

KRISZTIÁN SIMON

How I developed vocabulary tests using corpus-based word lists

Összefoglalás: Szókincsfejlesztés az elsőéves angol szakos hallgatóknak korpusz-alapú szószedet és interaktív tanulókártyás weboldal segítségével

A szókincsfejlesztés a nyelv elsajátításának lényeges kérdése. Az egyik első kérdés, amellyel a tanulók szembekerülnek, az, hogy mekkora legyen a szókincsük. Erre a kérdésre van többféle elméleti és gyakorlati szempontú válasz, figyelembe véve a korpuszkutatás eredményeit a szükséges szókincs terjedelméről és mélységéről. Webb (2008) azt állítja, hogy 2000–5000 szócsalád ismerete már elég a szövegek megértéséhez. A kellő szókincs elsajátítását segítő legmegfelelőbb szövegszintű lefedéshez számos korpuszt állítottak össze, amelyből Browne, Culligan és Phillips (2013) *New General Service List* szószedetét, valamint a Coxhead's *Academic Word List* (2000) szólistáját választottam osztálytermi megvalósításra, és később kiegészítettem a Nation és Beglar (2007) *Vocabulary size test* néhány részével. A projekt a Pécsi Tudományegyetem Anglisztika Tanszékének Listening and Speaking Skills (hallott szövegértés- és beszéd-készség-fejlesztés) kurzusán valósult meg. Blended tanulási módszerek mellett a szókincsfejlesztés online feladatmegoldással történt a *Quizlet* internetes tanulókártyás platformon. A hallgatók kettő csomag tanulókártyát kaptak hetente, az egyik a fenti szólisták egyes elemeit tartalmazta, a másik kulcsszavakat a hallásértést segítő feladatokból. A hallgatóknak továbbá tanulókártyákat kellett készíteniük a saját prezentációjuk alapján. A szókincsfejlődés méréséhez Nation és Beglar (2007) *Vocabulary size* tesztjét használtam a félév elején és végén. A hallgatók három online szókincstesztet is megoldottak az *Edmodo* oldalon, valamint egy záródolgozatot papíralapon, amely szavakat tartalmazott az egész kurzus során szereplő szókészletből Wesche & Paribakht (1996) *Vocabulary Knowledge Scale* formátumában. Összességében elmondható, hogy a projekt sikeres volt, az utólagos teszt eredményeit tekintve átlagosan mintegy 700 szóval gazdagodott a hallgatók szókincse négy hónap alatt.

Kulcsszavak: e-tanulás, blended tanulási módszerek, egyetemi szintű angol nyelvoktatás

Introduction

Vocabulary learning is a cornerstone of language mastery. Similarly to the productive and receptive skills, it requires training and scaffolding. It is common for students to have been socialized to some level to using vocabulary booklets in their primary and secondary studies which they also carry over to the tertiary level. However, this approach to lexical extension raises a number of problems. First, in terms of eventual vocabulary acquisition, simply keeping or handing out a list of words is archaic from a methodological standpoint. Second, in the current age of technological developments where various Web 2.0 solutions can be easily applied to meet language needs, the vocabulary booklet approach is severely outdated.

The project at hand is a follow-up study to an exploratory blended *Listening and Speaking Skills (LSS)* course conducted in the spring semester of the 2014/15 academic year. The purpose of the pilot study was to determine how well blended learning is suited for in- and out-of-class language skill development, what learning platforms are suitable for blending and which areas require further attention. The *LSSII* course gathered data in the form of a needs-analysis and a student satisfaction questionnaire, at the beginning and the end of the semester respectively. One of the main findings of that study was that students indicated a need for further vocabulary development which was the initiative of the present project.

The paper discusses possibilities of using corpus-based word lists in a first-year English-major blended *LSS* course. The project addressed vocabulary instruction, extension and testing through two websites: the social learning platform *Edmodo* and the interactive flashcard website *Quizlet*. Students were assessed four times during the course by three online and one final paper and pencil vocabulary tests. Overall, the course covered nearly 1,000 lemmas and the results of pre- and post-course testing found an average vocabulary increase of nearly 700 items. This project is part of a six-study approach that aimed to connect students' language development with the developing international trends in e-learning (see Simon 2016).

1. Literature review

1.1 Blended learning in group and individual learning

E-learning presents a number of language skill development opportunities that extend beyond the classroom limitations. Peachey (2013: 69) argues that online solutions can be applied to increase possible learning time. In cases where students meet regularly during a semester, a blended frame is especially well-suited to address learning needs. However, that is not the only way blended learning can support language skill development.

There are two major ways to approach blended learning in practice. It can either represent a combination of traditionally and online delivered lessons as seen in Sharma (2010: 456). The second realization of blended learning can be related to flipped classrooms. As Baepler, Walker, and Driessen (2014: 227) explain, this concept is an approach where teaching is redesigned at the lecture and seminar levels simultaneously. In Baepler et al.'s (2014) case, this meant that students were required to view a number of video lessons prior to the contact sessions. This approach can be translated well to the blended frame as well. This means that face-to-face time can be used for group-based learning while online environments, websites and mobile applications enable individual, self-paced learning.

1.2 Required vocabulary size and depth

Vocabulary acquisition is one of the major challenges language learners face. As Brezina and Gablasova (2015: 1) argue, the challenge for beginners is finding the best starting point. Questions like how big of a vocabulary should a learner have, how much should be taught about words and how can learners pick up vocabulary knowledge the easiest have been focal points of vocabulary research.

Nation (2006: 79) analysed the lexical coverage of different genres such as newspapers, novels, children's movies and unscripted spoken English and concluded that at the ideal coverage, which he marks at 98%, in written texts require 8,000–9,000; whereas in spoken language require 6,000–7,000 word families. Addressing the issue of vocabulary coverage, Schmitt (2008) summarized the findings of a number of research studies that aimed to find the threshold knowledge needed to be able to understand spoken and written discourse. Overall, Schmitt (2008) arrived at the same conclusion as Nation (2006).

