The summary of doctoral dissertation

Iranian and Hungarian EFL Students' Essays

A Corpus – Based Study

Akasha Ghaboosi

Supervisor: Dr. habil. Horváth József

Doctoral Program in English Applied Linguistics and TEFL/TESOL

Pécsi Tudományegyetem

2017

Table of contents

The dissertation theme and objectives

The study of learner language, mostly written, has become a major field within corpus linguistics in the past twenty years. Since the beginnings in the early 1990, with the development of the ICLE and the JPU Corpus, the field of learner corpus research (LCR) has established a number of standards specific to the design, analysis, and application of such corpora (Granger, 1998; Horváth, 2001, 2015). This dissertation aims to present the results of what can be considered a new vista in LCR: the qualitative and quantitative investigation of diachronic features of written language. I was interested in whether and how learner language has changed since the early years of the BA program, introduced in Hungary in 2006, on the one hand, and how the learner language in the BA programs in Hungary and Iran differ from each other. I have developed two small corpora; the first one is the *Happy Corpus* consisting of essays written at the University of Pécs as part of English majors' proficiency exams, and the second one is the *HI Corpus* encompassing the essays written by the Hungarian students at the University of Pécs and the Iranian students at Amol Islamic Azad University, who attended English writing courses. In this dissertation, I highlighted the most interesting results and suggest ways in which the results could inform future development of essay writing skills as well as testing procedures.

The literature review

While communicating our thoughts and ideas in a language in a readable form, we try to use proper grammar, punctuation and sentence structure beyond vocabulary, correct spelling and formatting. The better we write, the better we can impress our audience and make people get closer to what we mean in our interactions. The style and the proficiency in writing reflect our attitude, our voices and even our intentions. When we write a text, we produce a conceptual model of what is in our mind in different ways. One way is to expand the main idea as clearly as possible by applying micro-skills in writing skills such as aiming at an appropriate style, constituent sentences and coherent texts. Some researchers (Cumming & Riazi, 2000; Zimmerman, 2000) emphasize how people write and how different approaches and models are applied to writing process. By exploring EFL students' texts EFL students, teachers and scholars find what they need to know about learners' English language background knowledge (Bjork, Brauer, Reiecker & Jorgensen, 2003), their learning process and their writing problems. All of these factors may lead teachers and curriculum designers to provide more instructional practices for students' better understanding and their progress in learning English language skills, such as writing. By monitoring students' strategies for planning what they write, revising and editing their texts, teachers can find students' difficulties in using proper vocabulary, collocations and structures in more academic written products.

According to Goldsmith (2011, p. 21) "Writing involves notions of distribution while proposing new platform of receivership. Words might be written not to be read but rather to be shared, moved, and manipulated". EFL students need to understand what they want to write about and how to carry through writing process meaningfully and successfully. Such writing processes lead

EFL students to interact with others by printing their personal and social thoughts on a sheet of paper. Transmitting ideas to each other in written form can be inspected in variety forms of research including text-oriented research, writer-oriented research and reader-oriented research. These approaches will be discussed in what follows.

Hyland (2002, p. 8) describes "writing as a textual product, a coherent arrangements of elements structured according to a system of rules which makes human interaction possible by exchanging ideas from one person's mind to another via language." Paltridge and Phakiti (2010, p.192) also draw attention to descriptive, analytical and critical views of writing in which the emergence of words in a sentence, the reason it occurs, and its relation to social situations in a cohesive way are investigated. Moreover, they claim that a text is an independent sample which can be analyzed in various genres in different periods of time, written by individuals. Kastouli (2005) also mentions that the vital goal in doing text-oriented research is "the expansion of literacy to increase communicative competence […] in the view of various dimensions of social multiliteracy in a multicultural society." In general, text-oriented research enables us to focus on the final piece of writing and by studying the outcome we can measure the text(s) against criteria of vocabulary use, grammatical use, and mechanical considerations such as spelling and punctuation, and even its content and how sentences are formulated and organized.

