GLOKALde April 2016, ISSN 2148-7278, Volume: 2 Number: 2, Article 6
GLOKALde is official e-journal of UDEEEWANA

COMPUTERIZED CORPUS ANALYSIS OF KEY TERMINOLOGY USE IN ENGINEERING DEPARTMENTS IN TURKEY

Asst. Prof. Dr. Ali Sükrü OZBAY
Karadeniz Technical University,
Faculty of Letters, Department of English
Kanuni Campus Trabzon, TURKEY

ABSTRACT

The probable existing correspondence between computerized corpus based applications and the lexical analysis of the English language is far more obvious today, due to the potential of computerized corpus to offer new insights for the language teaching and learning opportunities. The aim of this descriptive study is to find out the most commonly used content specific lexical items in reference books of Engineering departments at KTU (Karadeniz Technical University, Turkey). Yet, another purpose of the study is to find out the existing lexical similarities among these departments (mainly Engineering) in an effort to match these departments according to the degree of lexical similarity. To this end, a descriptive study with eight engineering fields at KTU were done and introductory level reference books from each of these engineering fields were chosen in consultation with the content area instructors from each fields, They are, then, uploaded to AntConc 3.1.0 software for key word analysis. During the analysis procedures, content specific words were extracted for each field and these content specific words were presented in frequency order in the form of tables and graphics.

In the second part of the study, the findings (lexical items) were compared with each other and the fields which include the same or most similar content specific words were grouped together. These groups were, then, analyzed in terms of lexical properties and variations. The results of the lexical analysis showed that there are some correspondences among the various the fields in terms of shared lexical items. The correspondences observed between the above fields in terms of content specific lexical terms indicate that these similarities need to be considered seriously while preparing course content and teaching in the mixed classrooms.

Keywords: **computerized corpora, lexical analysis, terminology, engineering**

INTRODUCTION

It is a well-known fact that the integration of technologic tools, especially the computer, into the linguistics resulted in new discoveries about languages around the world.

These new discoveries were mostly made in the way that the languages are studied and thus the scope of the boundaries for the language studies have been extended far beyond.

One good example to this rising technology in linguistic studies is the emergence of computerized corpus and its widely available applications to the language studies. The advances in computer technology gave rise to the development of computerized corpus linguistics, through which it became largely possible to store, explore and analyze complex patterns of language use, or large language data in qualitative and quantitative terms through concordance component of corpus software.

Thus, it became possible to store and analyze a larger database of natural language. Furthermore, computers provide "consistent, reliable analyses" (Biber, Conrad and Reppen, 1998. p. 4). In fact, one purpose of this present study is to show the potential of computerized corpus based applications to give us qualitative data regarding the lexical usages in a language with different registers. In other words, rather than creating theories of what should be in the language computerized corpus based studies are interested in the patterns and variations in the empirical data and what they show to us about the language behavior.

It is perhaps Mukherjee`s (2005) comparison of computerized corpus with the invention of the microscope that best describes it to a novice person. He says that "how microscope helped to the creation of a new field, which is called microbiology, in the same way computer technology helps corpus linguistics, so it has chance to improve" (p. 86).

Simply defined as the electronic collections of spoken or written texts produced by native speakers in a variety of language settings, computerized corpus can make analysis of simple word count and display with the freely available concordancing software AntConc 3.0.1 that was designed by Lawrence Anthony in the year 2004.

The potential to provide most objective data about language under study is the best thing about a computerized corpus application. A computerized corpus based analysis can explore many aspects of a language, such as lexical, structural, lexico-grammatical, discoursal, phonological, and morphological aspects. It investigates not only these fields, but also such things as male versus female usage of tag questions, children's acquisition of irregular past participles or counterfactual statement error patterns of language learners.

"With the qualitative and quantitative information on the native speakers' use of language to be obtained in this way", researchers can discover both patterns of language use and the extent to which they are used by natives as well as the contextual factors that influence variability (Granger, 2007, p. 235).

Through computerized corpus based tools, it is for instance, possible to analyze present perfect to see how often it is used in various situations or to see the use of synonyms like 'begin' and 'start' or 'big/large/great' for the purpose of finding out their contextual preferences and frequency distribution (Krieger, 2003, p. 2).

