakers. Certain steps were taken to compare correlation of English and Arabic function word use across translation. For example, the correlation between the rates of pronouns used in male speeches and the rates of Arabic pronouns used in corresponding translations (r_{male} = 0.189, df=195, $p<0.05$) were firstly computed as labeled "Correlation 2" in the horizontal arrows in Model 1. Similarly, the correlation between the rates of pronouns used in female speeches and the rates of pronouns used in corresponding Arabic translations (r_f = -0.093) were calculated as labeled "Correlation 2" in the horizontal arrows in Model 1. Then, Fisher z-transformation (Equation 1) was applied to both r_m and $r_{f,}$

$$Z = \frac{1}{2} ln \frac{1+r}{1-r} \qquad \text{(Equation 1)}$$

Note that the number of the male and female speeches are not equal (n_m= 196 and n_f=132). Next, Z-statistics calculated (Equation 2) to check the existence of statistical significances between two correlations.

$$Z = \frac{Z_{r_m} - Z_{r_f}}{\sqrt{\frac{1}{n_m - 3} + \frac{1}{n_f - 3}}} \qquad \text{(Equation 2)}$$

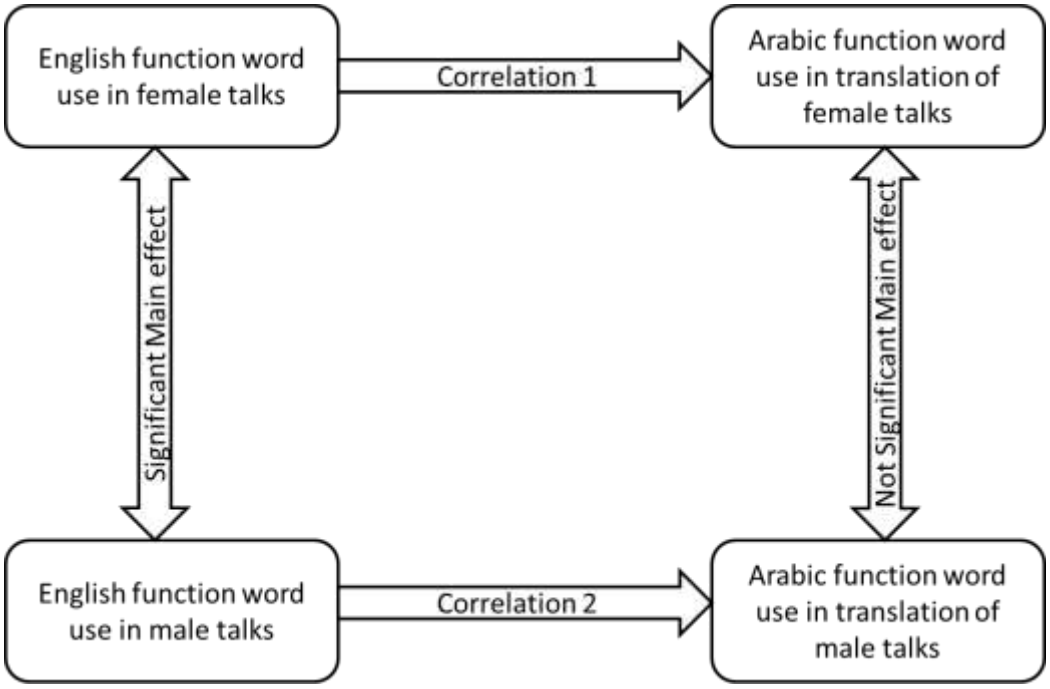
Finally, the *p* value associated with *Z* derived from the table. In case of English pronouns translated to Arabic, male speeches are translated significantly different from female speech ($r_m = 0.189$, r_f=-0.093, Z=-2.506, $p<.01$). It means that in the present sample, translators treated pronoun used by male and female speakers differently across translation.

This procedure of correlation analysis was repeated for all 18 categories of function words. The results are summarized in Table 4.9. Recall that, for each category of function words, two series of correlations for male and female speakers were compared to examine whether these two correlations for male speakers and female speakers are statistically identical. In other words, I examined the extent to which male and female talks are treated equally by translators.

