The Effectiveness of Corpus- Based Approach to Language Description in Creating Corpus-Based Exercises to Teach Writing Personal Statements

Norah Dhawi Almutairi[1]

[1] English Language Institute (ELI) – King Abdulaziz University (KAU), Jeddah, Saudi Arabia

Correspondence: Norah Dhawi Almutairi, English Language Institute (ELI) – King Abdulaziz University (KAU), Jeddah, Saudi Arabia. E-mail: ndalmutairi@kau.edu.sa

Abstract

Using corpora in language teaching has revolutionized language research with its 'authentic' appeal. Corpus tools have enabled linguistic researchers and teachers to investigate actual usages and the characteristics of certain genres in order to improve syllabus design and infer more effective classroom exercises. From this perspective, this paper attempts to use corpus tools to investigate the characteristics of one of the most important requirements of university programs admissions which is the *personal statement*. Despite the immense importance of writing a personal statement in the lives of students wanting to enroll in universities, little research has been conducted on its instructions. More importantly, teaching its features to university students has been neglected although personal statements are an essential genre that should be emphasized in academic writing classes or university preparation courses. The paper aims to investigate if compiling a corpus of personal statements can lead to creating an effective corpus-based activities to be taught in teaching writing a personal statement. Then the paper attempts to evaluate the pedagogical implications of using corpus-based activities and criticized the weaknesses and strengths of corpora as a resource in language teaching. This paper chose to focus on personal statements collected from law students due to the high demand on law colleges in Saudi Arabia and the difficulty of admission requirements. This study used Sketch Engine® to complie a corpus of sixty-seven personal statement with a total word count of 50, 691, then analysed the lexio-grammatical features. The results were used to create corpus-based excersises to be taught in writing courses teaching personal statements.

Keywords: corpus-based activities, personal statements, designing corpus, teaching writing, teaching and language corpora

1. Introduction

1.1 Introduce the Problem

Using corpora in language teaching is a recent phenomenon that only began in the late 1980s and mostly focused on English language. However, it grew immensely and revolutionized language research with its 'authentic' appeal. That is, corpus-based approach to language teaching is based on actual usage, real and authentic occurrences of language as it is uttered, written and used by native speakers in various situations. Therefore, corpus linguistics, being a computer-based tool, is centered on providing "a ready resource of natural, or authentic, texts for language learning" (Reppen, 2010, p. 4).

Corpus tools enable linguistic researchers and teachers to investigate actual usages or the characteristics of certain genres in order to improve syllabus design and infer more effective classroom exercises. Corpora can hugely impact language teaching in various fields because of its ability to present statistically proven evidence of the used language. This resulted in many incorporations of corpora in language teaching fields (Johansson, 2009; Leńko-Szymańska, 2014).

From this perspective, this paper will attempt to use corpus tools to investigate the characteristics of personal statements. A personal statement is considered one of the most important requirements of university programs admissions. However, despite the immense importance of writing a personal statement in the lives of students wanting to enroll in universities, little research has been conducted on personal statement's instructions. More importantly, teaching its features to university students has been neglected although personal statements are an

essential genre that should be emphasized in academic writing classes or university preparation courses.

This paper will compile a corpus of personal statements and provide a description of compilation procedures. Then, use the constructed corpus to investigate the most frequent lexico-grammatical features that can be introduced to language learners. Moreover, this paper will create corpus-based exercises that can help facilities the instructions of personal statements in EFL classrooms. Finally, this paper will attempt to evaluate and criticized the weaknesses and strengths of corpora as a resource in language teaching.

Personal statements, otherwise known as application letters for graduate schools, can be classified as an academic, promotional genre. A personal statement serves to win the student's admission to the desired university program or course, hence, it needs to have certain characteristics to convince the admission team to accept the applicant's proposal to join the program. However, I have noticed that in many academic writing courses, especially in Saudi Arabia, there is a lack of knowledge in the features of such genre. For example, personal statements serve to promote the applicant and persuade the audience (admission team) to accept her/him, thus, personal statements' structure allow narrating stories, listing qualifications, and encourage creativity and individuality.