Mukherjee (2005) reported that the first and most important effect that computerized corpus linguistics had on language teaching is the study of lexicography. In 1980s, Collins Birmingham University International Language Database (COBUILD) created the first corpus-based dictionary and John Sinclair initiated and organized this study. The corpus-based description of English and the systemic behind it brought about many new dictionaries that are prepared accordingly. These dictionaries have information which is not available in traditional dictionaries.

Computerized corpus linguistics has produced abundant findings in lexis, as reported in McCarthy & Carter (2002), in which "multi-word clusters in spoken British English are used as interactional patterns, showing that teaching should also be less focused on single, decontextualized words and more concerned with chunks of language as well as with how words are organized within specific contexts" (as cited in Mattos de Sa and Nunes, 2012, p. 181).

According to Sinclair's sense of lexical approach, a lexical item has priority in grammar. As Willis (1990) stated this concept is based on "grammaticalized" lexis that gave inspiration to the ELT lexical syllabus designs around frequency lists and their most important recurrent patterns.

He also noted that the most frequently used 700 words were chosen from the corpus for the first level of the COBUILD course, and these words indicate for around 70% of all English text (as cited in Flowerdew, 2013, p. 2).

Purpose of the Study

The idea behind this study is to find out the most commonly used content specific lexical terms in reference books of engineering departments at KTU. Another purpose is to find out the existing lexical similarities among these departments in an effort to match these departments according to the degree of lexical similarity. I think that finding out the lexical correspondences among the engineering fields in the context of the study is important for at least two reasons, these being that the knowledge of at least two fields having similar lexical properties may help students in the engineering departments to cooperate while learning these common content specific words.

Secondly, it is also the case the English teachers who are teaching English for specific purposes in engineering departments may cooperate and create a word list that is comprised in common content specific lexical items.

Rationale for the Study

One of the greatest benefits of corpus is that it gives us a true picture of many high and low frequency lexical items. I think that these lexical items should be considered carefully before any language course is prepared. Failure to do so, however, may create a language course that gives little pedagogical attention to lexical items with a high frequency of occurrence in corpora (Biber, Conrad and Reppen, 1994).

I also believe that awareness towards the shared lexical items among the engineering fields will help language teachers to explore the computerized corpus tools at least in terms of high and low frequency content specific lexical items.

This will, in turn, create a better learning situation particularly for intermediate and upper intermediate level learners who are from engineering departments. This was the main rationale for conducting this qualitative descriptive study, which I will present in the following section.

Research Questions

This study seeks to answer following questions;

- **What are the most commonly used content specific key terms in engineering reference books?**
- **How corpora can be used for understanding the extent of similarities that exist among the departments in terms of terminology and key words?**

METHODOLOGY AND DATA COLLECTION

In this qualitative descriptive study, first of all, engineering departments at KTU were selected and the introductory reference books from each field were chosen by the help of the content teachers from each field. These reference books were text filed and uploaded to AntConc 3.0.1 software program which is freely available on the internet. Content specific words were extracted for each field through the software.

First part of the study is entirely descriptive in that these content specific words were presented in frequency order as tables. In the second part of the study the findings were compared with each other and the fields which include the same or most similar content specific words were grouped. These groups were then analyzed in terms of lexical properties.

AntConc 3.0.1 is a software program developed by Laurence Anthony of Waseda University, Tokyo, Japan (Anthony, 2004). It is free for download from the author's homepage. AntConc is periodically refreshed with recent information. The version used in this paper is AntConc 3.2.1. AntConc does not require any installation on your computer but can be launched by simply double-clicking on the executable file (Römer and Wulff, 2010, p. 104). It provides word frequency (total number of types and tokens) and the proportion of types and tokens.

It can also display concordance lines (a specific word and its surrounding co-text of occurrence) and lists of collocates and clusters, among others (Mattos de Sa and Nunes, 2012, p. 179). According to the Laurence Anthony (2004) "AntConc is a freeware, multi-platform, multi-purpose corpus analysis toolkit, designed specifically for use in the classroom. It hosts a comprehensive set of tools including a powerful concordancer, word and keyword frequency generators, tools for cluster and lexical bundle analysis, and a word distribution plot" (p. 7). Two of the useful features of the Antconc are Word List and Keyword List tools. Composing a word list is important because it helps us to find interesting areas in a corpus and propose problem areas. We can also find lemmas of words or families of related forms in a corpus. The Word List tool of Antconc can sort words into alphabetical or frequency order. Also it has the ability to count words based on their 'stem' form (Anthony 2004, p. 10). Word list tools tell us little about how important a word is in a corpus.