The results in Table 4.9 indicate that, for most function word categories, there was a significant correlation between English and Arabic function word use. Considering the results of the t-tests of means across male and female speakers in the previous section (Table 4.4), I identified potential sources of similarities and differences in Arabic and English in terms of function word use. For those categories with non-zero correlations between English and Arabic, besides no significant difference between male and female speeches, the Arabic translations maintain their original English features (e.g. We, Adverb, and Inclusive Particle categories). On the other hand, some linguistic features

may be lost or gained in translation when categories with non-zero correlations between English and Arabic are significantly different between male and female correlations (e.g. Impersonal pronoun, Number, and Negative Particles categories)

As can be seen in Table 4.9, I identified categories of function words in which the similarities or differences between males and females in the original English speech are maintained in Arabic. Also, there are categories in which the similarities or differences between male and female speakers may be eliminated or added in Arabic translation.



Model 1. Comparison of English male and female speeches and corresponding translations to Arabic

	Correlation English and Arabic Function Words by Female Speakers	Correlation English and Arabic Function Words by Male Speakers	Z statistics	p	
Pronoun	-0.093	0.189	-2.506	0.006	*
Personal pronoun	-0.111	0.134	-2.166	0.015	*
I	0.155	0.363	-1.970	0.024	*
We	0.699	0.709	-0.181	0.428	
You	0.568	0.662	-1.324	0.093	
Shehe	0.006	0.065	-0.519	0.302	
They	0.694	0.652	0.672	0.251	
Impersonal pronoun	0.685	0.489	2.664	0.004	*
Negative Particles	0.953	0.829	5.963	0.000	*
Conjunction	0.298	0.133	1.520	0.064	
Preposition	0.580	0.303	3.073	0.001	*
Adverb	0.325	0.387	-0.626	0.266	
Quantity	0.396	0.427	-0.328	0.371	
Number	0.789	0.676	2.162	0.015	*
Auxiliary verb	0.278	0.359	-0.796	0.213	
Inclusive	0.119	0.177	-0.527	0.299	
Exclusive	0.626	0.476	1.916	0.028	*
Assent	0.446	0.207	2.372	0.009	*

Table 4.9: The correlation analysis of female speakers function words' ratio in English and Arabic versus male speakers function words' ratio in English and Arabic

TRANSLATION AND GENDER OF TRANSLATORS

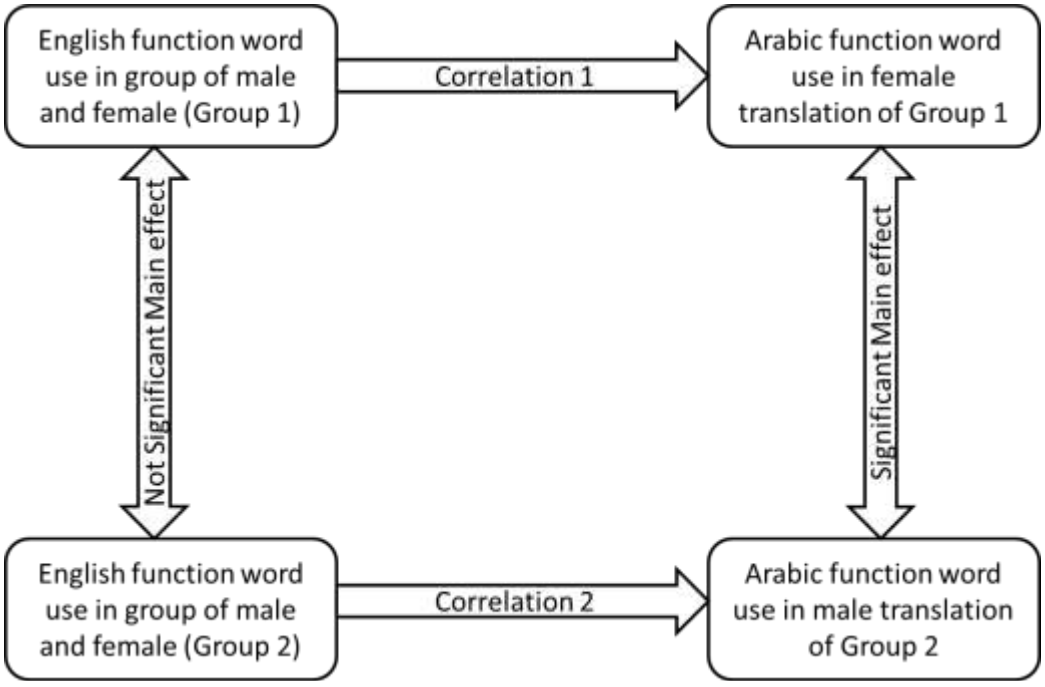
English and Arabic are two different languages in essence. Previous studies show that men and women speak differently in both English and Arabic. Also, male and female translators interpret differently. In order to examine all these findings in one setting, a series of correlation analyses were designed to investigate correlations between English and Arabic function word use in translation across gender of translators.

