Educational Research Association
The International Journal of Research in Teacher Education
2018, 9(3): 44-63
ISSN: 1308-951X

http://www.eab.org.tr

http://ijrte.eab.org.tr

# Corpus Linguistics Approach to In-Service Teacher Development

**Mogahed Abu Al-Fadl**[1]

**Abstract**

This study provided 10-week training in corpus linguistics and its relation to language teaching. Nineteen English teachers were randomly selected. It sought to give content for the training course in using corpus linguistics in language teaching. It also tried to find the effect of corpus linguistics on language teaching. The study adopts the quasi-experimental design in terms of using a pretest/ posttest design: paired sample t-test. A Corpus Linguistics Test was used as a pretest/posttest. It proved that there are statistically significant differences between the mean scores the corpus linguistics pretest and posttest in favour of the posttest scores.

**Keywords:** Teacher education, in-service, development, corpus linguistics.

[1] Assist. Prof. Dr., Qassim Private Colleges. mogahed72@windowslive.com

## Introduction

With the start of the 3rd millennium, fresh horizons have been opened before language Instruction. One of these is computer-assisted teaching (CAT), which has been promising teachers insightful techniques and teaching methods and tremendously giving a hand to inspect authentic language. in fact, technology has facilitated using authentic huge corpora to be accessed in the language classroom.

Society now demands multi-literacies which include a high proficiency in digital and on-line competencies: the use of electronically based language resources. Corpus literacy is the ability to use corpora—large, principled databases of spoken and written language—for language analysis and instruction. While linguists have emphasized the importance of corpus training in teacher preparation programs, few studies have investigated the process of initiating teachers into corpus literacy with the result that few guidelines exist for training teachers to make optimal use of corpus output (Heather & Helt, 2012).

Consequently, teachers need to have the necessary technical capabilities and these should be acquired during their formal language teacher education. It sensibly follows that language teacher educators have an important obligation in this regard. This will help provide a strong cognitive basis for one of the most crucial roles of language teachers in today's climate, that of lifelong learner (Egbert, Paulus & Nakamichi, 2002).

### Corpus linguistics and teacher education

The relevance of corpus linguistics to teacher education is under-explored, particularly in terms of teachers' language awareness. Tsui (2004) argues that more attention has been paid to the importance of raising teachers' language awareness. Teachers' language awareness is one area in which corpus linguistics has a unique contribution to make. It examines over one thousand grammar questions that English teachers in Hong Kong sent over a period of seven years to a website, TeleNex[2], to seek advice and demonstrates how empirical linguistic data which show the context and frequency of occurrence of the linguistic items in question can be a powerful tool to raise teachers' linguistic sensitivity, to help teachers question long-standing assumptions, and to gain new insights into language structure and use.

Whereas researchers have long seen the benefits of using corpora to enhance the description of language, the regular use of corpora in the EFL classroom is still a rare occurrence. One reason is likely to be that learning how to use corpora is seldom part of teacher training courses. As a result, teachers themselves, at university level and at lower levels, lack the skills needed to use this native-speaker consultant. If training in how to use corpora were integrated into university level courses such as syntax, written proficiency and translation, in time it could become just as natural to consult a corpus as to look up an item in a dictionary or a grammar book (Granath, 2009).

It is essential to train teachers in working with corpora so that they could design required materials themselves whenever they needed them. Mukherjee (2004) argue that corpora and concordance packages present very useful resources for the creation of exercises that motivate the learner and promote language awareness. Such courses could either be part of the general teacher training programme that every English language teacher has to do, or they could be offered to practicing teachers in the form of advanced teacher training workshops. However, since the schedules of teachers and teacher trainees tend to be rather full already, it might be more sensible to start this at the level of initial teacher training at universities and introduce future teachers to corpora and their pedagogical potential at this early stage. AnNayef (2001) adds that that the interest in concordancers rose up as language teachers came to realize that concordancers offer them and their students so many chances of approaching language that they

[2] http://www.telenex.hku.hk/telec/pmain/opening.htm

 

cannot have in traditional materials and approaches. A concordancer is mainly useful to the overloaded teacher. Concordance output lends itself readily to subdivision. This means that the output from concordance can simply be manipulated to create individual or group assignments. Apart from this, the teacher does not have to do most of the work as input from the teacher in this context is not central.

English Teachers have to use much advanced information technology into college English curriculum design, and develop a variety of computer and network courses. With the rapid development and wide application of computer technology, computer technology-based corpus technology becomes mature and become a powerful tool for the language study and teaching. Great worry given by the language researchers, contemporary corpora came into being (Guan, 2013). It is a clear that fluency in learner speech can only be achieved if the teacher model provides natural and fluent input. It is therefore vital for corpus-based insights into the nature of spoken language to play a much greater role in teacher education.

Huge banks of language data taking the form of language corpora are accessible now. These date can be analysed automatically using the appropriate software. Generally, there are three main stages when using this tool: extraction of data from texts, processing the output (reshaping according to your needs) and interpretation of output (asking the right questions). Corpus programmes have many advantages for language teachers: it creates word lists and counts occurrences of individual search items; it allows for the presentation and (re)organization of data in a way that facilitates the identification of patterns; it automatically produces cluster and collocation lists; and most software has a keyword tool, allowing a comparison of lexis between corpora to identify relatively significant items. In this way, it becomes easier to comment on the collocation, colligation, semantic preference (Sinclair, 1996).

Actually teachers often look for generalizations about grammar rules so that they can provide some guidelines to their students. This is perfectly legitimate especially in second language learning situations where learners do not have the same amount of exposure to the language as in first language learning situations. The problem is whether the rules and generalizations indeed capture how language is actually used rather than how language is perceived to be used, and whether they reflect the dominant patterns of use. The easy convenience of corpora lets teachers check set generalizations against linguistic data, inspires them to be sensitive to forms that arise from the data and to make their own generalizations of these patterns (Tsui ,2004).