Webb (2008: 80) addressed previous word family research in the literature which argued that smaller sizes ranging from 2,000 to 5,000 can also be acceptable for understanding texts. This argument does not contradict Nation's (2006) and Schmitt's (2008: 80) findings as they argued for an ideal coverage of 98% where the numbers they found are understandable. However, Webb (2008: 80) makes a reasonable point: lexical coverage studies had their primary focus on receptive knowledge and not productive.

1.3 Applying corpus linguistics to vocabulary development

Timmis (2015: 9–13) argues that there are two main types of datasets that can be gained from a corpus: quantitative and qualitative. The first category covers the frequency and token counts, as well as the type of collocations and grammatical structures present in the corpus. The second dataset concerns how specific differences emerge in the structure. This application of corpora can be highly beneficial in vocabulary teaching as it builds on the practical nature of corpora being collections of texts.

Larger corpora encompass the language use of different genres, including written and spoken corpora. Thus, it is a logical next step to generate frequency lists that represent the word families or the ideal coverage of language one should be familiar with. Brezina and Gablasova (2015: 1) state that wordlists are suitable for straightforward learner use and learning material creation. In the past, one of most widely used such lists was the *General Service List (GSL)* (West 1953).

Over the last sixty plus years a number of studies have updated West's original list. Two such projects were especially prominent. Brezina and Gablasova's (2015: 2–3) *new-GSL* is centred on the "existence and stability of general vocabulary" and Browne, Culligan and Phillips' (2013a) *New General Service List (NGSL)* aimed to construct a list of crucial high-frequency words to language learning (Brown 2014: 2). Similarly to the *GSL*, the *Academic Vocabulary List (AVL)* also underwent a number of updates. Three prominent approaches can be found in the form of Coxhead's (2000) *Academic World List (AWL)*, Gardner and Davies' (2014: 324) *new Academic Vocabulary List (new AVL)* for English for academic purposes settings and Browne et al.'s (2013b) *New Academic Word List (NAWL)*.

The question after analysing the aforementioned vocabulary lists is which ones to use. After careful consideration two were selected: Browne et al.'s (2013a) *NGSL* and Coxhead's (2000) *AWL* list. The reasons are as follows. The *NGSL* was designed to be used by language learners and is thus perfectly suited for the vocabulary development goals of the project at hand. Although the *AWL* has received criticism over the years, as

Gardner and Davies (2014: 307-311) emphasize, it was able to provide a basic academic terminology scaffolding for the participants.

2. Context

The context of the study was the *Listening and Speaking Skills I (LSSI)* course at the Department of English Studies at the University of Pécs in the fall semester of the 2015/16 academic year. The LSS seminar is part of six compulsory preparatory courses for a *Proficiency Exam* at C1 level. Students take this exam at the end of their first academic year.

The traditionally face-to-face course was turned into a blended one to address students' language needs, which in the present case concerned vocabulary development. Addressing this issue happened through the combination of *Edmodo* and *Quizlet*, a social learning platform and a flashcard website. During the course, the students were actively engaged with both sites. *Edmodo* provided the main hub for the course's online elements and provided the students with practice opportunities and it was also utilized for testing. *Quizlet* was used as an interactive space for vocabulary instruction, practice and flashcard set creation.

3. Research questions

The study aimed to answer two research questions:

- RQ1: How does vocabulary instruction using e-materials contribute to the vocabulary development of the students?
- RQ2: What kind of measurable changes take place in the vocabulary of the students?

4. Participants

There were overall 17 first- and 3 second-year English major participants in the project; 14 were female and 6 were male students. Twelve were from the three-year BA and eight from the five-year English teacher education programmes. The students were between the ages of 18 and 21 with an average of 20.22 years.

5. Data collection instruments

The main goal of the project was to apply corpus-based word lists and e-learning together to develop students' vocabulary. For this reason, the chosen approach was as follows. First, students completed an online needs analysis which highlighted the language areas that students felt need most attention. Second, the online platforms of the course were utilized in an interconnected manner, whereby instruction and practice was the focus of *Quizlet* and *Edmodo* was used for testing. Finally, next to the online tests, students also completed a final in-class vocabulary test.

5.1 Needs-analysis questionnaire

An essential part of the blended course design in the *Listening and Speaking Skills I* course was focusing on skill areas that students indicated to be problematic for them and therefore require further attention. For this reason, an online needs-analysis questionnaire was constructed on the *Qualtrics* website which asked students to assess their various language skills, what they do for their own development and list where they need help.

The questionnaire used four-point Likert-scales and although students indicated that they were most secure with their vocabulary knowledge (mean: 2.41), their overall written answers across the language skills pointed toward a strong need for vocabulary development. This finding is consistent with the results of the needs analysis and the student satisfaction questionnaire in the pilot blended *LSSII* course conducted prior to the present study. Thus, *Edmodo* and *Quizlet* were applied to meet student needs.