This paper chose to focus on personal statements collected from law students due to the high demand on law colleges in Saudi Arabia and the difficulty of admission requirements. I wish to be able to investigate the characteristics of the personal statements of law applicants to teach them to my foundation-year university students aiming for law courses.

2. Literature Review

Reppen (2010) defined the corpus as "a large and principled collection of naturally occurring texts (written or spoken) stored electronically" (p. 2). Analyzing this definition, two major characteristics of the corpus can be inferred. First, corpus collections need to be principled, i.e. shaped and directed by the researcher's goal of designing the corpus. For example, if a researcher or a teacher wanted to design a corpus of written language, then such corpus needs to be representative of the researcher's goal (written language) and contain a variety of written language situations. Second, corpus needs to consist of naturally occurring texts as in actual use of language in real situations such as letters, students' assignments, and books.

The use of corpus tools has immensely impacted linguistic research and second language (L2) learning and teaching. In the 1980s, the growth of corpora and corpus evidence have resulted in creating numerous corpus-based reference publications such as dictionaries and empirical grammar research. According to Partington (1980), language researchers and teachers started to compile mini corpora for specific purposes. Furthermore, Partington (1980) argued that these specially designed corpora are extremely relevant to language research. (p. 4). After that, many suggestions emerged about the creative use of corpora in language classrooms such as the creation of exercises that are directly extracted and driven from corpus tools. According to Römer (2011), the implementation of corpus tools and methods in L2 teaching can be classified to direct and indirect application. Indirect application of corpora means the instances in which corpora can provide information on "what to teach and how to teach it" (Römer, 2011, p. 206). Thus, affecting syllabus design and the improvement of teaching materials. On the other hand, direct application or data-driven learning (DDL) refers to introducing the learners to corpus tools where they can attempt the role of researchers by discovering and inferring meanings and grammatical rules.

Consequently, for genres that are infrequently investigated but highly influential such as personal statements, corpus tools proved to be useful. For example, Jones (2013) used corpus tools to analyze how personal statements differ according to applicants' educational background and recommended this genre to be given high attention as personal statements are being used as a way to select potential applicants from other "similarly qualified peers "(Jones, 2013, p. 401). Similarly, (Chiu, 2015) in his study of personal statements in PHD students recommended that the genre receive further research and attention in writing courses. He criticized the current teachings of this genre and noted "that certain features must be shared between the Personal Statement genre for doctoral study admissions and academic writing within the targeted academic setting" (Chiu, 2015, p. 72). Meaning that, writing courses are advised to study these shared features and implement them in their teachings.

From this perspective and based on the effective contribution of corpus into personal statements related researched as mentioned above. This study was conducted on this genre using personally complied corpus and implementing corpus tools.

3. Methodology

3.1 Data Collection

Sixty-seven law school personal statements have been collected from various public online websites, mostly from (www.law-schools.com/forums) where various law students post their personal statements for approval, criticism, and advice (see list of websites used for compiling this corpus in Appendix A). Thus, some of these personal statements are successful letters that have enabled their owners to secure a place in the desired law schools, some are deemed lacking, and others were posted by the forum admins as effective samples or edited ones. Additionally, all the personal statements were written to apply to law-colleges and mostly written by non-native applicants wanting to join law-schools in the United Kingdom or United States, while very few were written by native speakers in the mentioned countries.

Sketch Engine® was chosen as the corpus program for this study. It is a prominent corpus tool that enabled many around the world to study lexicography. Sketch Engine® provides researchers with important tools to study concordances, collocation, word list, key words, and word sketches among other tools. Moreover, it enables the users to compile their own corpus as well as gives free access to existing corpora. These features were needed to conduct this current study. Therefore, to compile my corpus, I uploaded a text file format that contained sixty-seven personal statements by sixty-seven students into the Sketch Engine®. The total number of words in my corpus are 50,691 words which mean that my corpus contained a large and representative sample. The corpus information is shown in table 1 below.