Making corpora a normal part of teacher education will certainly serve to establish corpora in the classroom, but to speed things up, in-service courses for practising teachers should be taken seriously (Mauranen, 2004). The picture of the future for corpora in teaching is bright although tempered by what we know about attitudes of teachers and learners. As Romer (2006) points out corpus linguists have a tough job to meet the challenges from teachers and students who are used to more traditional methods. Corpora draw attention to complex patterns and phraseology rather than regularities and supports the view of language learning as a complex process involving hypothesis formation and testing.

It is suggested that teacher training courses should include a course in corpus analysis in one of the last years of university studies. Students who have completed their language courses can be assumed to be equipped with the skills necessary for interpreting and evaluating corpus data. In such a course, it should also be possible to make the tasks directly relevant to classroom work, so that students would raise questions, based on problems they have come across when giving feedback to their own students. Such a course must clearly be designed to include exercises in using corpora to find out about lexicon, phraseology and grammar patterns, as well as background reading on corpora and corpus linguistics. The same kind of course can be offered to language teachers already in-service; this will give them access to a native speaker consultant who does more than any native speaker could do (Granath, 2009).

 

The internet has brought many corpora and dedicated tools within reach of all teachers and learners. However, Boulton (2012) points to a common criticism that it is still that many of them require considerable investment in terms of training for learners and teachers – in-service or out-of-service - to understand the rationale as well as how to use them efficiently. Cuban (2001) adds that despite the many incentives and opportunities afforded to teachers in more privileged environments to pioneer the use of technology, a surprising degree of resistance remains. Corpora have an obvious place in the classroom but cannot replace the teacher or language teaching. However, the teacher has an important role to guide the students to the use of corpora in the classroom. On the other hand, when the fruitful outcomes of corpus linguistics are known and realized, it is clear that it is worth the investment.

**Literature review**

Studies of applications of corpus linguistics to second language teaching and learning have emphasized the importance of adopting a data-driven approach to language learning so that learners go through a process of self-discovery. The discussion in this paper shows that it is equally important, if not more important, for teachers to go through this process of self-discovery and to experience formulating generalizations about linguistic patterns that they have observed so that they own the grammar as much as linguistic researchers.

Heather and Helt (2012) evaluated corpus literacy training for pre-service language teachers. This study uses a case study approach to examine six pre-service language teachers' development of multiple components of corpus literacy during a semester-long introductory grammar course through which corpus linguistics was threaded. Results revealed that while corpus literacy training was largely effective, that effectiveness was various among subjects. Examining the sources of that variation suggests several practices for teacher educators planning or modifying instruction in corpus literacy.

Egbert, Paulus and Nakamichi (2002) examined the effect of CALL teaching on classroom computer practice for rethinking technology in teacher preparation. They indicated that there is a dearth of evaluative research examining student teachers' perceptions of learning and teaching through corpus-based activities. Their investigation of these pertinent issues with a participant group of 25 student teachers led to the conclusion that there is generally a positive predisposition towards the use of corpora.

Leńko-Szymańska (2014) focuses on a teacher training course on the practice of corpora in language education offered to teacher students at the Institute of Applied Linguistics. In addition, it shows the results of two questionnaires distributed to the students before and after the second edition of the course. The course seeks to provide students with the concept of a corpus and its analysis; to familiarize them with a range of available corpora, corpus-based resources and tools; and to show them numerous applications of corpora in language education, emphasizing the in-house preparation of courses, teaching materials and class activities. In the first part of the study, the design, the syllabus, the progression and the outcomes of the course are presented. In the second part, the responses of thirteen students participating in the second edition of the course are analyzed. The analysis indicates that in general the students reacted positively to the course and they saw the benefits of corpus-based materials and tools in language teaching. Yet the students reported that they needed more time to gain full command of the resources and software and more guidance on the pedagogical issues related to corpus use. The study ends that fourteen sessions, designed as an overview of the whole range of corpus-based resources and applications, is not enough to encourage teacher trainees to use corpora in their future work if they have no contact with these resources and tools in other classes. Only extensive exposure to corpora by future teachers together with suitable teacher training in the applications of corpora in language education may bring a considerable change in the scope of corpus use in language classrooms in the wide educational context.

 

Farr (2008) explored the use of electronic corpora in teacher education programmes. In spite of arguments for and against their use, there is a lack of evaluative research investigating student teachers' perceptions of learning and teaching via corpus-based activities. This paper has two main foci. Firstly, it reports some of the ways in which corpora have been incorporated into a language systems module on an MA in English Language Teaching programme over a two-year period. More significantly, it outlines the findings from survey results, which uncover student teachers' perspectives on their experiences of using corpora. Additionally, it explores the potentials and problems foreseen by these practitioners in relation to using such an approach in their careers. The examination of these relevant issues with a group of 25 student teachers leads to the decision that there is generally a positive predisposition towards the use of corpora. These attitudes vary in relation to the projected adaptation in EL teaching, and the results also show that the real teaching scenario often does not permit the ideal of full application. The study concluded that the continued integration of corpus-based instruction in the language content component of language teacher education programmes should be encouraged despite some identified difficulties.

**Questions of the Study**

English teachers need to use corpora and corpus-based resources and tools in language teaching. Teachers of English are reluctant to use corpus linguistics because they lack confidence and skills of using it. So they need training in the use of it as found in the survey done by Tribble (2012). The problem which is probably at the heart of teachers' disinclination to exploit corpora in language instruction is their lack of knowledge about the different ways that large linguistic databases can be used in the classroom (Mukherjee, 2004; Romer, 2010). The problem of the study is stated in the following questions:

1. What is the content of the proposed training course in the use corpus linguistics in language teaching?
2. Does corpus linguistics have an effect on language teaching?