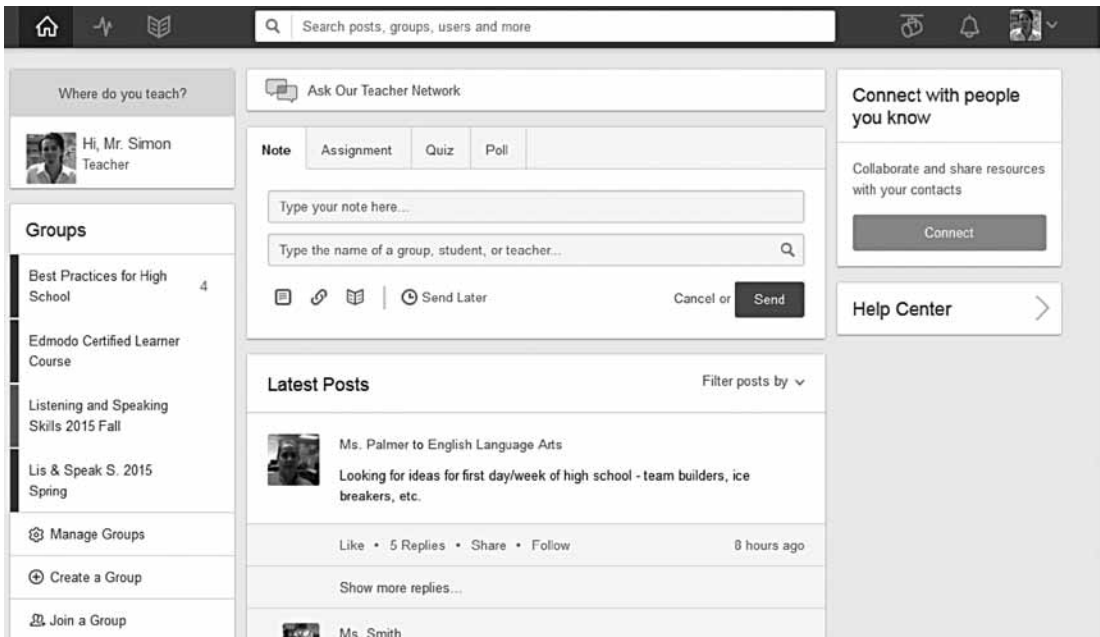
5.2 Online platforms

As stated previously, the *Listening and Speaking Skills I* course was delivered in a blended way. The inclusion of various e-solutions and e-materials was established from two sources. The first concerned the findings of a previous exploratory study using blended learning in a *LSSII* course (Simon–Kollárová 2015). The rationale behind exposing students to blended learning in their second *LSS* seminar was based on learners' prior experience with language skill development courses which would make the inclusion of an online element smoother.

The pilot project (Simon–Kollárová 2015) established that *Edmodo* is suitable for the creation of online learning environments due to its intuitive and easy to use interface and high number of task creation options (see *Figure 1*). The online element complemented

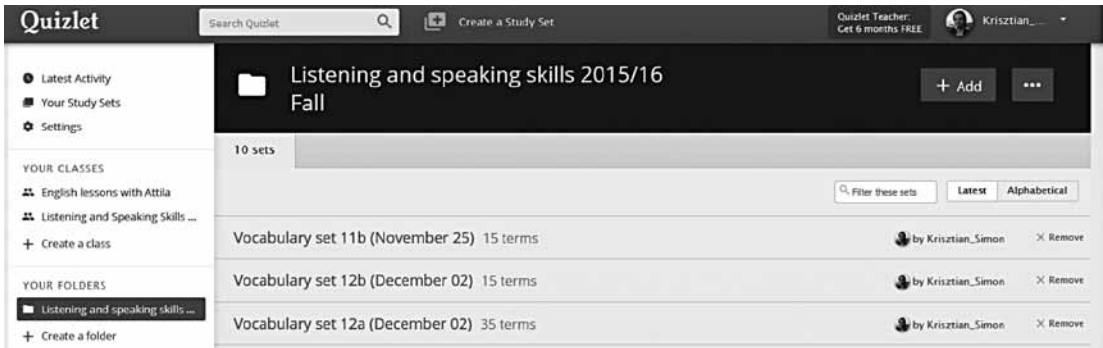
in-class work. Students were presented with additional tasks from which they rated the compulsory and home practice online listening tasks the highest on a four-point Likert scale (3.47 and 3.4, respectively). However, the end-of-term student satisfaction questionnaire also highlighted that vocabulary development needs further attention.

Figure 1: *Edmodo's Facebook-like interface*



Both *Edmodo* and *Quizlet* were selected for similar reasons. Both websites are available free of charge, they apply Web 2.0 frames to present users with intuitive interfaces. In the case of *Edmodo* the design of the interface is deliberately based on *Facebook's* layout. The underlying reason for this can be related to making the learning curve that comes with e-learning platforms as simple as possible. *Quizlet's* interface (see *Figure 2*) follows a straightforward logic, where users can access various folders on the left margin and can access their flashcard sets in the majority of the screen.

Figure 2: Quizlet's streamlined interface



The second reason of using *Edmodo* and *Quizlet* combined for vocabulary development lies in the number of task and test creation options they provide. *Edmodo* makes it possible to present students with timed quizzes. These can give immediate feedback on test completion in the case of pre-set answers. Moreover, these automated items also present tutors with pie-chart statistics and student high scores (see *Figure 3*).

Figure 3: Example for Edmodo's pie chart statistics for the individual test items

Question Breakdown

30 questions | 30:00



Question #1
acquire,felon,...
Matching



Question #2
There was no...
Fill in the blank



Question #3
Major TV cha...
Fill in the blank



Question #4
The oil comp...
Fill in the blank



Question #5
Whenever yo...
Fill in the blank



Question #6
I don't really ...
Fill in the blank



Question #7
Tokyo's ____...
Fill in the blank



Question #8
London, Pari...
Fill in the blank



Question #9
Whenever yo...
Fill in the blank

Overall, Edmodo's task creation tool makes practice, testing and self-paced learning possible, if the tutor prepares all of these activities. Quizlet on the other hand, present learners with a number of automated practice, self-testing and self-paced learning options (see *Table 1*).

Table 1: Description of how the two online platforms were utilized in vocabulary development


	Frame	Task creation	Practice	Testing	Self-paced learning
<i>Edmodo</i>	Web 2.0	tutor-based	tutor-based	tutor-based	tutor-based
<i>Quizlet</i>	Web 2.0	automated	automated	automated	automated

5.3 The vocabulary tests

The three online and one face-to-face vocabulary tests were the main data collection instruments of the study. The former served as formative assessment as they measured students' advances and the latter provided a detailed picture of learners' advancements throughout the semester in the form of a summative test (see Olrich–Harder–Callahan–Trevisan–Brown, 2010: 385–389). Concerning testing, Laufer, Elder, Hill, and Congdon's (2004: 204) points about assessing vocabulary knowledge in contexts as well as in a decontextualized manner served as a guideline for the design of the online vocabulary tests.

Each online test included 50 words from the previously covered items, from which 25 were gap-filling type of sentences and 25 matching tasks (see *Figures 4* and *5*). These tests measured how familiar the students were with the vocabulary items in the first, second and third months of the semester. The online tests were administered on *Edmodo*, as the site gave immediate feedback to students in the form of automated scoring of the items. Test completion was limited to a 30 minute window and students' completion times and scores provided feedback on the test items.

Figure 4: Example for the multiple matching question type in the online vocabulary tests


[Redacted]
38/50
Total Points:

Time Taken: 24:21 | Turned in Nov 18, 2015 @ 5:34 PM
Graded | Delete

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
more ▾

Question Total: 5 points

Match each letter with the correct answer.