Table 1. Corpus information

Name of corpus	Number of Personal statements	Total words	Tokens
Personal statements	67	50, 691	57, 019

4. Anaysis

As shown in table 2 below, the most frequent word in the whole corpus is the article 'the' with (2231) frequency, followed by the pronoun 'I' (2095), the preposition 'to' (1729) and the conjunction 'and' (1565). Although the articles (the, a, an), the prepositions (to, of, in), and the pronoun (I) are usually the most frequent words in most corpora, the pronoun I, in particular, carries a special significance in this corpus. That is, since this corpus is based on personal statements, which are self-promotional essays, the excessive occurrences of the pronoun I with frequency of (2095) is justified and even required. On the other hand, the verb 'was' with (616) frequency is the first verb to appear in the list indicating the dominating usage of past tense in this type of genre which depends on the writers' listing and narrating their qualifications and education. Further analysis of the most frequent words will be shown in the following sections.

Table 2. Most frequent words in personal statement corpus

Most frequent word	frequency	Most frequent word	frequency
The	2231	law	225
I	2095	school	174
To	1729	life	133
And	1565	work	101
My	1205	family	96
Was	616	legal	84
Me	442	goal	26

Another category that is worth investigating is the parts of speech through the 'word tags' feature in Sketch Engine. Table 3 below shows the most frequent tags in personal statement corpus:

Table 3. Most frequent tags in the personal statement corpus

Tags	Description	frequency	Tags	Description	Frequency
NN	Noun	8,132	VV	verb, base form	1,780
IN	preposition, subordinating conjunction	6,590	PP$	possessive pronoun	1,780
DT	Determiner	4,739	VVD	verb, past tense	1,749
JJ	Adjective	3,940	TO	infinitive 'to'	1,183
PP	personal pronoun	3,567	VVN	verb, past participle	1,125
NNS	noun plural	2,664	VBD	verb be, past tense	745

As shown in Table 3, the most frequent part of speech is nouns, followed by prepositions, then determiners and adjectives which is an expected result from essays that focus on narrating and have a self-promoting style. Thorough analysis will begin in the following sections by first focusing on pronouns, nouns and adjectives analyzing the frequency, collocates and the common lexical patterns. Then, verbs and grammatical patterns will be analyzed similarly

4.1 Pronouns and Nouns

Further analysis showed that the most frequent pronouns are I with (2231) frequency, my (1205), me (442), and myself (95). Interestingly, all four pronouns are about the applicant's self which is compatible with the self-promotion genre of personal statements. An applicant wanting to join a law school will write an essay showing off and selling himself in order to win the admissions team's approval. See Table 4 below:

Table 4. Most frequent pronouns in the personal statement corpus.

Most frequent pronoun	frequency	Most frequent pronoun	frequency
I	2095	me	442
My	1205	myself	95

Regarding nouns, the most frequent node is law occurring (225) times, followed by life (133) and family (96). It was expected for the noun law to have such high frequency since the corpus is compiled from law school applicants. After examining the concordances, it is noticed that it strongly collocated with school (65) times and is followed by the word (study) 13 times. As seen in in Figure 1 below.
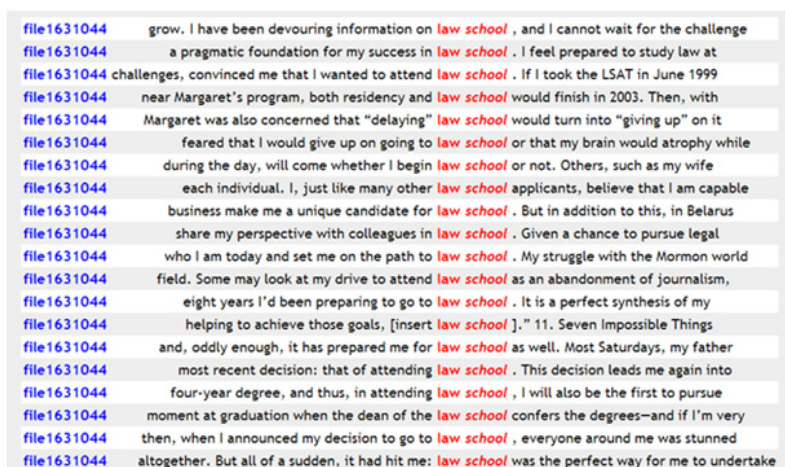