**Hypothesis of the study**

There are statistically significant differences between the mean scores the corpus linguistics pretest and posttest in favour of the posttest scores.

**Purpose of the Study**

This study aims to:

- Preparing a training course in the use corpus linguistics in language teaching?
- Determining the effectiveness of a training course in the use corpus linguistics in language teaching?

**Participants and Research Setting**

Nineteen faculty members, Department of English, College of Administration and Humanities, Kingdom of Saudi Arabia, were randomly selected for the second semester, 2017/2018 academic year. All the study participants agreed and welcomed participating in the study.

**Instrument of the Study**

A Corpus Linguistics Test was prepared and used by the researcher as a posttest. *(See Appendix 2)*

**Test Item Difficulty**

To specify the difficulty level of the test items, a measure named the *Difficulty Index* is used. This measure calculates the proportion of participants who answered the test item accurately. By looking at each alternative for multiple choices, we can also find out if there are answer choices that should be replaced. We can compute the difficulty of the item by dividing the number of

 

participants who choose the correct answer by the number of total students. A rough "rule-of-thumb" is that if the item difficulty is more than .75, it is an easy item; if the difficulty is below .25, it is a difficult item. Given these parameters, difficulty of the Corpus Linguistics Test items ranges from .75 to .25.

**Test Item Discrimination**

*Discrimination Index* refers to how well an assessment differentiates between high and low scorers. Then the assessment is said to have a *positive discrimination index* (between 0 and 1) indicating that participants who received a high total score chose the correct answer for a specific item more often than the students who had a lower overall score. If, however, it is found that more of the low-performing participants got a specific item correct, then the item has a *negative discrimination index* (between -1 and 0). Discrimination Index is determined by subtracting the number of participants in the lower group who got the item correct from the number of students in the upper group who got the item correct. Then, divide by the number of students in each group. Given these parameters, discrimination of the Corpus Linguistics Test items ranges from 0 to 1.

**Test Validity**

To achieve test validity, the test was submitted to a specialized jury in TEFL and linguistics to respond to some criteria for validating the test. The jury recommended making some modifications to the test and the researcher carried them out. Hence, the test is valid after introducing the jury's suggested modifications. (*See appendix 4 for Criteria of the Corpus Linguistics Test*)

**Test administration**

The Corpus Linguistics Test was administered to the study sample at the end of the 2nd term of the 2017/2018 academic year, following the training period. The training period lasted for 10 weeks during the 2nd semester.

**Methodology**

The study adopts the quasi-experimental design in terms of using a pretest/ posttest design: paired sample t-test, sometimes called the dependent sample t-test. It is done with one group (no comparison/ control group) of participants. Participants are pre-tested, receive an intervention/treatment and are post-tested. The purpose of the test is to determine whether there is statistical evidence that the mean difference between paired observations on a particular outcome is significantly different from zero.

**Procedures**

1. Reviewing the literature related to corpus linguistics and its relation to language teaching.
2. Selecting the sample randomly.
3. Preparing the Corpus Linguistics Test.
4. Submitting the Corpus Linguistics Test to a group of jurors for validity.
5. Administering the Corpus Linguistics Pretest to the study sample.
6. Providing the 10-week training period.
7. Administering the Corpus Linguistics Posttest to the study sample to measure the effectiveness of the training course.
8. Analyzing the data statistically using SPSS programme, version 16.
9. Reporting results, conclusions and suggesting recommendations.

 

**Intervention**

The participants have two hours a week of corpus linguistics training. The researcher gives explanation of the corpus linguistics topics and provides participants the opportunity to do exercises via the internet. Following is a description of the week by week training:

**Week 1:** An introduction to the course. The participants have an idea about the main features, types of activities and topics of the course. The participants are asked to have their laptops with them to do some practice and exercises during the training. Presenting key terms of corpus linguistics, such as collocation, concordance, corpus, corpus-based and corpus-driven.
**Week 2:** Giving a brief review of why we use a corpus and the benefits for language teaching.
**Week 3:** Annotation and mark-up: giving a brief overview of how corpus texts may be enriched with additional information to ease analysis. Examining a range of different types of corpora.
**Week 4:** Exploring the value of frequency data in corpus linguistics and explaining in detail a key concept in corpus linguistics: collocation.
**Week 5:** Using corpora in language teaching
**Week 6:** Looking at colligation and key concepts associated with collocation such as semantic preference and discourse prosody.
**Week 7:** Giving an introduction to a key method in corpus linguistics: keyword analysis. Discussing an extension of the notion of keywords, and looking at key words over time and the notion of lock words.
**Week 8:** Integrating the corpus method with other methods such as qualitative and quantitative analyses. Reviewing early studies which used corpora in the creation of language teaching materials.
**Week 9:** Lexical syllabus: discussing the work of Sinclair and Renouf, who argued for word frequency to be a central organising principle in language teaching.
**Week 10:** Giving an introduction to a proposal for language teaching: data driven learning.

**Results & Discussion**

Table 1 gives univariate descriptive statistics (mean, sample size, standard deviation, and standard error) for each variable entered. It shows that the posttest mean is higher than the pretest mean.

*Table 1: Paired samples statistics*

| | Mean | N | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| Pretest | 11.9474 | 19 | 1.35293 | .31038 |
| Posttest | 20.5789 | 19 | 2.24390 | .51479 |

Table 2 shows the bivariate Pearson correlation coefficient (with a two-tailed test of significance) for each pair of variables entered.

*Table 2: Paired samples correlations*

| | N | Correlation | Sig. |
|---|---|---|---|
| Pretest & Posttest | 19 | -.227- | .349 |

Table 3 gives the hypothesis test results. The estimated *t* value for the test as a whole is statistically significant at ($\alpha \leq .05$) level. From these results, the study alternative hypothesis is accepted that there are statistically significant differences between the mean scores the corpus linguistics pretest and posttest in favour of the posttest scores. This may be attributed to providing a training course in corpus linguistics and its relation to language teaching.