As in the previous section, the goal here is to examine the correlations between Arabic and English function word use in talks translated by male and by female translators, which are illustrated in the Model 2. For example, the average rate of conjunctions used in speeches translated by female translators was compared with the average rate of Arabic conjunctions used in corresponding translation (horizontal arrows in Model 2 labeled as *Correlation 1*). Similarly, the average rate of conjunctions used in speeches translated by male translators was compared with the average rate of Arabic conjunctions used in corresponding translation (horizontal arrows in Model 2 labeled as *Correlation 2*). Here, correlations between conjunction use in English and Arabic by female and male translators were $r_f=0.284$ and $r_m=0.228$ respectively. Then, using Equation 1 and Equation 2, Z statistics and p value calculated to test whether men and women translators were consistent in translating English conjunction to Arabic. The results show that, in current sample, there were no significant differences in translating conjunctions by male and female translators ($r_f=0.284$, $r_m=0.228$, $Z=0.477$, $p= 0.317$). This finding is important because it shows that regardless of possible differences in the

original text and final translation, translators were consistent in translating items in the conjunction word category.

This procedure was repeated for all 18 categories of function words. Two series of correlations for male and female translators were statistically examined to find the extent in which male and female translators treated the original texts equally in their translations in terms of function word use. The results are summarized in Table 4.10.

In the previous MANOVA analysis, the main effect of function word use for the talks selected by male and female translators in the original English talks was not significant. However, after translation to Arabic, a significant main effect was found with respect to Translator Gender. Comparing correlations in function word use between male and female translators, three potential sources were found in which males and females use function words differently in their translations such as You, Impersonal Pronouns and Inclusive Particle categories (Table 4.10). In other words, men and women translators were not consistent in translating these three categories of function words between English and Arabic. For example, there is a strong positive non-zero correlation between impersonal pronouns of English and Arabic used by male and female translators ($r_f = .35$ and $r_m=.65$). However, male translators tend to keep more impersonal pronouns in their translations to Arabic ($r_f = .35$, $r_m=.65$, $p <.01$). Also, the consistency in translation was seen in 15 categories of function words. It means that regardless of differences in original text and differences in language style of men and women translator, translators were consistent in their translation.



Model 2. Comparison of function word use in Group 1 speeches translated by Female versus function word use in Group 2 speeches translated by Male between English and Arabic

	Correlation English and Arabic Function words by Female Translators	Correlation English and Arabic Function words by Male Translators	z.stat	p.val	
Pronoun	-0.034	0.071	-0.828	0.204	
Personal pronoun	-0.033	0.037	-0.547	0.292	
I	0.327	0.242	0.728	0.233	
We	0.666	0.713	-0.703	0.241	
You	0.320	0.686	-4.002	0.000	*
Shehe	-0.008	0.034	-0.329	0.371	
They	0.679	0.668	0.145	0.442	
Impersonal pronoun	0.355	0.651	-3.195	0.001	*
Negative Particles	0.889	0.902	-0.503	0.307	
Conjunction	0.284	0.228	0.477	0.317	
Preposition	0.482	0.461	0.222	0.412	
Adverb	0.336	0.361	-0.231	0.409	
Quantity	0.362	0.426	-0.601	0.274	
Number	0.669	0.738	-1.069	0.143	
Auxiliary verb	0.318	0.322	-0.034	0.486	
Inclusive	-0.127	0.206	-2.649	0.004	*
Exclusive	0.537	0.552	-0.170	0.433	
Assent	0.276	0.339	-0.546	0.293	

Table 4.10: Correlation analysis of rate of function word use when female translate from English to Arabic versus male translating from English to Arabic

TRANSLATION SAMPLES

To illustrate the main effect of Translator Gender on translation, excerpts were selected from the sample of TED talks to illustrate the subtleties of Translator Gender effects on language style including female translations of female speakers, female translations of male speakers, male translations of female speakers and male translations of male speakers.. For each excerpt, I included the original English, the Arabic translation from TED talk corpus, and the literal word for word back translation of Arabic translations to English. Since no information exists about the gender of volunteers who transcribed the talks into English, no conclusion can be made about the difference in punctuation style in Arabic. However, it is clear that the Arabic translators segmented the text differently than the English speakers.

Female Translations of Female Speakers

In the Excerpt 1, a conditional type 2 sentences was used by female speaker in English used, however, female translator used real "if" in a construct. Also the first person pronoun "I" appeared as direct subject in English, across translation into Arabic, the female translator used the first person pronoun as object of preposition.