<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> A acquire </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> B felon </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> C prohibition </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> D haunted </div> <div style="border: 1px solid #ccc; padding: 5px;"> E regulation </div>	<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> 1 to come to have something Correct Answer </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> 2 a criminal Correct Answer </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> 3 a law that stops something from being used or done Correct Answer </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> 4 infested with ghosts Correct Answer </div> <div style="border: 1px solid #ccc; padding: 5px;"> 5 a law that says something should be done Correct Answer </div>
--	---

Figure 5: Example for the gap filling question type in the online vocabulary tests


[Redacted]
38/50
Total Points:

Time Taken: 24:21 | Turned in Nov 18, 2015 @ 5:34 PM
Graded | Delete

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 more ▾

Question Total: 1 point

There was no wind that day, nothing moved, even the flag was motionless.

Correct Answer

The final in-class test was intended to provide a picture of students' overall lexical development. For this reason, 25 items were selected from the 965 words discussed during the semester and presented in Wesche and Paribakht (1996: 37) *Vocabulary Knowledge Scale (VKS)* format. As Paribakht and Wesche (1997: 187) explain, the *VKS* measures vocabulary knowledge depth on a scale of five distinct points. These range from unknown words to correct usage in sentences. Furthermore, as Wesche and Paribakht (1996: 29) emphasize, the *VKS* is suited to pinpoint the first phases of vocabulary acquisition in the form of self-assessment. The aforementioned categories look the following way in Wesche and Paribakht's (1996: 37) *VKS*:

- I. I don't remember having seen this word before.
- II. I have seen this word before, but I don't know what it means.
- III. I have seen this word before, and I think it means _____. (synonym or translation)
- IV. I know this word. It means _____. (synonym or translation)
- V. I can use this word in a sentence: _____. (If you do this section, please also do section IV).

The *VKS* has five columns that test takers need to complete. These follow a logical format as illustrated above in Wesche and Paribakht (1996: 37). Originally, there is an instruction in the final column (write a sentence) to also complete the fourth one (give a synonym or translation). In the present application of the scale, one slight modification was made to the original test format. The instruction in column V to complete IV as well has been removed to gain an idea whether the participants would choose to answer both to show their understanding of the covered vocabulary (see *Table 2*). As this was a scored test and participants understood that completing both IV and V would yield the most points for each item, students still opted for this solution hence supporting the initial claim of the test designers to include both.

Table 2: *The slightly modified Vocabulary Knowledge Scale marking options used in the final vocabulary test*

I. I don't remember having seen this word before (✓)	II. I have seen this word before, but I don't know what it means (✓)	III. I have seen this word before, and I think it means: (synonym or translation)	IV. I know this word. It means: (synonym or translation)	V. I can use this word in a sentence . (Write a sentence.)
--	---	--	--	---

6. Procedures

The study at hand followed the blending frame established in a prior project (see Simon–Kollárová 2015), where all face-to-face sessions were kept and students were presented with obligatory and optional online tasks. During that project, the task types that could be moved to this online space had been piloted. After analysing the results of the student satisfaction questionnaire in a previous project and a needs-analysis in the present study (see section 5.1), the online element of the course was improved and extended to address vocabulary development needs. As it was a major factor indicated by the students, to achieve this goal, the *Quizlet* website was chosen as a way for interactive vocabulary instruction and practice and *Edmodo* was used for testing.

6.1 Vocabulary item selection

The baseline for vocabulary instruction and development was establishing a number of weekly items that the students could work with without being overwhelmed. From the start this meant three kinds of weekly vocabulary sets. The first set concerned the combined *NGSL* and *AWL* lists. The second flashcard set included items from the weekly listening tasks, which were in the second half of the course complemented by words from Nation and Beglar's (2007) diagnostic test. The final set was chosen from vocabulary lists put together by the students: they covered items students deemed necessary from their own presentations to their peers. As the number of student presentations in the face-to-face sessions ranged from one to two, the minimum number of weekly items was set at 70 and the maximum at 90 words. During the semester, this meant a total of 965 words (see *Table 3*), and an average of 74 words per week.

Table 3: *The distribution and number of the vocabulary items used during the semester*

Selected items from the NGSL	Selected items from the AWL	Selected items from the Nation and Beglar (2007) test	Vocabulary of the listening exercises	Vocabulary of the student presentations
1 st 1,000: 27	216	30	135	380
2 nd 1,000: 86				
3 rd 1,000: 91				
Total number: 965 words				

6.2 Vocabulary instruction

Due to the blended approach, a flashcard software was a logical choice to address lexical development as they would give students freedom and flexibility in learning and practicing. As argued previously, these issues were addressed on the interactive flashcard website *Quizlet*. This website has a built-in feature that pronounces the contents of the flashcards thus students can listen to not only the given lexical item but also to its definition. Furthermore, *Quizlet* also has a number of interactive practice modes. Sharing content and inviting students are also very intuitive on the website. Flashcards on *Quizlet* follow a straightforward principle. One can add the lexical item itself on one side of the card and the description on the other (see *Figure 6*).

Figure 6: *The two sides of a sample flashcard from the first NGSL+AWL set on Quizlet*



In order to provide students with further scaffolding for the vocabulary items, two features were added to the definitions of the words: the word-class and a sample sentence. The *Merriam-Webster* as well as the *Oxford* online dictionaries were used for the definitions. Although the *Corpus of Contemporary American English (COCA)* was chosen for more contemporary words, the bulk of the sample sentences were from *British National Corpus (BNC)*. This structure was applied for the *NGSL+AWL* sets, the vocabulary of the listening exercises plus the Nation and Beglar (2007) items. The students were also asked to follow this method of compilation for their flashcard sets. Although the learners received a list of reliable online dictionaries, they were allowed to use other ones as well as long as they could produce meaningful sets.

6.3 Vocabulary test development

The students were informed from the start of the semester that they would take four vocabulary tests: three online and one final in-class test. These were based on the flashcard sets that covered the *NGSL+AWL*, listening tasks, selected items from Nation

and Beglar's vocabulary levels test (2007) and presentations vocabulary. The online tests followed a pattern where the items were equally selected from the corresponding sets for that month, thus having 25 teacher- and 25 student-set-based words. The distribution of item types in the online tests was the same: 25 gap-filling and 25 multiple matching questions. However, the aim of representing each of the four word-classes found in the lists, nouns, verbs, adjective and adverbs, was not possible, as their distribution was far from equal (see *Table 4*).