*Table 3: Paired Samples Test*

| | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | Lower | Upper | | | |
| Pretest-Posttest | -8.631E0 | 2.87152 | .65877 | -10.01561- | -7.24755- | -13.103- | 18 | .000 |

As mentioned in the literature review, some studies support this result. Heather and Helt (2012) indicated that corpus literacy training was effective for teachers. Leńko-Szymańska's study (2014) showed that only extensive exposure to corpora with suitable teacher training in the applications of corpora in language education may bring a considerable change in the scope of corpus use in language classrooms in the wide educational context. Farr (2008) concluded that the continued integration of corpus-based instruction in the language content component of language teacher education programmes should be encouraged despite some identified difficulties. However, their studies were done with pre-service teachers.

**Conclusions**

My experience with teaching a course of this kind to faculty members, English Department, during the second term of the 2017/2018 academic term, was effective, and the feedback I received from the course participants, especially on the relevance of the topics for them as English faculty members was overwhelming.

This study is relevant to teacher education. There has been a new generation of learner dictionaries online. The the notion of "data-driven learning" is gaining in importance. Moreover, the compilation and analysis of learner corpora in language teaching. Conrad (2000) argues that corpus linguistics could revolutionize language teaching by changing the ways we approach all areas of teaching, such as materials development, curriculum design, teaching methodology and teacher training. Additionally, the associates between corpus linguistics and language teaching have been made. Therefore, it is vital to explore the key areas of interface between corpus linguistics and language teaching.

As native speaker corpora offer key information about the frequencies of linguistic features and their distributions across language use, Meunier (2002) contends that this information alone is not sufficient to inform curriculum and materials design: Within an EFL framework it is significant to achieve a balance between frequency, difficulty and pedagogical relevance. That is exactly where learner corpus research may have an effect in this regard. Learner corpus research offers further improvement in identifying those forms that are problematic for learners.

The contribution of this study has been to confirm that there should be more professional development for teachers, especially in modern innovations, such corpus linguistics due to the rapid changes in it in relation to language teaching. An implication of this study is that corpus exploration cannot be left to one course within a teacher training programme. As a result, there should be a follow-up to the topic until it enters mainstream education in language departments and teacher-training institutions on a large scale.

**References**

AnNayef, M. (2001). Corpora, concordances, and collocations in classroom teaching: Designing listening materials. *The First International TEFL Conference*, ESP Centre, Damascus University, Damascus, Syria, 27-29. Retrieved from: https://www.academia.edu/178284/Corpora_concordances_and_collocations_in_classroom_teaching_Designing_listening_materials

 

Boulton, A. (2012). *The Call Triangle: Student, teacher and institution: What data for data-driven learning?* EUROCALL (European Association for Computer-Assisted Language Learning). Retrieved from: http://webcache.googleusercontent.com/search?q=cache:K6PpB6Dx6wIJ:eurocall.webs.upv.es/documentos/newsletter/papers_20(1)/07_boulton.pdf+&cd=1&hl=ar&ct=clnk&gl=sa

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly 34*, 548–60.

Cuban, L. (2001) *Oversold and Underused: Computers in the Classroom.* Cambridge, MA: Harvard University Press.

Egbert, J., Paulus, T.M. & Nakamichi, Y. (2002) The impact of CALL instruction on classroom computer use: A foundation for rethinking technology in teacher education. *Language Learning and Technology, 6* (3), 108–126.

Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness,17* (1), 25-43.

Granath, S. (2009). Who benefits from learning how to use corpora? In Karin Aijmer (Ed.), *Corpora and language teaching*. Amsterdam: John Benjamins.

Guan, X. (2013). A Study on the Application of Data-driven Learning in Vocabulary Teaching and Leaning in China's EFL Class. *Journal of Language Teaching and Research, 4* (1), 105-112.

Heather, J. & Helt, M. (2012). Evaluating Corpus Literacy Training for Pre-Service Language Teachers: Six Case Studies. *Journal of Technology and Teacher Education, 20* (4), 415-440.

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*, (2), 1-19.

Mauranen, A. (2004). Spoken corpus for an ordinary learner. In Sinclair, H. M. (Ed.), *How to use corpora in language teaching*. Amsterdam: John Benjamins.

Meunier, F. (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In Granger, S., Hung, J. and Petch-Tyson, S., eds, *Computer learner corpora, second language acquisition and foreign language teaching.* Philadelphia: John Benjamins, 119–42.

Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In: Connor, U. and Upton, T. (Eds.), *Applied corpus linguistics: A multidimensional perspective*. Amsterdam/New York: Rodopi, 239–250.

Romer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik, 54* (2), 121–134.

Romer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. In Campoy-Cubillo, M.C., Bellés-Fortuño, B.and Gea-Valor., L.(Eds.), *Corpus-based approaches to English language teaching*. London: Continuum, 18–35.

Sinclair, J. (1996). The search for units of meaning. *Textus*, IX: 75–106.

Tsui, A. B. (2004). What teachers have always wanted to know. In Sinclair, H. M. (Ed.), *How to use corpora in language teaching*. Amsterdam: John Benjamins.