Excerpt 1:

" …And if I were to write a volume, …"

"...و إذا كان لي أن أكتب… "

/… and if(real) it-was for-me to write …/

(Thelma Golden, April 2010)

In the Excerpt 2, the female translator put more emphasis on the verb by using "will".

Excerpt 2:

" … it would be called …"

" … فسوف أطلق عليه …"

/ … then-will it-was-called to-it …/

(Thelma Golden, April 2010)

Female Translations of Male English Speakers

In the Excerpt 3, the sarcastic tone of the English is almost lost in translation, when the female translator focused on delivering meaning rather than the tone of the original speech. Interestingly, she added first person pronoun in her translation when she used "my talk is about …" instead of "this talk is about …". This observation supports the previous finding that women tend to use more first person pronouns than men in both English and Arabic.

Excerpt 3:

"…This talk is about righting writing wrongs …"

"…حديثي هو حول تصحيح الأخطاء المكتوبة …"

/… talk-my it-is around correction the-errors the-written .../

(Jamil Abu-Wardeh, August 2010)

In the Excerpt 4, again here in this short sentence, the first person plural pronoun appeared three times in the male talk in English, but the first person plural pronoun appeared five times in the corresponding Arabic translation by female translator.

Excerpt 4:

"… We need to take our responsibilities seriously, but not ourselves …"

"… ينبغي أن نعامل مسؤولياتنا بجدية ولكن علينا أن نكون أقل جدية مع أنفسنا …"

/… should to we-do responsibilities-our seriously, and-but to-us that we-were least responsibility with self-our …/

(Jamil Abu-Wardeh, August 2010)

Male Translations of Male Speakers

In the Excerpt 5, the translator was consistent respect to keeping the original English construct. He focused on the fluency of his translation into Arabic rather than keeping the informal tone of the original speech.

Excerpt 5:

> “As a magician I try to show things to people that seem impossible”
>
> “ بصفتي " ساحر " فأنا أحاول أن أري الناس الأشياء التي تبدو مستحيلة”
>
> /… as-I "magician" then-I I-try to I-see the-people the-things that it-seem impossible …/

In the Excerpt 6, it is cleared that the Arabic translator preferred not to use first person singular pronoun appeared in the English twice and he used third person singular instead. This observation also support the finding that men tend not to use first person pronouns than women use.

Excerpt 6:

> “… whether I'm holding my breath or shuffling a deck of cards, is pretty simple… ”
>
> “… سواء كان حبس الأنفس أو تقليب بعض أوراق اللعب هو أمر بسيط جدا …”
>
> /… whether it-was holding the-breath or shuffling some cards the-playing it-is task easy really…/

Male Translations of Female Speakers

In the Excerpt 7, the female speaker emphasized the original speaker's personality and experience by using "my" and "myself" in the first sentence. However, the male translator transformed the sentences to "I was failed to order …, not because of …". In other words, female speaker used first person singular pronoun twice, however, the same appeared only once in the Arabic translation by male.

Excerpt 7:

> "…My failure to procure myself a cup of sweet, green tea was not due to a simple misunderstanding"
>
> " لقد فشلت في الحصول على كوب من الشاي الحلو ليس بسبب عدم قدرتي على التعبير"
>
> /… if-have I-failed in the-gain of cup from the-tea the-sweet it-is-not because non- power-my over the-understanding …/

In the Excerpt 8, again similar to various examples pronoun "my" used by female speaker lost in translation. Also, the female speaker used feminine third person singular pronoun twice, but the male translator used masculine pronoun instead.

Excerpt 8:

> "…From my American perspective, when a paying customer makes a reasonable request based on her preferences, she has every right to have that request met …"
>
> للمفهوم الامريكي عندما يطلب الزبون طلبا معقولا تبعا لما يفضله هو فان طلباته يجب ان تلبى على الفور
>
> /… from-the-understanding the-American when he-request the-customer a-favor reasonable depending on whatever he-preferred he then-that requests-his it-must to fulfil over the-speed…/
>
> (Sheena Iyengar, July 2010)

In the Excerpt 9, the male translator focused more on the readability of his translation when he broke down the "that" clause into two separate sentences. Also, he used "don't match with" to translate "don't always hold true" to improve the strength of meaning in Arabic.