Table 4: *The overall number of word-class items in the whole vocabulary set*

Vocabulary set	Number of items			
	Noun	Verb	Adjective	Adverb
NGSL+AWL	188	98	86	19
Listening tasks	77	57	29	1
Items from the Nation and Beglar (2007) list	19	1	9	1
Student presentations	259	60	57	4
Overall ($n=965$)	543	216	181	25

Due to unequal word-class distribution, the test-item selection rationale was to select words that were prominent in the student lists and would represent the gist of their presentations. These items were used for the gap-filling sentences in the online tests. The second criterion was to select words that can be similar in meaning but stand for different notions (e. g. *regulation-prohibition*), which were used in the matching items (see *Figures 4* and *5*). This method was applied, instead of simply presenting the first letter of the given word as it was found to be the closest to Laufer et al.'s (2004) claim about testing vocabulary in context.

6.4 Evaluation of the vocabulary tests

The reason behind moving the tests to *Edmodo* was due to its automated assessing feature in the case of gap-filling and matching item types. In practical terms this means that once the items and their corresponding solutions are prepared, a given access date can be added and *Edmodo* will score the tests and provide feedback to students. Furthermore, a time limit also had to be added to each test. This was 30 minutes for each

50 word online quiz. Aside from counting the correct solutions, a post-test correction protocol was also used during data collection and a pattern of minor mistakes emerged. For this reason, if the only mistake of the student was the omission or incorrect doubling of a letter, the solution was accepted as correct in the post-correction phase. This was the case three times in test 1, four with test 2 and eight with test 3. As these corrections represent overall .5% of all solutions ($n=2,850$) they would not have made a statistical difference either way. However, they do show that the students indeed knew the items and their mistakes were typos or were related to problematic sentences.

In the case of the final in-class test, a different assessment method was implemented as it was a 30-minute paper and pencil test. Here the VKS test format (Wesche–Paribakht 1996: 37) categories were scored the following way:

- I. *I don't remember having seen this word before*: 0 points.
- II. *I have seen this word before, but I don't know what it means*: 0 points.
- III. *I have seen this word before, and I think it means (synonym or translation)*: 1 point if the synonym or translation was correct.
- IV. *I know this word. It means: (synonym or translation)*: 2 points if the synonym, translation or multi-phrase unit either in Hungarian or English was correct. However, no extra points were awarded for presenting both and students were also instructed not to fill III and IV together.
- V. *I can use this word in a sentence*: 4 points. The number of grammatical errors (e.g. wrong use of plural forms) was deducted from the overall score. Also, the synonyms and the sentences (III or IV + V) had to make sense together. However, in cases where the sentence was correct but the synonym was not, or vice versa, points were added depending on the grammatical acceptability of their solution.

The maximum graded score has been set for 100 points (25×4 points for the sentences), but most students completed both III or IV and V together which granted them 5-6 extra points per correct item. Thus, a large number of students were able to gather scores above the set graded maximum. Although this decision can be seen as a judgment error in test design, it was left in the test intentionally as a motivational factor to gather the most possible data about students' vocabulary knowledge.

7. Findings

This section presents the findings of the online and in-class vocabulary tests. It encompasses data from productive and receptive items. First, the three online vocabulary tests are analysed using quantitative tools. Second, the results of the in-class paper and pencil test are presented using quantitative and qualitative data. Finally, students' overall vocabulary development is analysed in the longitudinal project.

7.1 Analysis of the online vocabulary tests (RQ1)

During the analysis of the quantitative data of the online vocabulary tests, a number of patterns emerged. These were grouped according to the successful completion rate of the items covering the 0–20, 21–40, 41–60, 61–80 and 81–100% ranges of correct solutions. Although these were compulsory tests, only the second online test was completed by all 20 participants; in the first only 19 and in the third 18 students did so.

Two issues are immediately visible from Table 5 that presents the results of all 150 words included in the three online tests. First, vocabulary items are represented on various completion levels. The 81–100% group is the most prominent (81), followed by the 61–80% completion rate (25), the 41–60% (19) items, the 21–40% (19) words and finally the 0–20% (6) lemmas. The second issue that is striking in Table 5 is the abundance of matching items (words italicized in the table) in the 81–100% item pool. Some are also present in the 61–80% completion group, however, their numbers are minimal. These findings are in line with Webb's (2008: 80) points concerning how receptive vocabulary knowledge exceeds productive knowledge.

As the items were exclusively presented in online format, it can be stated that the *Quizlet* flashcard sets contributed to the greater vocabulary depth knowledge of the participants. Thus, e-materials have been identified as positively affecting students' vocabulary knowledge (RQ1).

Table 5: *The grouped results of the 150 vocabulary items used in the three online vocabulary tests throughout the semester*