 

## Appendix 1

## Description of Corpus Linguistics Test

### General Description

This test measures faculty members' understanding of corpus linguistics. It consists 25 multiple choice items. The test is designed to measure faculty members' understanding of corpus linguistics after the implementation of the training course. The topics to be tested are shown in Table 4:

*Table 4: Topics of the Corpus Linguistics Test*

| Question | Topic |
|---|---|
| .1 | Definition of a corpus. |
| .2 | The main reason for using corpora |
| .3 | Definition of corpus annotation |
| .4 | Definition of a specialized corpus |
| .5 | Types of corpus |
| .6 | Definition of British National Corpus |
| .7 | Definition of a monitor corpus |
| .8 | Definition of a concordance |
| .9 | Definition of collocation |
| .10 | Definition of frequency distribution |
| .11 | Definition of a lock work |
| .12 | Example of a collocation |
| .13 | Definition of a colligation |
| .14 | Example of a colligation |
| .15 | Definition of semantic preference |
| .16 | Example of a semantic preference |
| .17 | Definition of discourse prosody |
| .18 | Example of discourse prosody |
| .19 | Definition of a key word method |
| .20 | Relationship between corpora and language teaching |
| .21 | Relationship between corpus-based approaches and language teaching |
| .22 | Concept of a frequent verb |
| .23 | Definition of a lexical syllabus |
| .24 | Definition of lexical bundles |
| .25 | Definition of data-driven learning |

The test is constructed in the light of the following:

1. Reviewing related literature concerning corpus linguistics in relation language education.

2. Suitability and clarity of the test items.

Following are characteristics of the test in some detail so as to establish some criteria upon which the test should be designed.

### Prompt Attributes

With this test, test items are shown to faculty members and they are asked to choose the correct answer from four responses. Scoring is easy and reliable as it is an objective test. The post-test scores will indicate whether the participants' understanding of corpus linguistics has developed or not. The participants are asked to answer all 20 questions. Each question focuses on a separate point of corpus linguistics.

**Response Attributes**

There are answer keys to the questions. Each question has one answer. Therefore, Faculty members' responses should be the same. They should answer all questions. Answers are on the same sheets of questions. Scoring the test is objective.

## Appendix 2

## Corpus Linguistics Test

Time Allowed: 2 hours

Name:----------.

Instructions

- Following are 20 multiple choice questions.
- Read the questions carefully.
- Answer all questions.
- Answers are on the test's sheets
- Put a tick in the box next to the right answer.
- Abide by the time allowed.

**Choose the right answer from A, B, C and D.**

| | |
|---|---|
| • **What is a corpus?** | |
| A. A theory of language. | |
| B. A collection of texts stored on a computer. | |
| C. An electronic database similar to a dictionary. | |
| D. Any large collection of words such as a collection of books, newspapers or magazines. | |
| • **What is the main reason for using corpora?** | |
| A. Other methods of language analysis are not reliable. | |
| B. Computers can confirm our intuitions about language. | |
| C. Computers can help us discover interesting patterns in language which would be difficult to spot otherwise. | |
| D. With corpora we can answer all research questions about language. | |
| • **What is corpus annotation?** | |
| A. Adding an extra layer of information to the text to allow for more sophisticated searches. | |
| B. Separating text into sentences. | |
| C. Manual coding of text for parts of speech. | |
| D. Adding critical comments to a text. | |
| A. **What is a specialised corpus?** | |
| B. A corpus that is used for historical language investigations. | |
| C. A corpus that is composed of a large variety of genres. | |
| D. A corpus that is used by language specialists. | |
| E. A corpus that focuses on e.g. one type of genre, one period, one place etc. | |
| • **Which of these is NOT a type of corpus** | |
| A. Multilingual corpus | |
| B. Learner corpus | |
| C. Diachronic corpus | |
| D. Observer corpus | |
| • **What is the BNC?** | |

| | |
|---|---|
| A. A large general corpus of British English. | |
| B. A corpus of different genres of English writing. | |
| C. A large spoken corpus of British English. | |
| D. A specialised corpus representing the language of newspapers. | |
| • **Which of these statements is NOT true about a monitor corpus?** | |
| A. It is frequently updated. | |
| B. The Bank of English is an example of a monitor corpus. | |
| C. The BNC is an example of a monitor corpus. | |
| D. It is used to monitor rapid change in language. | |
| • **What is a concordance?** | |
| A. Information about word frequencies normalised per million words. | |
| B. Listing of examples of a word searched in a corpus with some context on the right and some context on the left. | |
| C. An alphabetical list of words that appear in a text. | |
| D. A list of words and their frequencies that can be used for identifying important words in a text. | |
| • **What is collocation?** | |
| A. The tendency of speakers to talk over each other. | |
| B. The tendency of words to co-occur with one another. | |
| C. The tendency of words to appear in unique, different contexts each time. | |
| D. The tendency of sentences to create meaning. | |
| • **What is a frequency distribution in a corpus?** | |
| A. Information about how frequent a word is in a corpus. | |
| B. Information about the frequency of use of a term across a number of different texts, corpus sections, speakers etc. | |
| C. Information about how frequent a word is per million words. | |
| D. Sociolinguistic information about the gender of the speakers that are represented in a corpus. | |
| • **What is a lock word?** | |
| A. A type of a keyword. | |
| B. A word that has more or less the same frequency over time. | |
| C. A word that steadily becomes more frequent. | |
| D. A word that steadily becomes less frequent. | |
| • **Which of these is a prototypical example of a collocation?** | |
| A. tell-story | |
| B. surprisingly-unsurprisingly | |
| C. a-the | |
| D. help-aid | |
| • **What is a colligation?** | |
| A. A strong affinity of a word for another word. | |
| B. A strong affinity of a word for a grammatical class. | |
| C. A statistic that compares the co-occurrence of two words. | |
| D. A grammatical category of a word. | |
| • **Which of these is an example of a colligation?** | |
| A. telephone-operator | |
| B. Mr-proper noun | |
| C. back-front | |
| D. Subject-Verb-Object | |
| • **What is a semantic preference?** | |
| A. A relationship between a word and a set of other words that form a semantic category. | |
| B. A relationship between a word and a set of other words that do not form a | |