Flower (1989) and Prior (1998) proposed that how writers perform in a text and how they reflect on an event, especially a 'social act' in a 'specific context' is more important than what their personal ideas are. Flower (1998, p. 288) explained that writing as a situated act is "the effect of prior knowledge, assumptions and expectations together with features of the writing environment which selectively tap knowledge and trigger specific processes". Such a 'social act' can also be seen in ethnographic texts which display people's culture and customs, and similarities and differences among them are described while each culture is observed in its own original environment (Ramanathan & Atkinson, 1999). In other words, how EFL students put their thoughts on paper to interact with their readers, how they transmit their messages to their audience, and how readers react to what is written based on their perception of vocabulary and grammar students use in the texts about their experiences, beliefs and/or favors are the reflections of social acts in students' essays which is the focus of writer-oriented research.

Authors may attempt to write effectively and professionally while not spending too much time writing and creating their texts. Creating clear and comprehensible texts requires an open lively mind which has a strong close connection with the reader. Such a connection makes the impact authors need, which is expressing their ideas clearly and drawing their thoughts confidently for readers' better understanding. Hyland (2010, p. 194) refers to reader-oriented research, as "the mediation between writers/institutions/ cultures, and conventions describing the stages that help writers to set out their thoughts in ways readers can easily follow and identifying salient features of texts which allow them to engage effectively with their readers." He also points out that for readers' deep understanding of writers' messages spread over the texts, writers can apply

different genres in scribing their texts such as genre portfolio, comparative tasks, and audience analysis (Hyland, 2007, p. 84).

Considering the constructive role of understanding the meaning of a text and what readers can create from it, Lewis (2000) indicated that texts help readers imagine particular moments, places and events by describing them in detail. Thus, it is understood that there is a direct relation between readers' personal and emotional life experiences and what they read and what they decode through the text they read.

Linell (1982) and Hyland (2007) stated that a text has a set of characters which specify its nature and it can be seen through different angles and all readers in different levels of social classes can understand and communicate with it in different locations. It is important for writers to have deep knowledge of writing to make a coherent passage covering their readers' expectations

When we observe English language learners, we recognize how different they are in term of needs, background experiences, cultures, native languages, personalities, and interests. Based on readers' visualization and what they picture in their minds and how texts reflect their experiences, Tomlinson and Imbeau (2010) state that there should be a balance between what EFL learners need to learn to produce the language they acquire properly and what they are looking for in learning the language parallel with their individual needs. Tomlinson and Imbeau (2010) point out that those language learners' interests are essential in motivating them to learn the language. It is also important to observe language learners closely to find out how they "approach the task of learning". To attract learners' attention and keep them highly motivated, a practical beneficial instruction is recommended, which not only covers teachers' expectations but also supports individual language learners' desire in mastering a foreign language. It is also necessary to make content understandable for learners so that they can develop their language skills by their own efforts and their teachers' instructions (Echevarria, Vogt & Short, 2008). In order to assist learners to save time and to keep their track of conveying their messages to readers, Hyland (2007) introduces a cycle which includes five stages in the process of learning how to write meaningful texts. The stages are: setting the context, modeling, joint construction, independent construction, and comparing. It is vital for writers to know how to lay out the content, how to choose a style to write texts with it, and how to elaborate on the content.

The topics of discussing how EFL learners acquire a language, the difficulties they encounter through developing such ability, and how EFL learners' writing problems can be solved attract scholars to do research in these areas. Sasaki (2000), among others, was interested in EFL students' writing fluency and how EFL learners act during writing learning process. Some others, such as Cumming (2001) explored the quality of texts students write in English and how social and cultural background affected their learning process and what they produced as texts. Linguistic and pedagogical experts were also keen on studying the factors which distinguish successful writers from less skilled writers (Kariminia, 2013; Sahragard & Mallahi, 2014), the strategies which EFL learners at different levels of proficiency apply in creating their texts

(Farhandezh et al, 2012), teachers' knowledge and ability in instructing students on employing right strategies in writing coherently and properly (Mozaheb, et al, 2013), how original students' essays are considering voice, personal narrative, and role-play (Horváth, 2015), and EFL students' attitudes toward writing in English (Gholaminezhad et al, 2013).

Academic writers are eager to make an effective piece of writing by brainstorming, planning, generating their ideas (describing, comparing, associating, analyzing, applying, and arguing their ideas to reflect their agreement or disagreement), questioning and revising, rearranging and deleting the irrelevant and incoherent parts in their text, re-reading/proof-reading and producing different drafts, and at last making their final draft. Skilled EFL learners take all aforementioned steps to create something which needs time and requires reliable feedback and information to convey their messages to their readers. There are various approaches which students use in writing their texts to not only evaluate their language learning potential and improvement as a communication tool, but to transfer their experiences and feelings to their readers.