Excerpt 9:

> "… these beliefs are based on assumptions that don't always hold true in many countries, …"
>
> "هذه الاعتقادات مبنية على افتراضات لا تنطبق على عدة دول .. "
>
> /this-*feminine* the-beliefs based on assumptions not it-match with many countries/
>
> (Sheena Iyengar, July 2010)

DISCUSSION AND CONCLUSIONS

The advent of computer aid text analysis tools shed new perspectives and stimulated new insights to scholars in corpus linguistics, discourse analysts and psychologists (Kelle, 1997). For example, the method of word counting introduced by Paulsen and Martino (2004) employed to detect the original language of a given text. Also, adopting another method of word counting by Newman and his colleagues revealed significant language differences between men and women in English discourses (Newman et. al 2008).

This study has shown that the gender language style of an original English language speaker is not maintained in translations into Arabic. Many language style features disappear in translations from English to Arabic to the degree that one cannot distinguish the gender of the original English language speakers by looking at the Arabic translations. This study has further shown that the gender language style of the translator appears in translations into Arabic. Maybe, it is better to consider a translation as an

original product of the translator. Indeed, one may find traces of the gender of the translator more than that of the original speaker in translations.

The results indicate that the translations of two sets of neutral texts from both genders lead us to have two highly gender oriented sets of text. The gender difference appears in translations when looking at the male and female translations separately. Translator's gender plays an important role in producing a new language product based on understanding someone else's ideas in another language. This finding suggests that the culture of the target language is dominant in most translations.

It is unknown whether language style in English translations reflects gender language style differences that may be found in original Arabic texts. For male translations, there are stronger correlations in Inclusive Particles, You and Impersonal Pronouns. And, for female translations, there are weaker relationships between function words of English and Arabic. This finding suggests that modern female translators are more likely to adopt a free style in their reproduction of ideas in target language.

Although there is not a perfect parallel relationship between English and Arabic function words, the analysis has captured small but significant language style differences between male and female translators. First person singular pronouns, second person pronouns, conjunctions, and prepositions were used more by female translators, and quantity words were used more by male translators. It seems that women tend more to keep fidelity of a talk and men focus more on transparency and readability of a translation product regardless of the gender of the original talk.

Moreover, regardless of the translators' proficiency, there is gender difference across Arabic translations. Although, it was considered that all translators selected for this study had an acceptable credibility for Arabic translation, it seems that in some of the translations, the translators had difficulty to understand the meaning in English and/or had difficulty to render his/her understanding to Arabic. Therefore, the findings cannot be generalized to any translations from any languages; however, the findings give an insight of how gender may alter an accurate translation.

Future studies required to address whether level of proficiency in original and target language interact with gender as a factor. In other words, whether having higher level of proficiency can help translators to maintain the original tone of speaker across translation. Also, it will be more interesting to examine male and female translation of same pieces in several languages, and see whether the current finding still valid in other languages as well.

SUMMARY

The present chapter has attempted to address whether gender of a translator can affect the tone of translations into Arabic. To do that, the main focus was on the rate of function words use in 328 talks from English translated into Arabic in TED conference project. The sample consist of 196 male, 132 female, 241 male translations and 87 female translations.

A 2 by 2 MANOVA was conducted on Speaker's Gender and Translator's Gender (two independent variables) and 18 LIWC function word categories (dependent variable) in English. The overall main effect on SG was significant in English, meaning males and females speak differently when using function word categories as a group. However, there were not any significant main effects on TG and interaction between SG and TG.

Later, a 2 by 2 MANOVA was conducted on Speaker's Gender and Translator's Gender (two independent variables) and 18 LIWC function word categories (dependent variable) in Arabic. The overall main effect on SG and interaction term between SG and TG were not significant in Arabic; meaning having a translation in hand one cannot predict the gender of the original speaker and therefore there are aspects of language that get lost in translation. However, there was a significant main effect on TG in Arabic; meaning males and females translated differently when using function words in Arabic.

Chapter Five: Conclusions

In the process of translation, many features of the original language disappear. To study lost/gained features, I focused on function words of English and Arabic to examine whether such changes across translation are apparent. The Arabic LIWC, a computer software, was introduced which can count the occurrence of function words of Arabic and English texts in 18 different categories. The tools described in this paper can give people who work with translators and translations an insight in dimensions of a culture that may be invisible to someone not familiar with the other language or culture.

First, I translated the LIWC text analysis program, (Pennebaker, Booth, & Francis, 2007), into Arabic. Then, the grammatical dimensions of Arabic function words were determined that served as a basis for the Arabic LIWC designed for Arabic text. These same Arabic dimensions were used to force English words into the same categories. A large corpus of Modern Standard Arabic and English text files that have been translated in both directions were used to establish the equivalence of the translated dictionaries.