| | |
|---|---|
| semantic category. | |
| C. A relationship between a word and other words that form a grammatical category. | |
| D. A relationship between a word and a set of other words that do not form a grammatical category. | |
| • **Which of these is an example of semantic preference?** | |
| A. He-Verb | |
| B. Drinkable Liquids-Precious Stones | |
| C. Corpus-Linguistics | |
| D. Glass Of-Drinkable Liquids | |
| • **What is discourse prosody?** | |
| A. Collocations that reveal speaker socio-economic status. | |
| B. The way that words in a corpus can collocate with a related set of words or phrases, often revealing (hidden) attitudes. | |
| C. Intonation patterns in speech. | |
| D. Two words that do not go together. | |
| • **Which of these is an example of discourse prosody?** | |
| A. New-York | |
| B. Happen-Precious Stones | |
| C. Happen-Unpleasant Things | |
| D. Hair-Length | |
| • **What is the keyword method?** | |
| A. A method of identifying words that are statistically significantly more frequent in one corpus as compared with another corpus. | |
| B. A method of identifying words that speakers find important in a corpus. | |
| C. A method of identifying words that are 'key' for a particular corpus. | |
| D. A statistical procedure which identifies words which co-occur with other words. | |
| • **What new information can corpora bring to language teaching?** | |
| A. Information about what forms are common (and therefore useful) in language. | |
| B. Information about which grammatical structures are correct and which incorrect. | |
| C. Information about all possible words in a language. | |
| D. Information about what teaching approaches are efficient. | |
| • **What did the early corpus-based approaches to language teaching focus on?** | |
| A. Grammar | |
| B. Vocabulary | |
| C. Pragmatics | |
| D. Stylistics | |
| • **What is the most frequent verb form according to George (1963)?** | |
| A. Plain stem after don't | |
| B. Simple present actual | |
| C. Past participle of state | |
| D. Simple past narrative | |
| • **What is "the lexical syllabus"?** | |
| A. A syllabus for teaching new words to native speakers. | |
| B. A syllabus for teaching English vocabulary proposed by Sinclair and Renouf. | |
| C. A syllabus for language teaching suggested by Sinclair and Renouf that is based on frequency information derived from a corpus. | |
| D. A syllabus for teaching English grammar proposed by Sinclair and Renouf. | |
| • **What are "lexical bundles"?** | |
| A. Simple grammatical structures that consist of a subject and a verb. | |

| | |
|---|---|
| B. Clusters of frequent letters that form a word. | |
| C. Sequences of words that occur frequently in language. | |
| D. Clusters of sentences that express similar meanings. | |
| • **What is data-driven learning?** | |
| A. Direct use of corpora and corpus-generated concordances in the language classroom. | |
| B. Language learning based on grammar books produced with the aid of corpora. | |
| C. Language learning based on dictionaries produced with the aid of corpora. | |
| D. Use of learner corpora as data in linguistic research. | |

***End of the Test***

## Appendix 3

## Answer Key

**Choose the right answer from A, B, C and D.**

| | |
|---|---|
| • **What is a corpus?** | |
| A. A theory of language. | |
| B. A collection of texts stored on a computer. | √ |
| C. An electronic database similar to a dictionary. | |
| D. Any large collection of words such as a collection of books, newspapers or magazines. | |
| • **What is the main reason for using corpora?** | |
| A. Other methods of language analysis are not reliable. | |
| B. Computers can confirm our intuitions about language. | |
| C. Computers can help us discover interesting patterns in language which would be difficult to spot otherwise. | √ |
| D. With corpora we can answer all research questions about language. | |
| • **What is corpus annotation?** | |
| A. Adding an extra layer of information to the text to allow for more sophisticated searches. | √ |
| B. Separating text into sentences. | |
| C. Manual coding of text for parts of speech. | |
| D. Adding critical comments to a text. | |
| • **What is a specialised corpus?** | |
| A. A corpus that is used for historical language investigations. | |
| B. A corpus that is composed of a large variety of genres. | |
| C. A corpus that is used by language specialists. | |
| D. A corpus that focuses on e.g. one type of genre, one period, one place etc. | √ |
| • **Which of these is NOT a type of corpus?** | |
| A. Multilingual corpus | |
| B. Learner corpus | |
| C. Diachronic corpus | |
| D. Observer corpus | √ |
| • **What is the BNC?** | |
| A. A large general corpus of British English. | √ |
| B. A corpus of different genres of English writing. | |
| C. A large spoken corpus of British English. | |
| D. A specialised corpus representing the language of newspapers. | |
| • **Which of these statements is NOT true about a monitor corpus?** | |
| A. It is frequently updated. | |

| | |
|---|---|
| B. The Bank of English is an example of a monitor corpus. | |
| C. The BNC is an example of a monitor corpus. | √ |
| D. It is used to monitor rapid change in language. | |
| • **What is a concordance?** | |
| A. Information about word frequencies normalised per million words. | |
| B. Listing of examples of a word searched in a corpus with some context on the right and some context on the left. | √ |
| C. An alphabetical list of words that appear in a text. | |
| D. A list of words and their frequencies that can be used for identifying important words in a text. | |
| • **What is collocation?** | |
| A. The tendency of speakers to talk over each other. | |
| B. The tendency of words to co-occur with one another. | √ |
| C. The tendency of words to appear in unique, different contexts each time. | |
| D. The tendency of sentences to create meaning. | |
| • **What is a frequency distribution in a corpus?** | |
| A. Information about how frequent a word is in a corpus. | |
| B. Information about the frequency of use of a term across a number of different texts, corpus sections, speakers etc. | √ |
| C. Information about how frequent a word is per million words. | |
| D. Sociolinguistic information about the gender of the speakers that are represented in a corpus. | |
| • **What is a lock word?** | |
| A. A type of a keyword. | |
| B. A word that has more or less the same frequency over time. | √ |
| C. A word that steadily becomes more frequent. | |
| D. A word that steadily becomes less frequent. | |
| • **Which of these is a prototypical example of a collocation?** | |
| A. tell-story | √ |
| B. surprisingly-unsurprisingly | |
| C. a-the | |
| D. help-aid | |
| • **What is a colligation?** | |
| A. A strong affinity of a word for another word. | |
| B. A strong affinity of a word for a grammatical class. | √ |
| C. A statistic that compares the co-occurrence of two words. | |
| D. A grammatical category of a word. | |
| • **Which of these is an example of a colligation?** | |
| A. telephone-operator | |
| B. Mr-proper noun | √ |
| C. back-front | |
| D. SUBJECT-VERB-OBJECT | |
| • **What is a semantic preference?** | |
| A. A relationship between a word and a set of other words that form a semantic category. | √ |
| B. A relationship between a word and a set of other words that do not form a semantic category. | |
| C. A relationship between a word and other words that form a grammatical category. | |
| D. A relationship between a word and a set of other words that do not form a grammatical category. | |
| • **Which of these is an example of semantic preference?** | |
| A. He-verb | |