Expressive writing is the way of writing writers choose to discover their deepest emotions and feelings about their past, present and/or future, people they know including their family members and friends, what they like or dislike. Authors reveal their ideas to themselves at first place without being evaluated or judged by others. They attempt to make their personal thoughts, observations and analyses visible in texts they create (Foulk & Hoover, 1996).

Beaugrande (1982, 1984), Calkins (1983), Emig (1983) and Graves (1984) noted that the cognitive approach is a psychologically - and writer-based and scientific aspect in which writers' personally performance, pre-planning and editing, can be directly observed. Their studies showed that the errors they make during writing texts provides a useful source of data for language teachers' better comprehension of how writing activity takes place in EFL learners' mind and action. Teachers can also detect the difference between what weak and skilled writers do and why there are various styles differing from person to person. By focusing on learners' understanding of proper writing, teachers can help learners identify the structure and lexis of the language they learn beside the style they feel more comfortable in expressing their intentions and thoughts in their texts. Teachers present the rules, especially the grammatical ones, in a good writing, and give tasks to learners to practice their writing skill which leads to learners' production.

There are two distinctive outlooks of a social viewpoint to the process of writing. One of the perspectives is 'educational ethnography' which goes beyond reporting events and describing personal experiences in real. It refers to how something in a social community is presented while explaining people's cultures, habits and interactive social differences in that specific group. Graves (1984) and North (1987) propounded that if writers intend to create a piece of writing without its natural context, the main purpose of such a text disappears and it loses its powerful influence on readers' mind. The second effective social perspective in the process of writing is sociological linguistics in which Grabe and Kaplan (1997, p. 97) defined "writing as a process of

discovery […] and its development progresses as a problem-solving activity […] and drawing and talking are means for pre-writing and rehearsal of aspects of society encompassing cultural norms, individuals' expectations and how language usage affects societies and communities". It is assumed that writers are inspired by whatever they observe in the environment in which they are in contact with its members. They have their own ideas and viewpoints, try to elaborate them by thinking loudly on pieces of paper, examine their debates, and finally develop them by omitting the irrelevant and inept parts or adding more logical and convenient parts matching the acceptable norms in their social community. Taking all these steps makes writers' texts more highlighted and noticeable in readers' mind.

One of the main goals EFL learners attempt to achieve is producing meaningful and real communication in writing when they are keen to find out about the environment or to express themselves properly and correctly. Some EFL learners can acquire much better when there is a model text which is a pattern of an acceptable professional text. Modeling is a product-based approach which is defined by Tangermpoon (2008) as EFL learners' production from pre-writing to composing and correcting.

As a communication tool, a language can provide combination of words and expressions in millions of sentences and help people interact with each other and talk about what they feel and what they believe in life. Language reflects the nature of the real world, where people live and their lifestyles, their social identities, and even their specific dialects in a specific area of a country. Corpus linguistics is a method which assists researchers to get advantage of real life texts, both spoken and written, to analyze a language in terms of its vocabulary.

Sinclair (2005) and Taylor (2009) used corpus analysis to do their research in the different fields such as education and social sciences. Kennedy (1998, p. 7) stated that "Corpus linguistics is based on bodies of texts as the domain of study and as the source of evidence for linguistic description and argumentation." Corpus linguistics is a method in which a language is analyzed to give logical answers to linguistic questions and presents a unique view of how dynamic a language is by studying the rules which exist in a language and its patterns, such as grammatical or lexical. Kennedy also referred to the use of corpus linguistics in showing the 'ongoing progress in language' and investigating specific linguistic issues including how a language is acquired and changed during periods of time. Taylor (2009) emphasized the use of corpus as a tool or a method which helps researchers figure out how languages are studied. According to McEnery & Wilson (2001, p.197), "corpus is anybody of text" which is the collected samples and examples of spoken or/and written language.