Online vocabulary test #1			Online vocabulary test #2			Online vocabulary test #3		
Items with 81–100 % completion			Items with 81–100 % completion			Items with 81–100 % completion		
<i>admire</i>	<i>distant</i>	<i>offspring</i>	<i>accommodation</i>	<i>contemporary</i>	<i>intricate</i>	<i>acquire</i>	<i>elaborate</i>	<i>pallor</i>
<i>appealing</i>	<i>domestic</i>	<i>qualitative</i>	<i>accomplish</i>	<i>drudgery</i>	<i>irrelevant</i>	<i>assure</i>	<i>examination</i>	<i>population</i>
<i>competent</i>	<i>entire</i>	<i>recruit</i>	<i>anniversary</i>	<i>enchant</i>	<i>questionnaire</i>	<i>atop</i>	<i>extraordinary</i>	<i>prohibition</i>
<i>confuse</i>	<i>evidence</i>	<i>reliable</i>	<i>bilingual</i>	<i>exaggeration</i>	<i>quote</i>	<i>canonical</i>	<i>felon</i>	<i>regulation</i>
<i>content</i>	<i>exchange</i>	<i>reputation</i>	<i>blazing</i>	<i>exhibition</i>	<i>remedy</i>	<i>capital</i>	<i>haunted</i>	<i>settlement</i>
<i>crash</i>	<i>fault</i>	<i>strive</i>	<i>broadcast</i>	<i>herd</i>	<i>restore</i>	<i>clunky</i>	<i>hutch</i>	<i>suggestion</i>
<i>decipher</i>	<i>faulty</i>	<i>sustain</i>	<i>compensation</i>	<i>host</i>	<i>scholarship</i>	<i>constituent</i>	<i>limpid</i>	<i>upbeat</i>
<i>declare</i>	<i>habitat</i>	<i>tongue</i>	<i>competition</i>	<i>illuminate</i>	<i>trophy</i>	<i>cranny</i>	<i>monotonous</i>	<i>wise</i>
<i>decrease</i>	<i>judgment</i>	<i>witness</i>	<i>complaint</i>	<i>influence</i>	<i>weep</i>	<i>egalitarian</i>	<i>mussel</i>	
Items with 61–80% completion			<i>consumerism</i>			Items with 61–80% completion		
<i>browse</i>	<i>impertinent</i>	<i>mishaps</i>	Items with 61–80% completion			<i>attraction</i>	<i>discipline</i>	<i>revise</i>
<i>CEO</i>	<i>independent</i>	<i>praise</i>	<i>ancestor</i>	<i>outskirts</i>	<i>screening</i>	<i>banish</i>	<i>motionless</i>	<i>sniff</i>
<i>genuine</i>	<i>merge</i>	<i>trait</i>	<i>audience</i>	<i>reminiscent</i>	<i>volunteer</i>	<i>crowbar</i>	<i>prehistoric</i>	<i>tribe</i>
Items with 41–60% completion			<i>bankrupt</i>			Items with 41–60% completion		
<i>appropriate</i>	<i>consumer</i>	<i>majority</i>	Items with 41–60% completion			<i>common</i>	<i>fortress</i>	<i>request</i>
<i>attitude</i>	<i>interfere</i>	<i>proceed</i>	<i>depict</i>	<i>landscape</i>	<i>preparation</i>	<i>concern</i>	<i>proof</i>	<i>residence</i>
<i>competent</i>			<i>distinguish</i>	<i>manufacture</i>	<i>soothing</i>	Items with 21–40% completion		
Items with 21–40% completion			Items with 21–40% completion			<i>gap</i>	<i>prominent</i>	<i>scatter</i>
<i>bargain</i>	<i>principle</i>	<i>sequence</i>	<i>attach</i>	<i>likelihood</i>	<i>privilege</i>	<i>initiate</i>	<i>remarkable</i>	<i>spokesman</i>
<i>fascinate</i>	<i>retirement</i>	<i>statement</i>	<i>barely</i>	<i>origin</i>	<i>shore</i>	<i>permanent</i>		
Items with 0–20% completion			Items with 0–20% completion			Items with 0–20% completion		
<i>comprehensive</i>			<i>advertising</i>	<i>scenario</i>	<i>verdict</i>	<i>flagship</i>	<i>grand</i>	

Legend: *italics:* matching items (see Figure 4 for example), regular print: gap-filling items (see Figure 5 for example)

7.2 Analysis of the final in-class vocabulary test results

The goal of the final in-class paper and pencil vocabulary test was to assess students' mastery of the selected items. Overall, 25 words were selected from which 13 appeared at least twice in the whole vocabulary set. These include: *audience*, *bargain*, *common*, *contemporary*, *discipline*, *ensure*, *establish*, *genuine*, *manufacture*, *primarily*, *purpose*, *scholarship* and *volunteer*. The reason for including these is that their second or higher listings could imply that they were seen as scaffolding words in a number of student presentation lists. As already discussed previously, the *Vocabulary Knowledge Scale* test format (Wesche–Paribakht 1996: 37) was chosen to measure depth of knowledge with the 25 items.

Table 6 presents the findings of the final test which was completed by 19 students. The results of the VKS were analysed as follows. While correct synonyms and translations received different scores in the test depending on whether the students put them in category III ('I think I know') or IV ('I know'), here they were rated equally and listed in the corresponding columns based on the word-classes they used. It was often the case that students used a number of synonyms and descriptions, even both in Hungarian and English, thus they were counted individually. This resulted in the total number of these solutions being higher than the overall number of participants. Correct sentences were also categorized based on the word-classes of the items used in them.

Two additional columns were added to Table 6 where all the sentences for the given item, even the incorrect ones, were analysed using the N-gram and keyword tool on the *Compleat Lexical Tutor* website. Only two-string N-grams with the highest number were listed together with the most frequent keywords which were not the target items. In cases where the assessed words appeared, their boxes were left blank.

There are a number of patterns visible from Table 6. *Audience* seems to have been the least problematic item on the test as all participants were able to write a meaningful and grammatically correct sentence with it. *Justification* is on the other end of this continuum. Twenty out of the 25 items show a rather equal distribution of world-classes both in the synonyms/translations and the sentences section. The words where larger mismatches can be found are *leisure*, *neglect*, *quote* and *waste*.

Table 6: Summary of the results of the final in-class vocabulary test in relation to the VKS and corpus data