| | |
|---|---|
| B. DRINKABLE LIQUIDS-PRECIOUS STONES | |
| C. corpus-linguistics | |
| D. glass of-DRINKABLE LIQUIDS | √ |
| • **What is discourse prosody?** | |
| A. Collocations that reveal speaker socio-economic status. | |
| B. The way that words in a corpus can collocate with a related set of words or phrases, often revealing (hidden) attitudes. | √ |
| C. Intonation patterns in speech. | |
| D. Two words that do not go together. | |
| • **Which of these is an example of discourse prosody?** | |
| A. New-York | |
| B. happen-PRECIOUS STONES | |
| C. happen-UNPLEASANT THINGS | √ |
| D. hair-LENGTH | |
| • **What is the keyword method?** | |
| A. A method of identifying words that are statistically significantly more frequent in one corpus as compared with another corpus. | √ |
| B. A method of identifying words that speakers find important in a corpus. | |
| C. A method of identifying words that are 'key' for a particular corpus. | |
| D. A statistical procedure which identifies words which co-occur with other words. | |
| • **What new information can corpora bring to language teaching?** | |
| A. Information about what forms are common (and therefore useful) in language. | √ |
| B. Information about which grammatical structures are correct and which incorrect. | |
| C. Information about all possible words in a language. | |
| D. Information about what teaching approaches are efficient. | |
| • **What did the early corpus-based approaches to language teaching focus on?** | |
| A. Grammar | |
| B. Vocabulary | √ |
| C. Pragmatics | |
| D. Stylistics | |
| • **What is the most frequent verb form according to George (1963)?** | |
| A. Plain stem after don't | |
| B. Simple present actual | |
| C. Past participle of state | |
| D. Simple past narrative | √ |
| • **What is "the lexical syllabus"?** | |
| A. A syllabus for teaching new words to native speakers. | |
| B. A syllabus for teaching English vocabulary proposed by Sinclair and Renouf. | |
| C. A syllabus for language teaching suggested by Sinclair and Renouf that is based on frequency information derived from a corpus. | √ |
| **D.** A syllabus for teaching English grammar proposed by Sinclair and Renouf. | |
| • **What are "lexical bundles"?** | |
| A. Simple grammatical structures that consist of a subject and a verb. | |
| B. Clusters of frequent letters that form a word. | |
| C. Sequences of words that occur frequently in language. | √ |
| D. Clusters of sentences that express similar meanings. | |
| • **What is data-driven learning?** | |
| A. Direct use of corpora and corpus-generated concordances in the language classroom. | √ |
| B. Language learning based on grammar books produced with the aid of corpora. | |

| | |
|---|---|
| C. Language learning based on dictionaries produced with the aid of corpora. | |
| D. Use of learner corpora as data in linguistic research. | |

## Appendix 4

### Criteria of the Corpus Linguistics Test

### Jury Members' Evaluation Sheet

Dear Professor,

This evaluation sheet is part of a study entitled "Corpus Linguistics Approach to In-Service Teacher Development." The study aims to introduce corpus linguistics to in-service university teachers to enhance their professional development. The evaluation sheet consists of 25 multiple choice questions. The bold statements are the correct answers. You are kindly asked to determine whether the questions are suitable or not. Any suggestions are most welcome. Thank you for your fruitful cooperation.

The researcher

Choose the right answer:

| | Suitability | |
|---|---|---|
| | Suitable | Unsuitable |
| **What is a corpus?** | | |
| A theory of language. | | |
| A collection of texts stored on a computer. | | |
| An electronic database similar to a dictionary. | | |
| Any large collection of words such as a collection of books, newspapers or magazines. | | |
| **What is the main reason for using corpora?** | | |
| Other methods of language analysis are not reliable. | | |
| Computers can confirm our intuitions about language. | | |
| Computers can help us discover interesting patterns in language which would be difficult to spot otherwise. | | |
| With corpora we can answer all research questions about language. | | |
| **What is corpus annotation?** | | |
| Adding an extra layer of information to the text to allow for more sophisticated searches. | | |
| Separating text into sentences. | | |
| Manual coding of text for parts of speech. | | |
| Adding critical comments to a text. | | |
| **What is a specialised corpus?** | | |
| A corpus that is used for historical language investigations. | | |
| A corpus that is composed of a large variety of genres. | | |
| A corpus that is used by language specialists. | | |
| A corpus that focuses on e.g. one type of genre, one period, one place etc. | | |
| **Which of these is NOT a type of corpus?** | | |
| Multilingual corpus | | |
| Learner corpus | | |
| Diachronic corpus | | |
| Observer corpus | | |
| **What is the BNC?** | | |
| A large general corpus of British English. | | |
| A corpus of different genres of English writing. | | |