While looking for a word in a database, we see some comments, explanations or presentational markups which are attached to the original text. Such machine-readable information provided by special software and programs are called annotations which is the practice of adding interpretative linguistic information to a corpus (Leech, 2005). Annotations usually refer to a specific part of the original data. After compiling samples used in a corpus,

researchers need to add further information to the "raw corpus", giving linguistic information at different levels such as parts of speech, syntactic, supersegmental, semantic, discoursal, and stylistic (Sinclair, 1991; Baker, 2006). Such levels can be analyzed through the matter of annotating texts in a corpus. Moreover, annotation can also be studied in corpus linguistics as a manual examination of corpus, an automatic analysis of corpus, functional segments in texts stored in a corpus (Leech, 2005). Corpus annotation can be classified into 'part-of-speech (POS), lemmatization, syntactical (parsing), semantic (domain classifications), co-reference (discourse), pragmatic (speech acts – dialogues), stylistic, and research specific (ad hoc) aspects (Leech, 2005; McEnery and Hardie, 2012).

As a reflection of real life and how language is produced among both native and non-native speakers, scholars need to have a database in which written texts and spoken transcripts are gathered and stored for researchers' studies in different fields like linguistics, focusing on issues such as lexico-grammar. In order to get proper and adequate results while applying the categories above, a learner corpus should be designed well and carefully considering ten key design principles proposed by Sinclair (2005), consisting of content selection, representative, contrast, structural criteria, annotation, sample size, documentation, balance, topic, and homogeneity.

Studying learners corpora is useful in figuring out the difficulties which language learners encounter through the process of a language acquisition, and as Kennedy (1998, p. 182) pointed out, '… one can comprehend language transfers, errors and other difficulties non-native speakers may have when learning English'. Hunsten (2002) also touches on learners' corpora, their focus on specific aspect of EFL learners' language and the development of peculiar features in languages. It seems that the combination of these definitions can fulfill the deep vision of what learner corpora are, and why they are useful in the field of applied linguistics. Therefore, one of the essential issues in a corpus-based study is to build learner corpora "underlying representations of learners at a particular stage in the process of L2 acquisition and of the development constraints that limit L2 production. The language produced by learners is the central source for these mental processes, whether spontaneously or through data elicitation procedures" (Myles, 2005, p. 374).

The collected data from EFL learners is stored in electronic collections of authentic foreign/second language texts which can deal with various target languages such as English (Granger, 2003, p. 465) and Dutch (Cucchiarini et al., 2008). Beside the importance of students' texts collection, Botley and Dillah (2007, p.75) uttered that "such a collection is not considered a corpus proper unless it is planned and collected according to clear and sound design principles". Pravec (2002, p. 81) stated that learner corpora is a database which provides a "deviation from the standard, i.e. the language of the native speakers of a particular language". Learner corpora can even provide a real data of how language learners use linguistics variables in formal or informal status (Preston & Fasold, 2007; Tarone, 2009). Many language corpora have been conducted for various purposes which are not only in English but also in other target languages such as French. Cambridge Learner Corpus (CLC), Corpus of Academic Learner English

(CALE), the International Corpus of Learner English (ICLE), British Academic Writing English Corpus (BAWE), and the JPU Corpus can be considered as useful learner corpora models.

Tsui (2005) is one of the scholars studying about the use of corpora in language teaching. Her observations reveal that while students are in the process of acquiring English language, they encounter a lot of problems in understanding the meaning of confusing words, using connective markers in sentences, applying countable and non-countable nouns, and following the correct form of number agreement in their produced sentences. Tsui (2005) uttered that corpora can be a practical source for teachers which can give teachers precise information about the aforementioned aspects above.

Another use of corpora in language teaching and learning is when teachers lead students to use data in different corpora to be introduced with the texts native speakers compiled from native speakers. Teachers can teach grammatical patterns to students who can practice such rules by making sentences themselves and compare their sentences with the ones produced by native-speakers and gathered in corpora such as BNC. In their research on the role of corpora in grammar teaching, McEnery, Wilson and Baker indicated that "a corpus should be at least integrated into teaching by which grammar teaching may be more effective … and more importantly be rated more positively by learners" (1997, p. 15).

After compiling the data and identifying the area of research and research questions or hypotheses, scholars decide which statistical or logical techniques they tend to apply in describing, analyzing and evaluating the data in the corpora they design on their own or the ones which are available in different sources such as online databases. Corpora can be analyzed qualitatively and/or quantitatively which is explained in the following sections. According to Hasko (2012, p.1) "qualitative corpus analysis is a methodology for pursuing in-depth investigations of linguistic phenomena, as grounded in the context of authentic, communicative situations that are digitally stored as language corpora and made available for access, retrieval, and analysis via computer." Given (2008) defined quantitative research as the systematic empirical study in which the data are analyzed statistically by means of software, such as Excel 2007, used in the current study. While trying to examine their hypotheses or answers their questions in a study, researchers focus on more general aspects at first and then they narrow their research and get to more specific conclusion. Therefore, it is implied that "quantitative research is deductive: based on already known theory we develop hypotheses, which we then try to prove (or disprove) in the course of our empirical investigation" (Rasinger, 2008, p. 9).