The Arabic LIWC provides tables of function words rate within a given texts in Arabic, texts translated to Arabic and texts translated from Arabic into English. Although the current Arabic LIWC is limited only to function word categories, scholars can add any category of interest to the program based on their research (i.e. emotion words, motion verbs, so on so forth).

Then, in order to show one application of such a text analysis tool within and across cultures, the influence of gender on translation has studied using TED talks translated from English into Arabic. I identified differences in language style between men and women in their English language TED talks, and examined whether these features are faithfully maintained or changed in translations to Arabic.

One of my goals was to introduce the Arabic LIWC as a tool which may capture slight linguistic differences of homogeneous group of subjects. Although our studied sample size was not large (N=328), I identified small but significant impact of gender on translation.

In summary, I found that there is a gender difference in language style in both English and Arabic. The distinguishing linguistic features of male and female speech are disappeared in Arabic translation when there is no translator bias (equal number of males and female translating).

There is a significant gender bias between male and female translators work. The function words of English and Arabic are highly correlated. I identified sources of similarity and differences in the linguistic style of female and male translators. In other words, I found the extent in which the gender of a text is maintained, gained or lost in translation with respect to my sample; and finally, the gender of author could affect a translation.

I showed that there are some areas in which the mind of men and women may work differently while translating oral talks from English to Arabic. Also, I addressed the

question of how men and women choose their own style to reproduce an idea in other language. Although I cannot explain why such phenomena exist, the method provided a new avenue in the study of good translation. In addition, the presented quantitative analysis may be useful for future human behavior studies adapting an Arabic corpus.

TED Open Translation Project provides rich resources of translated materials from English to several languages. These resources may be employed for future translation research.

Studying translation is one aspect of the Arabic LIWC. The Arabic LIWC can be used in several areas such as social science, political science and Arabic education. Since Egyptian Revolution of 2011, scholars in social and political science have contacted the author to use the Arabic LIWC monitoring events in the Arab worlds. For example, using the same methodology in studying gender across translation, one can perform a study to see whether the Arabic function word use in politicians' speech could predict the speaker political party. In other words, do people from different political party speak differently? The same set up is possible to examine whether function word use could be a predictive of speakers religion.

The Arabic LIWC can be used in Arabic educations in several ways. For example, a study could be conducted to examine whether function word use in a student writing could predict student's final grade (Arabic language assessment). Also, a research could be designed to examine whether any relation exists between function word use in students' essay and level of anxiety in the Arabic language classroom/test.

SUMMARY

Function words reflect the linguistic style of a language. A computer text analysis software was developed to count function words of Arabic in 18 different categories. The tools could bring an insight into dimensions of language transformation in Arabic-English, English-Arabic translations.

Studying function words of 328 talks translated from English to Arabic, there is significant difference between men and women speech in English. This language difference was disappeared across translation to Arabic. However, there is significant difference between male and female translation into Arabic. In summary, I found that there is a gender difference in language style in both English and Arabic and the gender of translator may affect Arabic translations.

References

Abu-Haidar, F. (1988). Male/female linguistic variation in a Baghdad community. In A. K. Irvine, R. B., Serjeant, & G. R. Smith (Eds.), *A Miscellany of Middle Eastern Articles: In memorian Thomas Muir Johnstone* (pp.1924-1983). Harlow: Longman.

Abu-Haidar, F. (1989). Are Iraqi women more prestige conscious than men? Sex differentiation in Baghdadi Arabic. *Language in Society*, 18, 471-481.

Agorni, Mirella, (1998). The voice of the 'Translatress': From Aphra Behn to Elizabeth Carter, *Yearbook of English Studies*, 28 pp.181-195.

Ahmed-Ullah, N. S. (2007, April 10). A new look at a holy text. *Chicago Tribune*. Retrieved May 1, 2014, from http://articles.chicagotribune.com/2007-04-10/news/0704100141_1_translation-muslims-and-non-muslims-arabic

Akram, M., and Sa'adeddin A. M., (1987) Three problem areas in teaching translating to native Arabic literates, *Anthropological Linguistics*, 29/2, 181-93.

Al-Ameed H. K., Al-Ketbi S. O., Al-Kaabi A. A., Al-Shebli K. S., Al-Shamsi N. S., Al-Nuaimi N. H., and Al-Muhairi S. S., (2005) Arabic Light Stemmer: A New Enhanced Approach. The Second International Conference on Innovations in Information Technology (IIT'05).

Al-Sughaiyer I & I. Al-Kharashi (2004). Arabic Morphological Analysis Techniques: A Comprehensive Survey, *Journal of the American Society for Information Science and Technology*, Vol. 55, Issue 3, PP. 189 – 213.