Figure 1. Law school appearing in concordances lines, taken from Sketch Engine

Additionally, Life and family are common nouns in this corpus due to the applicant's need to impress the

admission team in the desired law school by narrating about life's accomplishments, obstacles and family's influence. The same result can be inferred from the frequent noun 'experience' (56) which collocated with work and rewarding, e.g. my diverse work experience, rewarding experience. Another highly frequent nouns are time references such as years, months, hours, and days which is compatible with the applicants' purpose of narrating events that helped shaping their characters or time sequences of obtaining experiences and qualifications.

4.2 Adjectives

While examining the most frequent adjectives, it is expected that the adjective legal is very frequent since all personal statements used in this corpus are from law school applicants. Legal commonly collocated with legal education, legal system, and legal career.

However, the adjective 'first' is the most frequent with (86) occurrences in the corpus. It frequently collocated with the article the and the nodes member and times, hence creating lexical patterns such as the first time and the first member. These examples serve to narrate and express the sequence of events in the applicant's essay because personal statements often contain the applicants' experiences and small extracts of their lives' stories.

Another interesting adjective is personal which strongly collocated with growth creating an appealing quality of the applicant. Similar result is found with the adjective unique which collocated with perspective, and positive which collocated with impact. See frequency list in Figure 2 below.
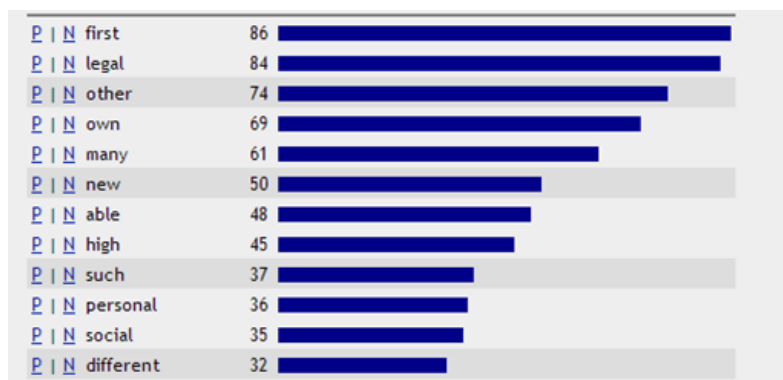
Figure 2. The most frequent adjectives arranged by frequency, taken from Sketch Engine

4.3 Verbs and Grammatical Patterns

The most common verb in the whole corpus is was with (616) occurrences which is due to was being a frequent helping verb in many grammatical patterns such as passive voice, simple past and past continuous tense. The grammatical pattern (I+was+ past participle) which formed a passive voice was highly frequent. For example, I was given the opportunity, I was not allowed, I was voted, and I was assigned. Another grammatical pattern that involved was is (I+ was+ adjective) such as I was honored, I was excited, and I was proud.

One of the most frequent lexical verbs is 'work' occurring (101) times. More importantly, it is observed by closely studying the concordances lines that work is often preceded by to (to work) with (24) frequency, e.g. to work in the legal field. Similar findings are noticed with the verb learn which highly collocated with to and with about as in the following example: to learn about the process. Another frequent lexical verb is 'make' (59) which collocated with decision, e.g. make a decision and with me, e.g. make me a unique candidate. Other examples of frequent lexical verbs are pursuing (21), succeed (14) and achieve (11).