Due to the importance and impact of the English language as a global language, EFL learners are eager to study English as a foreign language at colleges to gain undergraduate or even graduate degrees by applying to local and/or international schools. They are also keen on getting to know more about various cultures and communicating with other people who speak English as their native language or the second language in their countries as a lingua franca. By learning the English language, Iranian and Hungarian students have the opportunity to study and work in the

English language spoken countries, especially the Great Britain and the United States. Since studying or working abroad requires improving and demonstrating reasonable and sufficient language skills, especially writing, which is the main purpose of the current study, EFL instructors intend to provide useful materials for their students and ensure that EFL learners get access to the core curriculum to acquire such skills better in a more academic way and not to encounter serious problems in interacting with English language speakers.

The studies showed that English departments of Iranian universities pay more attention to more general aspects in each field, particularly linguistics and literature. Unlike Iranian higher education materials design, the Hungarian curriculum focuses on more specific areas: American, British, Canadian and Irish cultures, history and civilizations. Based on my understanding of the Iranian government's ideology, belief, and conservative manner about Islamic regulations, I think the education directors try to keep students away from foreign cultures such as American and British to preserve their established Islamic perspectives and to avoid their interference on the current policies in Iran. However, it seems impossible to decline the presence of the internet and its virtual world which makes quick access for its users to find answers to their questions by a click. Besides the differences, findings presented the importance of writing skills in academic curricula in both Iranian and Hungarian universities where essay, academic and advanced writing skills and styles, syntax, lexicon, and morphology play a vital role in each English field of study and their effective influences on students' goals.

The study of learner language and how students make progress in the process of English language acquisition, especially writing which plays a vital role in academic stages, has become a major field of corpus linguistics in the past twenty years. Since the beginning of 1990, with the development of the ICLE and the JPU Corpus, the field of learner corpus research (LCR) has established a number of standards specific to the design, analysis, and application of such corpora (Granger, 1998; Horváth, 2001, 2013). Considering those standards I present the initial results of what can be considered a new vista in LCR: the qualitative and quantitative investigation of diachronic features of written language. I was interested in how EFL learners' writing skills have changed since the early years of the BA program in Hungary and Iran by introducing two corpora: *Happy Corpus*, and *HI Corpus*.

The Happy corpus

My main focus in the *Happy Corpus* analyses was on the learner language changes since the early years of the BA program in 2006 when the first BA English programs were offered in Hungary. I have developed a small corpus of students' essays written in English proficiency exams held at the University of Pécs. The corpus encompasses two subcorpora: the 2009 and the 2014 samples. Students were instructed to write a 300-word essay about one of the two topics presented to them. I assumed that the topic of "Don't Worry; Be Happy" would be more popular than the other topic among the proficiency exam attendants in 2014 since I had seen the same result among the participants in the 2009 exam. The content and language features of the two

subcorpora were analyzed to determine how students used linguistic patterns in the texts syntactically and semantically. The emphasis on such aspects could help examiners make a distinction between less and more skilled students based on assessment criteria applied in the exam. I pinpointed how the corpus was developed, and the results that the analysis in proficiency exams in 2009 and 2014 yielded: the focus was on the vocabulary profile, the ordering patterns, keywords and collocations, and the annotated patterns.

The diachronic study of Hungarian students' writing proficiency level between 2009 and 2014 revealed that there was just 1% difference between the learners' vocabulary lexical density in 2009 and 2014. There was a 1% difference of K1 between the 2009 and 2014 subcorpora essays. There was also a minor increase in using academic words between essays collected in 2009 and 2014, which reflected little growth in students' advanced writing skills level. The content and the length of texts turned my attention to students' interest in writing long essays, time to time, and their desire to write more about others than themselves, especially the ones about the students' friends in 2009. The analysis of common keywords statistic significance showed a large chi value of 59.35 for preceding, and chi value of 106.50 for following parts of speech in 2009 and 2014 samples and rejected their equality hypothesis of those parts of speech distribution in the 2009 and 2014 subcorpora.