Hence, AntConc, also, offers us Keyword List Tool. Thus, we can find which words frequently appear in our corpus and these words are compared with the same words in a reference corpus. The software calculates the 'keyness' of the words by using 'chi-squared' of 'log likelihood' statistical measures as well as hiding the infrequent words or negative-keywords (Anthony 2004, p. 11).

DESCRIPTIVE DATA ANALYSIS OF MOST COMMONLY USED CONTENT SPECIFIC TERMS

In this section, the below fields from the engineering departments were chosen as the target for the descriptive lexical classifications of the first fifty lexical items from each:

- **Civil Engineering,**
- **Architecture,**
- **Material and Metallurgical Engineering,**
- **Mechanical Engineering,**
- **Geology Engineering,**
- **Mining Engineering,**

Table 1. Material and Metallurgical Engineering

line	antline	freq.	words
1	28	937	temperature
2	32	788	energy
3	35	709	material
4	46	629	properties
5	47	621	polymer
6	53	543	stress
7	55	538	phase
8	62	518	materials
9	65	498	fiber
10	66	496	surface
11	68	485	rate
12	75	467	strength
13	80	447	point
14	81	441	process
15	83	435	magnetic
16	88	402	thermal
17	89	394	structure
18	93	386	glass
19	97	377	high
20	100	371	solid
21	102	360	polymers
22	103	357	liquid
23	107	347	constant
24	110	341	density
25	113	335	matrix
26	115	330	conductivity
27	117	327	pressure
28	118	324	metal

Table 1: Mechanical Engineering

line	antline	freq.	words
1	30	843	stress
2	41	625	system
3	43	593	force
4	45	565	natural
5	46	563	indenter
6	53	510	frequency
7	54	492	mass
8	55	488	contact
9	61	448	load
10	64	430	excitation
11	67	426	frequencies
12	68	418	inertia
13	69	418	surface
14	70	415	energy
15	72	405	crack
16	74	404	moment
17	77	399	motion
18	79	394	spring
19	81	374	specimen
20	83	370	elastic
21	84	369	model
22	86	352	values
23	93	330	time
24	94	329	damping
25	95	329	vibrations
26	98	321	equations
27	100	319	material
28	105	310	matrix

29	122	318	modulus	29	106	310	point
30	128	312	fibers	30	112	295	value
31	130	308	flow	31	114	289	mode
32	132	306	reaction	32	115	288	body
33	133	304	atoms	33	117	282	parameters
34	136	301	metals	34	120	274	plastic
35	138	298	Engineering	35	122	268	function
36	139	298	molecular	36	123	268	rigid
37	140	297	reprinted	37	125	267	calculated
38	141	296	heat	38	127	265	machine
39	142	292	strain	39	129	260	velocity
40	143	292	viscosity	40	130	256	constant
41	150	274	shear	41	131	256	indentation
42	152	270	composite	42	134	251	Oscillators
43	153	270	volume	43	135	248	ratio
44	155	268	mechanical	44	138	243	radius
45	156	265	time	45	140	240	coordinates
46	157	259	weight	46	143	239	calculation
47	160	257	diffusion	47	144	239	moments
48	163	253	cell	48	147	231	resonance
49	165	251	gas	49	149	227	axis
50	167	249	diagram	50	150	227	direction

First fifty content specific key words from the each engineering fields listed below were selected by means of AntConc 3.0.1. The tables were then created showing the lists from each field separately. First part of Table 1. shows the first fifty selected content words from the Mechanical Engineering field and these words were extracted through the AntConc 3.1.0 software program. In Ant line section, the words are given according to their sequence in Antconc program. In the second part, the frequency level of the words or how many times a word is used in the reference books are shown. For example, the very first word in the table is 'stress' and it was used 843 times in the selected reference books, and it comes in the t thirtieth word order in Antconc concordance lines.

The table presents us the most commonly used content words in the related field and the word "stress" come in first in rank order.

If we look into the first fifty content specific words it is possible to conclude that these words are also used in another related discipline namely, Material and Metallurgical Engineering.

Second part of the table consists of key words selected from the Material and Metallurgical Engineering reference books.