Ba'albaki, R. (2005). *Al-Mawrid: A Modern Arabic-English Dictionary*. Beirut, Lebanon: Dar El-Ilm Lilmalayin.

Bakhtiar, L. (2006). *The Sublime Quran*. Kazi Pubns Inc.

Berners-Lee, T. (2009). Tim Berners-Lee on the next web. TED Talks.

Biber, D. (1995). *Dimensions of Register Variation: A Cross-linguistic Comparison.* Cambridge: Cambridge University Press, Cambridge.

Bloom, A. H. (1981). *The Linguistic Shaping of Thought: A Study in the Impact of Language on Thinking in China and the West*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bloom, A. H. (1984). Caution - the words you use may affect what you say: A response to Au. *Cognition*, 17(3), 275-287.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43, 1-22.

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought*, pp. 61-79. Cambridge, MA, US: MIT Press.

Brown, S. A. (2005). Women Translators, in Roger E., Gillespie S.& D. Hopkins (2005). *The Oxford History of Literary Translation in English*. Oxford: Oxford UP, 111-20. Print.

Chalmers, H. (2004). *Royalist Women Writers, 1650-1689*. Clarendon Press.

Chung, C. K., & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality*, 42, 96-132.

Chung, C.K., & Pennebaker, J.W. (2007). The psychological functions of function words. In K. Fiedler (Ed.), *Social Communication* (pp. 343-359). New York: Psychology Press.

Darwish K., (2002). Al-stem: A light Arabic stemmer [Online]. Available: http://www.glue.umd.edu/~kareem/research.

Elbeheri, G., Everatt, J., Reid, G. and Mannai, H. a. (2006), Dyslexia assessment in Arabic. *Journal of Research in Special Educational Needs*, 6: 143–152. doi: 10.1111/j.1471-3802.2006.00072.x

Ibramhim, M. H. (1986). Standard and prestige language: A problem in Arabic sociolinguistics. *Anthropological Linguistics*, 28, 115-126.

Kashima, Y., & Kashima, E. S. (2003). Individualism, GNP, climate, and pronoun drop: Is individualism determined by affluence and climate, or does language use play a role? *Journal of Cross-Cultural Psychology*, 34, 125-134.

Kelle, U. (1997). *Computer-assisted analysis of qualitative data*. Methodology Institute.

Khoja S. (2001) APT: Arabic Part-of-Speech Tagger. *Proceedings of the Student Workshop at the Second Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL2001)*, Pittsburgh, Pennsylvania, pp. 20-26.

Krontiris, T. (1997). *Oppositional Voices: Women as Writers and Translators of Literature in the English Renaissance*. Routledge.

Lardiere, D. (1992). On the linguistic shaping of thought: Another response to Alfred Bloom. *Language in Society*, 21, 231-251.

Larkey L., Ballesteros L., and Connell M.E. (2002) Improving Stemming for Arabic Information Retrieval: Light Stemming and Co-occurrence Analysis. In SIGIR'02, August 11-15, 2002, Tampere, Finland, pages 275–282.

Long L. (2011), Women translators of Sacred Texts- female voices in translation, In Palusci O (2011) *Traduttrici: Female Voices across Languages*. Trento: Tangram Edizioni Scientifiche.

Ma'luf, L. (2000). *Al-Munjad fi Al-loqa wa Al-a'lam*. Beirut, Lebanon: Al-Maktaba Al-Sharqiyya.

Macfarquhar, N. (2007, March 24). New Translation Prompts Debate On Islamic Verse. *The New York Times*. Retrieved May 1, 2014, from http://www.nytimes.com/2007/03/25/us/25koran.html?ei=5088&en=835924e5b6d16c52&ex=1332475200&adxnnl=1&partner=rssnyt&emc=rss&pagewanted=all&adxnnlx=1398917519-ePZ1e8UIspPCLy85Pic1ig

McNamara, R. (2003). *Britain, Nasser and the Balance of Power in the Middle East, 1952-1967 from the Egyptian Revolution to the Six-Day War*. London, Portland, OR: Frank Cass.

Mulac, A., & Lundell, T. L. (1994). Effects of gender-linked language differences in adults' written discourse: Multivariate tests of language effects. *Language & Communication*, 14, 299–309.

Mulac, A., Bradac, J. J., & Gibbons, P. (2001). Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. *Human Communication Research*, 27, 121-152.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45, 211-236.

Ono, T., & Thompson, S. A. (2003). Japanese (w)atashi/ore/boku 'I': They're not just pronouns. *Cognitive Linguistics*, 14, 321-347.