The HI corpus

The study of the *HI Corpus* showed the results of diachronic features in language learners' essays, investigated both qualitatively and quantitatively. I was eager to observe how diversely English program students apply English lexicons to express their ideas in the scene they create in essays in Iran and Hungary. I have compiled a bicultural corpus in which I collected and gathered both Iranian and Hungarian BA English program students' essays. The students were participating in English writing courses. There are two subcorpora in the *HI* Corpus: Iranian subcorpus and Hungarian subcorpus. The Iranian data included 67 texts written by students participating in English writing programs in Amol Azad University in 2013 spring, analyzed in autumn 2013. The Hungarian data contains 47 texts created by students attending English writing courses in the University of Pécs in winter 2014, analyzed in spring 2015. I was interested to discover the content and language features students, with different cultural backgrounds, use in their essays syntactically and semantically. I explained how the corpus was developed, and the essays analysis results of Iranian and Hungarian subcorpora was demonstrated to discuss vocabulary occurrences, keywords and collocations students used more in essays.

The comparative research of the *HI Corpus*, demonstrated a 1% or 2% difference between the vocabulary lexical density of the Iranian and Hungarian subcorpora, and the use of academic words percentages in both Hungarian and Iranian students' texts in all categories, their anxiety, their dreams, their memories, their favorite English texts, and their purpose of studying English. The percentages of K1 and K2 of students' texts varied in different categories. The analyses showed that there was a low difference of academic words in favorite English texts, the higher

percentage of functional words in Hungarian dream and fantasy subcorpora, and the higher percentage of functional words in the Iranian memories subcorpus. The length and content comparison revealed the students' favor in creating long stories or wrap everything up in just short paragraphs. The texts also showed that Hungarian and Iranian students had different tastes in choosing a topic to write and the stories they made up based on their ordinary life, which might be as a result of their different thoughts, lifestyles and customs. The analysis of common keywords statistic significance showed a large chi value of 62 for preceding collocations, and chi value of 52 for following parts of speech in the Iranian and Hungarian samples and rejected their equality hypothesis of those parts of speech distribution in the Iranian and Hungarian subcorpora.

Conclusions and pedagogical implications

The keywords and the collocation unit analyses in both *HI Corpus* and *HAPPY Corpus* have demonstrated various possibilities of using keywords different parts of speech and collocation units which can assist pedagogy experts and linguists pinpoint the students' difficulties and misunderstandings in applying a word or a combination of words correctly in sentences. Detecting the problematic points can give the curriculum designers the idea of providing more practical and useful tasks in which students' involvement in writing and developing their skills take the higher levels of proficiency.

Further empirical work will be conducted in the area of international learner corpora. Such a project can make a vast field of study to understand how different EFL learners acquire and use EFL not only in academic fields but also in their daily lives.

References

Baker, P., Wilson, A. & McEnery, T. (1997). Teaching grammar again after twenty years: Corpus based help for grammar teaching. *RECALL Journal*, *9*(2), 8 - 17.

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Beaugrande, R. (1982). Psychology and composition: Past, present, future. In Martin Nystrad (Ed.) *What writers know: The language, process and structure of written composition.* (pp. 211 - 267). London: Academic Press.

Beaugrande, R. (1984). *Text production: Toward a science of composition.* Norwood, NJ: Ablex.

Björk, L., G. Bräuer, Reinecker, L. & Jörgensen, P. S. (2003). *Teaching academic writing in European higher education.* Kluwer: Dordrecht.

Botley, S., & Dillah, D. (2007). Investigating spelling errors in a Malaysian learner corpus. *Malaysian Journal of ELT Research*, *3*, 74 - 93.

Calkins, L. (1983). *Lessons from a child.* London, NH: Heinemann

Cucchiarini, C., Doremalen, J. J. H. C. & Strik, H. (2008). DISCO: Development and integration of speech technology into courseware for learning language. *In proceeding of interspeech* (pp. 2791 - 2794). Brisbane.

Cumming, A. & Riazi, A. (2000). Building models of adult second language writing instruction. *Learning and Instruction, 10*(1), 55 - 71.

Cumming, A. (2001). Learning to write in a second language: Two decades of research. *IJES, 1*(2). 1 – 23.

Echevarria, J., Vogt, M. E. & D. Short. (2008). *Making context comprehensible for English language learners: The SIOP model.* Boston, MA: Allyn & Bacon.