Another interesting grammatical pattern that was observed by studying concordances is (I+will+ verb) which convey the future determination and dedication of the applicants to pursue and fulfill the academic requirements of the chosen field. For example, I will approach the study, and I will reach the top. See Figure 3 below for more examples:

Figure 3. I will appearing in concordances lines, taken from Sketch Engine

5. Pedagogical Implications

This study stresses the importance of teaching personal statements in academic courses especially to non-native speakers of English. The lexical findings, frequency and tag lists showed that personal pronouns, nouns, and adjectives are very important. That is, it showed that personal statements are types of essays where writers narrate their own personal stories that are relevant to academic qualification, education, and career experiences. Using the personal pronouns in this genre of academic writing is strongly favorable because it serves the self-promotion purpose. Moreover, writers, in their attempt to promote themselves, positively describe their personality and vision. For example, when talking about themselves, writers use words such as personal growth, unique perspective, and positive impact to create a favorable appeal to the university's admission team. On the other hand, grammatical features showed the dominance of past tense when listing academic qualifications. In addition, the passive voice is used commonly because the focus should be in the student and not on whoever the subject is, for example, *I was voted*, and *I was assigned*.

These findings which were discovered using the concordance lines can help teachers design various classroom activities to equip students with the appropriate academic vocabulary and grammatical features that suit personal statement genre. The goal of these classroom activities is to encourage students to present themselves, their achievements, goals, and qualification positively and favorably in an academic language to sell themselves and win a place in the desired school.

I created few classroom activates using concordance lines to demonstrate using the corpus findings in classroom exercises. See Appendix B, C and D for examples of these classroom activities.

6. Conclusion: Benefits and limitations of corpora in language teaching

Corpus-based approach to language teaching has proved its benefits in various research. O'Keeffe, McCarthy and Carter (2007) proved that language taught on textbooks is frequently based on native speakers' intuition about how they use the language rather than the actual evidence of usage (p. 21). On the contrary, corpus can present statistically proven evidence of the language actually used which explains the new approaches relying on corpora in language teaching field. For example, according to Johansson (2009), corpora can hugely impact language teaching in the various fields summarized in Figure 4 below.

By showing how corpus can influence the preparation of tests, textbooks, grammar books, dictionaries, classroom activities and in addition to syllabus design. Johansson (2009) is strongly arguing for the effective relevance of corpora in language teaching. Furthermore, frequency data combined with lines of concordances that expose the verbal environment allow great opportunities for linguistic research. Corpus tools have contributed in discovering the behavior of various lexical and grammatical features. Thus, immensely influencing language teaching pedagogy.
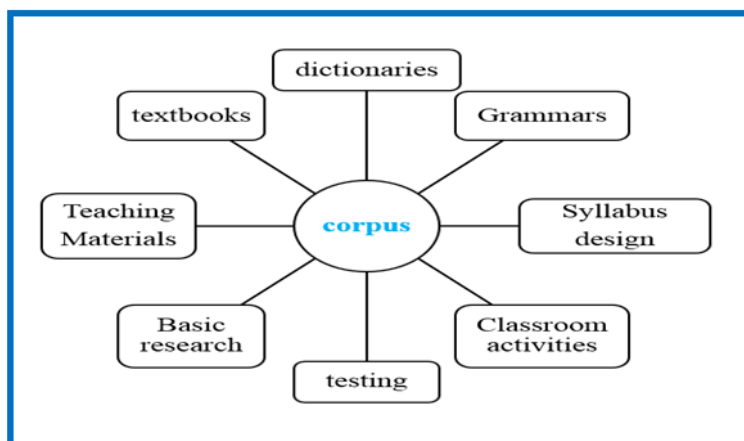
Figure 4. Uses of corpora of relevance for language teaching. Adapted from Johansson (2009)