Item	Does not know/recognize	Correct synonyms and translations (word-classes)				Correct sentences (word-classes)	Most frequent 2 string N-gram	Most frequent keyword in the text after the item (keyword ratio)
		EN synonyms.	EN phrase	HU phrase	HU translation			
1. audience	0 0 1 (n)	4 (n)	0	15 (n)	19 (n)	the audience: 16	perform (0.028)	
2. bargain	0 1 2 (n)	2 (np)	2 (n)	11 (n), 3 (v)	11 (n), 3 (v)	a bargain: 5	market (0.011)	
3. common	0 0 3 (ad)	1 (ad), 1 (vp)	0	20 (ad)	17 (ad)	a common: 7	-	
4. contemporary	1 0 0	1 (vp), 1 (np)	0	12 (ad)	14 (ad)	contemporary authors: 5	author (0.019)	
5. discipline	3 0 0	1 (np)	0	13 (n), 1 (v)	12 (n), 1 (v)	a discipline: 3	-	
6. ensure	0 0 0	3 (vp)	0	13 (v)	15 (v)	ensure you: 5	safe (0.014)	
7. establish	0 0 2 (v)	1 (vp)	0	16 (v)	15 (v)	was established: 9	company (0.016)	
8. former	0 0 2 (av)	1 (avp)	0	12 (av)	13 (av)	the former: 5	-	
9. genuine	0 0 6 (ad)	0	0	15 (ad)	11 (n)	a genuine: 3	gold (0.014)	
10. interior	0 0 0	2 (np)	0	15 (n), 2 (ad)	11 (n), 2 (ad)	the interior: 8	design (0.026)	
11. justification	1 0 1 (n)	0	0	4 (n)	3 (n)	the justification: 3	prison (0.023)	
12. leisure	1 0 1 (n), 2 (ad)	1 (vp)	0	7 (n), 3 (ad)	8 (ad), 4 (n)	leisure time: 6	-	
13. manufacture	0 0 1 (v), 2 (n)	0	0	9 (v), 3 (n)	7 (v), 6 (n)	the manufacture: 5	product (0.02)	
14. neglect	0 2 0	2 (vp)	3 (v)	3 (n), 5 (v) (ad)	10 (v), 1 (n), 1 (ad)	been neglecting: 2	-	
15. obtain	0 2 5 (v)	0	0	6 (v)	10 (v)	to obtain: 5	-	
16. opportunity	0 0 5 (n)	1 (np)	0	15 (n)	18 (n)	opportunity to: 9	travel (0.021)	
17. primarily	0 0 6 (av)	0	0	16 (av)	12 (av)	concentrate on: 3	concentrate (0.019)	
18. purpose	0 0 7 (n)	0	0	14 (n)	17 (n)	purpose of: 6	life (0.012)	
19. quote	0 0 4 (v)	1 (vp)	0	7 (n), 9 (v)	13 (v), 4 (n)	quote from: 3	Shakespeare (0.02)	
20. regulation	0 0 2 (n)	0	0	18 (n)	16 (n)	the regulation: 6	strict (0.03)	
21. scholarship	0 0 1 (np)	0	0	18 (n)	18 (n)	a scholarship: 9	semester (0.015)	
22. tension	0 2 0	0	0	12 (n)	11 (n)	the tension: 6	tense (0.007)	
23. ultimate	0 0 3 (ad)	0	0	12 (ad)	12 (ad)	the ultimate: 9	solution (0.013)	
24. volunteer	0 0 1 (n)	3 (np)	0	15 (n), 1 (v), 2 (ad)	14 (n), 2 (v), 1 (ad)	a volunteer: 8	shelter (0.023)	
25. waste	0 0 1 (n)	0	0	16 (n), 8 (n)	11 (n), 6 (v)	waste your: 4	-	

Legend: n: noun, v: verb, ad: adjective, av: adverb, np: noun phrase, vp: verb phrase, avp: adverbial phrase

7.3 Students' overall vocabulary development (RQ2)

The final part details the students' overall vocabulary profile to determine how the treatment affected their vocabulary knowledge. *Table 7* summarizes the findings of the semester-long project. At the beginning of the semester, students had to complete Nation and Beglar's (2007) diagnostic test and scored an average of 10,000 words. At the end of the semester, the test was repeated which resulted in an average increase of 694.64 words. Unfortunately, it was not completed by all students in either instance so a side-by-side comparison was only possible in ten cases. From this small sample eight participants show vocabulary growth (bold italics) and two decrease (bold).

All three online vocabulary tests show similar standard deviation and mean values, despite the varying number of participants. The SD values of 7.521, 5.369 and 5.583 are quite reasonable and the test scores of 37.368, 37.75 and 37.333 mean that on average each student could complete the tests with a 74% pass grade. The results of the VKS in the final in-class vocabulary test show a somewhat different picture. As already discussed, the graded maximum score was set at 100 points, however, by adding correct synonyms and writing sentences they could achieve 150 points. This explains the SD value of 25.567 due to some tests having a difference of almost 90 points. The average score of 109.157 illustrates that students were able to surpass the set maximum score. However, it also means that the scoring might have to be reconsidered.

Finally, every answer by each student was analysed with Davies' (2012) *Words and Phrase* tool to calculate frequency and vocabulary distribution on the 1-500, 501-3,000, 3,000+ and academic registers. The word counts varied from student to student, however, from the overall of 2,374 about 125 words can be seen as average. The categories were not all equally represented with the 1-500 one covering about 60% of each student's solutions. Finally, the 501-3,000 registers and above categories contained about 15-15% of the words, while the academic register was responsible for the remaining 10%.

In terms of test completion, the above results mean that students were able to successfully complete most of the assessed items on the tests. Therefore, scaffolding their vocabulary development with e-materials was successful. This development is further evidenced by a measurable average increase of almost 700 words by the end of the semester (RQ2) proving that the project reached its goal.

Table 7: The overall vocabulary profile of the students in the Listening and Speaking I course