Emig, J. (1983). *The web of meaning.* Upper Montclair, NJ: Boynton/Cook.

Farhandezh, F. S. & J. Othman. (2012). An investigation into writing strategies of Iranian EFL undergraduate learners. *World Applied Sciences Journal, 18*(8), 1148 - 1157.

Flower, L. (1989). Cognition, context and theory building. *College Composition and Communication*, *40*, 282 - 311.

Foulk, D. & Hoover, E. (1996). *Incorporating expressive writing into the classroom.* University of Minnesota, Minnesota: The Board of Regents.

Gholaminezhad, R., Moinzadeh, M. Youhanaee, M & Ghobadirad, H. (2013). Writing attitudes of Iranian EFL students: A qualitative study. *Journal of Language Teaching and Research, 4*(5), 1138 - 1145.

Given, L. M. (2008). *The Sage encyclopedia of quantitative research method*. Los Angeles, California: Sage Publications.

Goldsmith, K. (2011). *Against expression: an anthology of conceptual writing.* Co-edited with Craig Dworkin. Evanston, Illinois: Northwestern University Press.

Grabe, W. & Kaplan, R. B. (1997). *Theory & practice of writing.* New York: Longman.

Granger, S. (1998). *Learner English on computer*. London: Longman.

Granger, S. (2003). Error-tagged learner corpora and CALL: A promising Synergy. *CALICO Journal*, *20*(3), 465 - 480.

Graves, D. (1984). *A researcher learns to write.* London, NH: Heinemann.

Hasko, V. (2012). Quantitative corpus analysis. In carol A. Chappelle (Ed.), *The Encyclopedia of applied linguistics*. Malden, MA: Wiley Blackwell.

Horváth, J. (2001). *Advanced writing in English as a foreign language: A corpus-based study of processes and products*. Pécs: Lingua Franca Csoport. http://bit.ly/9ionQx.

Horváth, J. (2015). "Two worlds and us" Originality in university students ebooks. *Modern Trends in Foreign Language Teaching and Applied Linguistics*, 90 - 100.

Hunston, S. (2002), *Corpora in applied linguistics*. Cambridge: Applied Linguistics Press.

Hyland, K. (2002). *Teaching and researching writing.* London: Pearson Education Limited.

Hyland, K. (2007). Genre Pedagogy: Language, Literacy and L2 writing instructions. *Journal of Second Language, 16,* 148 - 164.

Hyland, K. (2010). Researching writing. In B. Paltridge & A. Phakiti (eds.) *Continuum companion to research methods in Applied Linguistics* (pp. 191 - 204). London: Continuum International publishing.

Kariminia, A. (2013). Writing research articles in English: Insights from Iranian university teachers' of TEFL. *Procedia – Social and Behavioral Sciences,* 70, 901 - 914.

Kastouli, T. (2005). *Writing in contexts textual practices and learning process in sociocultural settings.* New York: Springer Science and Business Media Inc.

Kennedy, G. (1998). *Direct use of corpora in teaching*. London: Longman.

Leech, G. (2005). Adding linguistic annotation. In M. Waynne (Ed.), *Developing linguistic corpora: a good guide to good practice* (pp. 17 - 29). Oxford: Oxbrow Books.

Lewis, C. (2000). Critical issues: Limits of identifications: The personal, pleasure, and critical in reader response. *Journal of Literacy Research, 32*(2), 253 - 266.

Linell, P. (1982). *The written language bias in linguistics* (SIC = Studies in Communication 2). Linkoping: Department of communication studies.

McEnery, T. & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.

McEnery, T. & Hardie, A. (2012). *Corpus linguistics*. Cambridge: Cambridge University Press.

Mozaheb, A. M, Seifoori, Z. & Beigi, A. B. (2013). Effective Iranian EFL writing teachers. *Procedia Social and Behavioral Sciences, 70*, 18 - 27.

Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, *21*(4), 371 - 391.

North, S. (1987). *The making of knowledge in composition.* London, NH: Heinemann.

Paltridge, B. & Phakiti, A. (2010). *Continuum companion to research methods in applied linguistics.* London: Continuum International publishing.

Pravec, N., (2002). Survey of learner corpora. *ICAME*, *26*, 81 - 114.

Preston, D. & Fasold, R. (2007). The psycholinguistic unity of inherent variability: Old occam whips out his razor. In R. Bayley & C. Lucas (Eds.), *Sociolinguistic variation* (pp 45 - 69). Cambridge: Cambridge University Press.