However, Corpus-based approaches to language teaching have limitations. For example, Flowerdew (2011) insisted that even the largest corpora contain "less language than the average user will have experienced in their daily life" (p. 328). Moreover, the authenticity of the linguistic content in corpora is challenged when Flowerdew (2011) claimed that the language found in corpora is different from the language that native speakers experience in their daily life. The reason of these claims is due to the corpora being compiled mostly from written language, hence, corpora do not reflect the spoken language correctly. Additionally, using corpora requires technical skills in using computer programs, access to computers and internet, and knowledge of related software that might require subscription fees or license. Such aspects might not be reachable by teachers and many learners, hence, obstructing the spread of corpora's usage in classrooms and research. Also, lower proficiency students might not be able to deal with the overwhelming data provided by corpora.

Nevertheless, it has been proved that Corpus-driven approaches to language teaching and learning can immensely help both teachers and learners to deduce and explore authentic language usages. Corpora's potential to investigate grammatical and lexical patterns are well proven and established. This paper hopes that the technological advancement can popularize corpus by facilitating and motivating the full exploitation of its benefits.

Acknowledgments

References

Chiu, Y.-L. T. (2015). Personal statement in PhD applications: Gatekeepers' evaluative perspectives. *Journal of English for Academic Purposes, 17*, 63-73. http://dx.doi.org/10.1016/j.jeap.2015.02.002

Flowerdew, J. (2011). Corpora in Language Teaching. In M. H. Long, & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 327-351). Chichester: Wiley-Blackwell. http://dx.doi.org/10.1017/S026144481000042X

Johansson, S. (2009). Some thoughts on corpora and second language acquisition. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 33-44). Amsterdam: John Benjamins. http://dx.doi.org/10.1075/scl.33.05joh

Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL, 26* (Special Issue 02), 260-278. http://dx.doi.org/10.1017/s095834401400010x

O'Keeffe, A., Michael McCarthy, & Carter, R. (2007). *From Corpus to Classroom*. Cambridge University Press.

Partington, A. (1980). *Patterns and meanings: using corpora for English language research and teaching*. Amsterdam: John Benjamins. http://dx.doi.org/10.1017/cbo9780511497650

Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge: Cambridge University Press.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31*, 205-225. http://dx.doi.org/10.1017/S0267190511000055

Appendix A

List of websites used for compiling the personal statement corpus:

http://www.top-law-schools.com/statement.html

http://blog.owu.edu/writingcenter/files/2012/07/Law-School-Personal-Statements.pdf

http://www.law-schools.com/forums/viewtopic.php?f=18&t=224237

www.law.utoronto.ca/.../UofT_Law_Personal_Statements.pdf

https://www.washington.edu/uaa/advising/.../personal_statement_law.pdf

http://www.powerscore.com/lsat/law-school-admissions/content/Law-School-Admissions-Essay.pdf.

Appendix B

Classroom activity 1:

• Objective: students will be able to observe academic vocabulary and grammatical features and use them correctly to express self-promotions for their personal statements.

> • Study the verb 'pursuing' in these examples.
> • What does it convey? How is it different from the verb 'continuing'?
> • In your notebook, produce your own sentences using 'pursue' while expressing your goals in enrolling in a law school.
> • Exchange your notes with a partner and comment on hers.

Appendix C

Classroom activity 2:

• Objective: students will be able to use academic vocabulary correctly to express self-promotions for their personal statements.

• All the examples are taken from the concoradances lines in the personal statement corpus

a. Fill in the gaps using the words in the box:

> positive impact- hard work- personal growth- unique perspective- a legal career- law school- - unique perspective- look forward

1. I can be a lawyer who makes a direct and _____ on people's lives.
2. I seek more challenge and_____in a field that calls on my written skills.
3. this experience itself gives me _____ on law
4. I _____ to beginning my legal career at Columbia
5. After my year of tutoring, I began to consider _____
6. I believe that the privilege of attending _____ will best allow me to share my vision of justice.

Use the words in the box to write a personal statement promoting your personal traits and academic achievements to win a place in a law course

Copyrights