The word 'temperature" has the highest frequency order and it is used 937 times in reference books and it means the degree of hotness or coldness of a body or environment (corresponding to its molecular activity)

In Table 2, content specific key words from the Architecture field were analyzed. The first word 'data' is used 1655 times and it is in the twenty-eighth line in Antconc.

The second word 'buildings' is used 1470 times in Architecture reference books. Another frequently used word is 'function' and it is used 1349. Function means a math ematical relation such that each element of one set is associated with at least one element of another set.

In Table 2. Civil Engineering reference books were used. First fifty words were chosen as the most frequently used content words in the reference books.

The very first word in the table is 'rock' and it is used 1286 times in reference books, it is eighteenth word in Antconc line.

According to table, it can be said that this word has an important place in Civil Engineering.

In Table 3, the first word is 'figure' from Geology Engineering appears and it was used 3576 times in reference books.

It means a diagram or picture illustrating textual material. The second word 'rock' is used 2164 times in introductory reference books.

It naturally indicates the proficiency of the engineering.

Third word is 'heat' and this word is used 2159 times. Another frequently used word is water and it is used 1945 times.

In Table 3, the word 'rock' from Mining Engineering was used 2400 times in reference sources.

Table 2. Architecture

Line	Antline	freq.	words
1	28	1655	Data
2	31	1470	buildings
3	36	1349	function
4	41	1199	Code
5	44	1142	process
6	49	1057	Cisco
7	54	944	architecture
8	55	936	system
9	56	926	memory
10	58	905	application
11	67	854	Time
12	73	812	Set
13	76	770	Label
14	80	719	List
15	83	706	struct
16	85	681	version
17	86	675	File
18	88	666	structure

Table 2. Civil Engineering

Line	Antline	freq.	Words
1	18	1286	rock
2	23	976	ground
3	27	873	flow
4	29	795	tunnel
5	31	781	Design
6	34	733	Section
7	35	708	water
8	37	680	diameter
9	41	657	Hydraulic
10	48	562	Equation
11	49	557	Manual
12	50	556	depth
13	54	516	strength
14	55	514	section
15	59	490	construction
16	60	478	conditions
17	61	478	support
18	62	465	mass

19	89	655	client	19	65	454	area
20	93	627	design	20	66	451	strength
21	96	621	void	21	67	450	pressure
22	98	607	device	22	68	442	time
23	106	579	strength	23	75	398	system
24	109	566	information	24	79	378	surface
25	113	562	service	25	84	360	slope
26	118	550	interface	26	85	352	velocity
27	125	514	guide	27	86	350	concrete
28	127	505	form	28	88	349	channel
29	130	476	create	29	89	342	culvert
30	131	466	value	30	90	342	material
31	132	461	functions	31	92	339	hydraulic
32	134	459	size	32	93	338	earthquake
33	138	441	Guide	33	95	337	flood
34	139	437	applications	34	98	334	cross
35	140	436	architectural	35	100	327	Engineering
36	142	432	queue	36	102	321	soil
37	143	428	object	37	104	314	discharge
38	153	407	access	38	107	308	data
39	154	406	platform	39	108	307	face
40	155	405	pool	40	115	294	stress
41	157	393	copy	41	116	294	value
42	161	385	network	42	117	293	method
43	166	367	routine	43	118	293	seismic
44	171	353	buffer	44	120	289	lining
45	174	350	port	45	121	289	site
46	175	350	style	46	128	276	length
47	177	349	index	47	129	276	stream
48	178	347	class	48	130	271	analysis
49	179	347	walls	49	134	267	inlet
50	182	345	way	50	135	266	shear

Table 3. Geology Engineering

Table 3. Mining Engineering

Line	Antline	freq.	words	line	Antline	fre.	words
1	24	3576	figure	1	11	2400	rock
2	37	2164	rock	2	17	1681	slope
3	39	2159	heat	3	19	1488	web
4	45	1945	water	4	23	1359	mining
5	50	1749	surface	5	25	1135	analysis
6	52	1732	stress	6	30	1065	figure
7	53	1725	strain	7	31	1048	failure
8	54	1652	rocks	8	40	822	data
9	60	1475	plane	9	41	779	safety
10	61	1448	line	10	42	773	equipment
11	75	1271	crust	11	44	745	water
12	77	1243	deformation	12	48	696	design
13	79	1237	shear	13	49	686	strength
14	80	1218	earth	14	50	685	surface