Prior, P. (1998). *Writing/disciplinarity: a sociohistoric account of literature activity in the academy*. Hillsdale, NJ: Lawrence Erlbaum.

Ramanathan, V. & Atkinson, D. (1999). Ethnographic approaches and methods in L2 writing research: a critical guide and review. *Applied Linguistics*, *20*(1), 44 - 70.

Rasinger, S. M. (2008), *Quantitative research in linguistics*. London: Continuum International Publishing Group.

Sahragard, R. & Mallahi, O. (2014). Relationship between Iranian EFL learners' language learning styles, writing proficiency and self-assessment. *Procedia-Social and Behavioral Sciences*, *98*, 1611 - 1620.

Sasaki, M. (2000). Toward an empirical model of EFL writing process. *Journal of second language writing. 9*(3), 259 - 291.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair. J. (2004). *Trust the language: language, corpus and discourse*. London: Routledge.

Sinclair, J. (2005). How to build a corpus. In M. Waynne (Ed.), *Developing linguistic corpora: a guide to good practice* (pp. 79 - 83). Oxford: Oxford books.

Tangpermpoon, T. (2008). Integrated approaches to improve students' writing skills for English major students. *ABAC Journal*, *28*(2), 1 - 9.

Tarone, E. (2009). A Sociolinguistic perspective on interaction in second language acquisition. In A. Mackey and C. Polio (Eds.), *Multiple perspectives on interaction: second language research in honour of Susan M. Gass*. New York: Routledge.

Taylor, C. (2009). What is corpus linguistics?. *ICAME Journal. 32*, 179-200.

Tsui, A. B. M. (2005). ESL teachers' questions and corpus evidence. *International Journal of Corpus Linguistics. 10*(3), 335 - 356.

Zimmermann, R. (2000). L2 writing: Subprocesses, a model of formulating and empirical findings. *Learning and Instruction. 10*(1), 73 - 99.

Doctoral candidate list of presentations

Ghaboosi, A. (2014, May). Six years of UPRT research papers: A corpus linguistics analysis. Előadás a University of Zagreb Round Table Conference 2014 c. konferencia során. Zágráb.

Ghaboosi, A. (2014, September). Keywords and collocations in Iranian students' written texts: Corpus linguistics analysis. 18th Hungarian Sociolinguistics Conference. Nitra.

Ghaboosi, A. (2014, November). The frequency, distribution and functions of OK in academic speech. Kodolányi János University of Applied Sciences. ANYK 2014. Budapest.

Ghaboosi, A. (2015, November). Speech acts in *Madagascar* 1: A discourse analysis. Kodolányi János University of Applied Sciences. ANYK 2015. Budapest.

Ghaboosi, A & Horváth, J. (2015, May). The Happy Corpus: A diachronic study of university students' written proficiency in EFL. Előadás a University of Pécs Round Table Conference 2015c. konferencia során. Pécs.

Ghaboosi, A. (2016). Testing English language among Iranian EFL learners at English institutes. Kodolányi János University of Applied Sciences. ANYK 2016. Budapest.

Doctoral candidate list of publications

Ghaboosi, A. (2014). The frequency, distribution and functions of OK in academic speech. In T. Gecső & Cs. Sárdi, Cs. (Eds.), *Nyelv és kép*. (pp. 67 - 79). Tinta Könyvkiadó, Budapest.

Ghaboosi, A. (2015). Speech acts in *Madagascar* 1: A discourse analysis. In T. Gecső & Cs. Sárdi, Cs. (Eds.), *Nyelv*. *Kultúra, Társadalom.* (pp. 89 - 97). Tinta Könyvkiadó, Budapest.

Ghaboosi, A & Horváth, J. (2015). The Happy Corpus: A diachronic study of university students' written proficiency in EFL. In M. Lehmann, R. Lugossy & J. Horváth (Eds.), *UPRT 2015: Empirical studies in English applied linguistics* (pp. 141 151). Lingua Franca Csoport, Pécs

Ghaboosi, A. (201x).Testing English language among Iranian EFL learners at English institutes. In T. Gecső & Cs. Sárdi, Cs. (Eds.), *Az ellentét nyelvi és nonverbális kifejezésének lehetőségei.* (pp. xx - xx). Tinta Könyvkiadó, Budapest. (In progress)