Student	Vocabulary level at the start of the semester		Online vocabulary test #1		Online vocabulary test #2		Online vocabulary test #3		Final (in-class) vocabulary test	Frequency lists (%) word and phrase 1-500, 501-3000, >3,000, academic		Vocabulary level at the end of the semester
	Score	Time	Score	Time	Score	Time	Score	Time		1-500	501-3000+ ac.	
S1	36/50 (72%)	17:22	36/50 (72%)	21:49	37/50 (74%)	24:48	67	14	130/100	6	12,400 words	
S2	48/50 (96%)	17:52	41/50 (82%)	29:33	41/50 (82%)	30:02	70	13	136/100	9	not completed	
S3	36/50 (72%)	30:25	32/50 (64%)	30:25	35/50 (70%)	30:08	59	18	80/100	11	not completed	
S4	not completed	-	28/50 (56%)	20:30	not completed	-	not completed	-	not completed	-	not completed	
S5	46/50 (92%)	20:27	44/50 (88%)	29:59	41/50 (82%)	30:03	68	13	122/100	6	12,700 words	
S6	30/50 (60%)	12:27	32/50 (64%)	18:27	not completed	-	63	15	96/100	8	not completed	
S7	28/50 (56%)	31:05	40/50 (80%)	30:33	38/50 (76%)	31:09	66	11	62/100	9	11,000 words	
S8	33/50 (66%)	27:46	41/50 (82%)	34:34	36/50 (72%)	24:26	64	14	107/100	7	9,500 words	
S9	26/50 (52%)	19:44	27/50 (54%)	19:28	30/50 (60%)	23:55	61	13	83/100	8	9,200 words	
S10	23/50 (46%)	27:28	36/50 (72%)	29:56	35/50 (70%)	29:52	61	17	100/100	11	9,100 words	
S11	33/50 (66%)	28:33	33/50 (66%)	30:33	34/50 (68%)	30:31	74	14	62/100	9	11,600 words	
S12	43/50 (86%)	23:39	42/50 (84%)	18:22	39/50 (78%)	30:27	58	15	117/100	8	10,900 words	
S13	42/50 (84%)	30:30	48/50 (96%)	27:05	41/50 (82%)	13:11	63	16	117/100	8	not completed	
S14	42/50 (84%)	30:12	40/50 (80%)	30:17	44/50 (88%)	30:24	61	14	143/100	7	12,800 words	
S15	46/50 (92%)	22:57	40/50 (80%)	18:06	39/50 (78%)	24:21	67	13	150/100	9	10,700 words	
S16	31/50 (62%)	21:21	42/50 (84%)	25:56	34/50 (68%)	28:05	66	14	82/100	8	9,100 words	
S17	38/50 (76%)	28:05	38/50 (76%)	30:39	40/50 (80%)	30:15	65	12	125/100	10	9,700 words	
S18	44/50 (88%)	22:49	38/50 (76%)	21:07	44/50 (88%)	24:05	62	16	144/100	9	11,100 words	
S19	46/50 (92%)	16:06	42/50 (84%)	30:02	43/50 (86%)	30:14	67	12	110/100	8	not completed	
S20	39/50 (78%)	30:54	35/50 (70%)	33:11	21/50 (42%)	30:23	71	11	108/100	7	10,100 words	
SD/	7.521		5.369		5.583		25.567	64	14	22	8	1317.611
mean	37.368		37.75		37.333		109.157	Total: 2,374 words				10707.14

Legend: *bold italics*: decrease, **bold**: increase

8. Conclusion

The study presented the details of a semester-long vocabulary development project. It was discussed how the project fits into the general frame of corpus-based vocabulary instruction in the case of a university level skill development course by establishing blended learning as a pedagogical organizational framework. The vocabulary items were presented using interactive flashcard sets on *Quizlet*. The project included three formative online and one summative paper and pencil tests to measure development and mastery. The findings indicate that the e-materials successfully contributed to vocabulary development (RQ1) and by the end of the semester students' measured vocabulary levels increased by almost 700 words.

9. Limitations

The study discussed the data collected from the productive and receptive vocabulary test items that twenty learners completed during a semester. However, two limitations need to be noted. One, this sample is not representative for the whole first- or second-year English major population in Hungary and should therefore be viewed as more representative to the students at the University of Pécs. Two, although this major limitation is present, it was never a goal of the project to be generally representative, however, it is representative for the tested vocabulary items.

References

- Baepler, P. – Walker, J. D. – Driessen, M. (2014). It's not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Computers & Education*, 78, 227–236.
- Brezina, V. – Gablasova, D. (2015). Is there a core general vocabulary? Introducing the *New General Service List*. *Applied Linguistics*, 36(1), 1–22.
- Browne, C. (2014). A new general service list: The better mousetrap we've been looking for? *Vocabulary Learning and Instruction*, 3(2), 1–10.
- Browne, C. – Culligan, B. – Phillips, J. (2013a). *The New General Service List*. Available online at: <http://bit.ly/1j6EaZJ>.
- Browne, C. – Culligan, B. – Phillips, J. (2013b). *The New Academic Word List*. Available online at: <http://bit.ly/2aN7Hr1>.
- Coxhead, A. (2000). *The Academic Word List*. Available online at: <http://bit.ly/1iAfNzm>.
- Davies, M. (2012). *Word and Phrase*. Available online at: <http://bit.ly/1R7cPPP>.

- Gardner, D. – Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327.
- Laufer, B. – Elder, C. – Hill, K. – Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, 21(2), 202–226.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59–82.
- Nation, P. – Beglar, D. (2007). *Vocabulary size test (1-14k)*. Available online at: <http://bit.ly/1NLLDrS>.
- Olrich, D. C. – Harder, R. J. – Callahan, R. C. – Trevisan, M. S. – Brown, A. H. (2010). *Teaching strategies: A guide to effective instruction* (9th ed). Boston, MA: Wadsworth Cengage Learning.
- Paribakht, T. S. – Wesche, M. B. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In Coady, J. – Huckin, T. N. (eds.). *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge: Cambridge University Press, 174–200.
- Peachey, N. (2013). A blended learning teacher development course for the development of blended learning in English Language Teaching. In Tomlinson, B. – Whittaker, C. (eds.). *Blended Learning in English Language Teaching: Course Design and Implementation*. London: British Council, 65–74.
- Schmitt, N. (2008). Instructed second language vocabulary learning. Review article. *Language Teaching Research*, 12(3), 329–364.
- Sharma, R. (2010). Blended learning. *ELT Journal*, 64(4), 456–457.
- Simon, K. (2016). *Blending is trending. Applying blending learning to meet EFL students' language needs in listening and speaking skills development*. University of Pécs. Unpublished doctoral dissertation. Available online at: <http://bit.ly/2mSkeQu>.
- Simon, K. – Kollárová, K. (2015). Blending with Edmodo: The application of blended learning in a listening and speaking skills development course. In Lehmann, M. – Lugossy, R. – Horváth, J. (eds.). *UPRT 2015: Empirical studies in English applied linguistics* Pécs: Lingua Franca Csoport. ISBN 9789636429799. 193-217. Available online at: <http://bit.ly/2ohjwbQ>.
- Timmis, I. (2015). *Corpus Linguistics for ELT: research and practice*. NY: Routledge.
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(1), 79–95.
- Wesche, M. – Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, 53(1), 13–40.
- West, M. (1953). *A general service list of English words*. London: Longman, Green & Co.