15	81	1215	faults	15	56	591	stability
16	82	1203	fold	16	58	582	time
17	83	1198	temperature	17	61	571	face
18	85	1175	time	18	62	557	reliability
19	88	1165	plate	19	67	545	shear
20	89	1081	slip	20	71	524	plane
21	90	1080	folds	21	72	515	factor
22	91	1077	point	22	77	499	model
23	92	1056	direction	23	78	494	system
24	93	1047	zone	24	80	488	angle
25	95	1043	angle	25	84	463	number
26	103	969	flow	26	86	444	ground
27	106	944	ground	27	88	439	slopes
28	109	925	continental	28	91	428	conditions
29	112	904	dip	29	93	416	set
30	114	886	data	30	96	409	information
31	118	853	strike	31	97	402	mass
32	121	834	thrust	32	98	397	maintenance
33	122	814	parallel	33	99	388	dip
34	124	796	chapter	34	101	386	discontinuities
35	129	770	pressure	35	103	362	sliding
36	134	735	model	36	107	344	pit
37	135	723	section	37	112	336	friction
38	141	685	km	38	114	334	methods
39	144	673	normal	39	117	330	section
40	147	662	points	40	122	319	stress
41	149	657	axis	41	123	318	engineering
42	150	656	north	42	124	315	probability
43	151	653	circle	43	125	313	block
44	153	648	depth	44	126	313	wedge
45	154	646	system	45	129	304	maintainability
46	155	643	orientation	46	130	304	values
47	158	634	mantle	47	131	303	cost
48	159	632	structural	48	134	301	toppling
49	162	609	energy	49	136	298	value
50	163	606	structures	50	137	297	Hole

COMPARATIVE DATA ANALYSIS OF MOST COMMONLY USED CONTENT SPECIFIC TERMS

In this section, the fields which include similar content specific words were classified and presented together in the tables below.

The words were chosen according to their rank order or frequency in the concordance lines in AntConc program.

In Table 4, words are shown according to their frequency level in reference books of each department.

As it seen 'energy", "material", "stress"', and "strain" are the most common three words in Material Engineering vs. Mechanical Engineering departments.

Table 4. Material Engineering vs. Mechanical Engineering

		Material	Mechanical	Material	Mechanical
line	words	Antline	Antline	freq.	freq.
1	energy	32	70	788	415
2	material	35	100	709	319
3	stress	53	30	543	843
4	elastic	243	83	178	370
5	matrix	113	105	335	310
6	constant	107	130	347	256
7	ratio	224	135	194	248
8	surface	66	69	496	418
9	pressure	117	188	327	182
10	strain	142	168	292	199

Architecture and Civil Engineering reference books were analyzed and findings were compared to obtain the following Table 5. "Structure", "design" and "building" are the most commonly used words both in Architecture and Civil Engineering. In the rest of the table it is seen that there are many other correspondences between the two fields in terms of lexical usages.

These lexical similarities indicate that these fields can be treated together while language content is being prepared for the learners.

This will help language teachers to cooperate with other content teachers as well in terms of creating word lists to be taught.

Table. 5 Architecture vs. Civil Engineering

		Architecture	Civil en.	Architecture	Civil en.
Line	words	Antline	Antline	freq.	freq.
1	structure	88	199	666	198
2	design	93	31	627	781
3	building	31	163	1470	237
4	construction	230	59	297	490
5	wall	179	185	347	216
6	time	67	68	854	442
7	strenght	106	54	579	516
8	size	134	198	459	198
9	control	193	146	329	253
10	material	185	90	339	342

In Table 6. common content words from both fields were given. The first word 'rock' is the most commonly used word in both departments.

It is used 2164 times in Geology Engineering and 2400 times in Mining Engineering.