| | | | |
|---|---|---|---|
| A large spoken corpus of British English. | | | |
| A specialised corpus representing the language of newspapers. | | | |
| **Which of these statements is NOT true about a monitor corpus?** | | | |
| It is frequently updated. | | | |
| The Bank of English is an example of a monitor corpus. | | | |
| The BNC is an example of a monitor corpus. | | | |
| It is used to monitor rapid change in language. | | | |
| **What is a concordance?** | | | |
| A. Information about word frequencies normalised per million words. | | | |
| B. Listing of examples of a word searched in a corpus with some context on the right and some context on the left. | | | |
| C. An alphabetical list of words that appear in a text. | | | |
| D. A list of words and their frequencies that can be used for identifying important words in a text. | | | |
| **What is collocation?** | | | |
| A. The tendency of speakers to talk over each other. | | | |
| B. The tendency of words to co-occur with one another. | | | |
| C. The tendency of words to appear in unique, different contexts each time. | | | |
| D. The tendency of sentences to create meaning. | | | |
| **What is a frequency distribution in a corpus?** | | | |
| A. Information about how frequent a word is in a corpus. | | | |
| B. Information about the frequency of use of a term across a number of different texts, corpus sections, speakers etc. | | | |
| C. Information about how frequent a word is per million words. | | | |
| D. Sociolinguistic information about the gender of the speakers that are represented in a corpus. | | | |
| **What is a lock word?** | | | |
| A. A type of a keyword. | | | |
| B. A word that has more or less the same frequency over time. | | | |
| C. A word that steadily becomes more frequent. | | | |
| D. A word that steadily becomes less frequent. | | | |
| **Which of these is a prototypical example of a collocation?** | | | |
| A. tell-story | | | |
| B. surprisingly-unsurprisingly | | | |
| C. a-the | | | |
| D. help-aid | | | |
| **What is a colligation?** | | | |
| A. A strong affinity of a word for another word. | | | |
| B. A strong affinity of a word for a grammatical class. | | | |
| C. A statistic that compares the co-occurrence of two words. | | | |
| D. A grammatical category of a word. | | | |
| **Which of these is an example of a colligation?** | | | |
| A. telephone-operator | | | |
| B. Mr-proper noun | | | |
| C. back-front | | | |
| D. SUBJECT-VERB-OBJECT | | | |
| **What is a semantic preference?** | | | |
| A. A relationship between a word and a set of other words that form a semantic category. | | | |
| B. A relationship between a word and a set of other words that do not form a semantic category. | | | |
| C. A relationship between a word and other words that form a grammatical category. | | | |

| | | | |
|---|---|---|---|
| D. A relationship between a word and a set of other words that do not form a grammatical category. | | | |
| **Which of these is an example of semantic preference?** | | | |
| A. He-verb | | | |
| B. DRINKABLE LIQUIDS-PRECIOUS STONES | | | |
| C. corpus-linguistics | | | |
| D. glass of-DRINKABLE LIQUIDS | | | |
| **What is discourse prosody?** | | | |
| A. Collocations that reveal speaker socio-economic status. | | | |
| B. The way that words in a corpus can collocate with a related set of words or phrases, often revealing (hidden) attitudes. | | | |
| C. Intonation patterns in speech. | | | |
| D. Two words that do not go together. | | | |
| **Which of these is an example of discourse prosody?** | | | |
| A. New-York | | | |
| B. happen-PRECIOUS STONES | | | |
| C. happen-UNPLEASANT THINGS | | | |
| D. hair-LENGTH | | | |
| **What is the keyword method?** | | | |
| A. A method of identifying words that are statistically significantly more frequent in one corpus as compared with another corpus. | | | |
| B. A method of identifying words that speakers find important in a corpus. | | | |
| C. A method of identifying words that are 'key' for a particular corpus. | | | |
| D. A statistical procedure which identifies words which co-occur with other words. | | | |
| **What new information can corpora bring to language teaching?** | | | |
| A. Information about what forms are common (and therefore useful) in language. | | | |
| B. Information about which grammatical structures are correct and which incorrect. | | | |
| C. Information about all possible words in a language. | | | |
| D. Information about what teaching approaches are efficient. | | | |
| **What did the early corpus-based approaches to language teaching focus on?** | | | |
| A. Grammar | | | |
| B. Vocabulary | | | |
| C. Pragmatics | | | |
| D. Stylistics | | | |
| **What is the most frequent verb form according to George (1963)?** | | | |
| A. Plain stem after don't | | | |
| B. Simple present actual | | | |
| C. Past participle of state | | | |
| D. Simple past narrative | | | |
| **What is "the lexical syllabus"?** | | | |
| A. A syllabus for teaching new words to native speakers. | | | |
| B. A syllabus for teaching English vocabulary proposed by Sinclair and Renouf. | | | |
| C. A syllabus for language teaching suggested by Sinclair and Renouf that is based on frequency information derived from a corpus. | | | |
| D. A syllabus for teaching English grammar proposed by Sinclair and Renouf. | | | |
| **What are "lexical bundles"?** | | | |

| | | | |
|---|---|---|---|
| A. Simple grammatical structures that consist of a subject and a verb. | | | |
| B. Clusters of frequent letters that form a word. | | | |
| C. Sequences of words that occur frequently in language. | | | |
| D. Clusters of sentences that express similar meanings. | | | |
| **What is data-driven learning?** | | | |
| A. Direct use of corpora and corpus-generated concordances in the language classroom. | | | |
| B. Language learning based on grammar books produced with the aid of corpora. | | | |
| C. Language learning based on dictionaries produced with the aid of corpora. | | | |
| D. Use of learner corpora as data in linguistic research. | | | |

If there is something else to be added, omitted, modified, from your point of view, would you provide it, please?

I think the following should be added:

……
……
……

I think the following should be omitted:

……
……
……

I think the following should be modified:

……
……
……

Additional comments:
Please add any items and/or comments that you consider important for the test validation.
……
……
……