Oren, M. (2002). *Six Days of War : June 1967 and the Making of the Modern Middle East*. Oxford: Oxford University Press.

Palusci, O. (2011). *Traduttrici: Female Voices across Languages*. Trento: Tangram Edizioni Scientifiche, Print.

Paulsen Jr, R. C., & Martino, M. J. (2004). *U.S. Patent No. 6,704,698*. Washington, DC: U.S. Patent and Trademark Office.

Pennebaker, J. W. (2011). *The Secret Life of Pronouns: What Our Words Say about Us*. New York, NY: Bloomsbury Press.

Pennebaker, J. W., Booth, R. J. & Francis, M. E. (2007). Linguistic Inquiry and Word Count: LIWC2007. Austin, TX: LIWC.net (www.liwc.net)

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic Inquiry and Word Count: LIWC2001. Austin, TX: LIWC.net (www.liwc.net)

Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A., & Booth, R.J. (2007). The development and psychometric properties of LIWC2007. Austin, TX: LIWC.net

Pennebaker, J.W., Mehl, M.R., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54, 547-577.

Peretz, M. (1992). A female poet and her male translator: A case study. *Trabalhos em Lingüística Aplicada*, 19(1).

Quotes by Gamal Abdel Nasser, 1954-1967., the center for online Judaic studies, from http://cojs.org/cojswiki/Quotes_by_Gamal_Abdel_Nasser,_1954-1967.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ramirez-Esparza, N., Pennebaker, J.W., Garcia, F.A., & Suria, R. (2007). La psicología del uso de las palabras: Un programa de comutadora que analiza textos en Español (The psychology of word use: A computer program that analyzes texts in Spanish). *Revista Mexicana de Psicología*, 24, 85-99.

Rumsey, A. (2003). Language, desire, and the ontogenesis of intersubjectivity. *Language & Communication*, 23, 169-187.

Saeed, A. T., & Fareh, S. (2006). Difficulties encountered by bilingual learners in translating Arabic "fa" into English. *The International Journal of Bilingual Education and Bilingualism*, 9, 19-33.

Schulz, E., Krahl, G., & Reuschel, W. (2000). *Standard Arabic: An Elementary Intermediate Course*. Cambridge: Cambridge University Press.

Scott, M., 2008, WordSmith Tools version 5, Liverpool: Lexical Analysis Software.

Selkirk, E. (1996). The Prosodic structure of Function Words. In J. L. Morgan & K. D. Mahwah (Eds), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. NJ: L. Erlbaum Associates.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. Language and Society, 1, 179-195.

Venuti, L. (1998). *The Scandals of Translation: Towards an Ethics of Difference.* New York: Routledge. (Winterson, Jeannette. *Written on the Body*. New York: Knopf, 1993.)

Walter, K. (1996). Gender, identity, and the political economy of language: Anglophone wives in Tunisia. *Language and Society*, 25, 515-555.

Wehr, H. & J. M. Cowan (1976). *A Dictionary of Modern Written Arabic*. Ithaca, NY: Spoken Language Services.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104, 7780-7785.

Wolf, M., Horn, A., Mehl, M., Haug, S., Pennebaker, J. W. & Kordy, H. (2008). Computergestützte quantitative Textanalyse: Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry and Word Count [Computer-aided quantitative text analysis: Equivalence and reliability of the German adaptation of the Linguistic Inquiry and Word Count]. *Diagnostica*, 2, 85-98.

Xu J., Fraser A., and Weischedel R., (2002) Trec 2001 cross-lingual retrieval at bbn. In E.M. Voorhees and D.K. Harman, editors, *The Tenth Text Retrieval Conference* (TREC 20).

Vita

Navid Hayeri was born in Tehran, Iran in December 1977, the son of Nasrin and Mohammadtaghi. Completing his work at Shahid Modarres high school in Tehran, Iran, in 1995, he entered the University of Tehran pursuing a degree in Mathematics. He received the degrees of Bachelor of Science in Mathematics in 2000. From 2000 to 2005, he worked for Iranian National Oil Company. Also, he enrolled in the University of Damghan, Iran, from 2002 to 2004 where he received the degree of Master of Science in Mathematics. In August 2005, he enrolled the Graduate School of The University of Texas at Austin. He received the degree of Master of Arts in Middle Eastern Studies on May 2008. He continued his study at The University of Texas in Foreign Language Education program. He studied under Professor James W. Pennebaker and Professor Michael C. Hillmann involved in various corpus linguistic and lexicography projects.

Email: navid.hayeri@gmail.com

This dissertation was typed by the author.