The second one is 'water' and this word is used 1945 times in Geology, and it is used 745 times in Mining Engineering

Table 6. Geology Engineering vs. Mining Engineering

		Geology	Mining Eng.	Geology	Mining Eng.
line	words	Antline	Antline	freq.	freq.
1	rock	37	11	2164	2400
2	water	45	44	1945	745
3	ground	106	86	944	444
4	surface	50	50	1749	685
5	angle	95	80	1043	488
6	flow	103	221	969	193
7	shear	79	67	1237	545
8	time	85	58	1175	582
9	stress	52	122	1749	319
10	direction	92	173	1056	233

CONCLUSION

In this study descriptive research methodology was used. Content specific key terms were sought with the help of AntConc 3.0.1.
This study aimed at finding answers to these questions;

- **What are the most commonly used content specific key terms in engineering reference books?**
- **How corpora can be used for understanding the extent of similarities that exist among the departments in terms of terminology and key words?**

The first question was asking the most commonly used content specific key terms in engineering reference books. The first fifty terms from each engineering field were collected and tabled in data analysis section.

When the tables were analyzed it was seen that there were some lexical correspondences among the fields in terms of their use of content specific terms. These fields were given as follows:

- **Material Engineering vs. Mechanical Engineering**
- **Architecture vs. Civil Engineering**
- **Geology Engineering vs. Mining Engineering**

The possible correspondences observed between the above fields in terms of content specific lexical terms indicate that these similarities need to be considered seriously while preparing course content and teaching in the mixed classrooms. Once the shared lexical items are revealed through such computerized corpus based applications, then it may be possible for language teachers to group their students according to their departments or fields and thus the quality and content of teaching and learning may be much more increased.

The second question was asking how corpora can be used for understanding the extent of similarities that exist among the departments in terms of terminology and key words.

The answer is that through the concordance software program introduced before, the most popular content specific key words which are commonly shared in engineering disciplines can be found by means of concordances.

And, thus, the researchers who will be interested in making further discoveries of lexical complexities among the related disciplines in this respect can definitely use computerized corpus based tools in order to obtain reliable and verifiable findings.

BIODATA and CONTACT ADDESSES of the AUTHOR

Ali Şükrü ÖZBAY is an Assistant Professor in the Department of English Language and Literature at Karadeniz Technical University, Turkey. He holds his BA in English Language and Literature from Ankara University and MA and Ph.D. from KTU, Applied Linguistics Program. His main areas of interest include Corpus Linguistics and Learner Corpora, translation studies, listening comprehension strategies, academic writing and error analysis. He has presented papers at several international conferences. He Ph.D. thesis (2015) is titled as "Corpus analysis of the support verb construction development of the advanced learners of English in a tertiary level EFL setting in Turkey".

Asst. Prof. Dr. Ali Sükrü OZBAY
KTU Faculty of Letters
Department of Western Languages and Literature
English Language and Literature
Kanuni Campus Trabzon, TURKEY
OfficePhone(s): +904623772493 +904623773781
Fax: +904623253248
Mobile: +905353093010
Email(s): ozbay@ktu.edu.tr alisukruozbay@gmail.com
URL'S: https://twitter.com/alisukruozbay
https://www.facebook.com/alisukruozbay

REFERENCES

Anthony, L. (2004). AntConc (Version 3.0.1) [Software]. Retrieved July 10, 2008, from http:// www.antlab.sci.waseda.ac.jp/antconc_index.html

Anthony, L. (2004) *AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit*. Center for English Language Education in Science and Engineering School of Science and Engineering, Waseda University, Japan.

Biber, D., S. Conrad & R. Reppen (1998): *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press

Flowerdew, L. (2013) *Corpus-based Research and Pedagogy in EAP: From Lexis to Genre*. Language Teaching, Available on CJO 2013.

"Granger, S., Olivier K., Claude,P., Georges A., and Virginie Z., (2007). Integrating learner corpora and natural language processing : A crucial step towards reconciling technological sophistication and pedagogical effectiveness*. ReCALL*, 19, pp.34

Krieger, D. 2003. *Corpus Linguistics: What It Is and How It Can Be Applied to Teaching* Siebold University of Nagasaki, Japan.

Mattos de Sa, Elisa, Nunes, Leonardo Pereira, (2012) *Language and learner corpora in ELT: reviewing the impact of Corpus Linguistics in the teaching of English,* BELT Journal. Porto Alegre. p. 176-187.

Mukherjee, Joybrato. 2005. *English ditransitive verbs: Aspects of theory, description and a usage-based model.* Amsterdam and New York: Rodopi.

Römer, U. & Wulff, S. (2010). *Applying corpus methods to writing research: Explorations of MICUSP*. Journal of Writing Research, 2(2), 99-127.