advances in Technology

Adv. Technol. **2022**, *2(4), 402-405*

Communications

Application of Digital Corpus Analysis Technologies to address Low English Proficiency' in undergraduates in English Medium Higher Education

Zareena D. Hussain* and Savitri Y. Dias

Department of English Language Teaching, Faculty of Arts, University of Colombo, Colombo 00300, Sri Lanka.

*Email Correspondence: zareenadh@gmail.com (Z. D. Hussain)

Received: 20 August 2022; Revised: 23 October 2022; Accepted: 29 October 2022; Published: 09 November 2022

Undergraduates with Low English Proficiency (LEP) in faculties that teach exclusively in the English medium would be linguistically ill-prepared to follow an English medium degree. One of the immediate needs created is the need to develop high discipline and course content specific (content related to the undergraduates' year of study, such as the first year) English proficiency within an extremely limited time [1]. This need could be more acute in first-year LEP students who need to transition to English Medium Instruction (EMI) in higher education. So is the need to develop discipline-specific (content related to the undergraduates' subjects of study) English proficiency for the first time after completing school education in the local languages [2]. Based on observations gathered through an endeavour to prepare English language assistance course material for English preliminary level undergraduates in EMI, this paper proposes digital Corpus Analysis technology as a feasible tool that can be used as a fast-track method to develop a discipline and course-specific English proficiency in LEP undergraduates.
This paper intends to highlight the following advantages of using digital Corpus Analysis technology, envisioned by the course designers in an English language course.

- High potential for use in providing discipline, subject, and course content-specific language training [3][4]
- User-friendly technology for both teachers and students
- Generating of editable data that can be word processed to create teaching-learning activities
- Teach authentic language in the context
- Free and Easy access to the digital Corpus Analysis technology [5]
- High potential for use in self-access and open and distance learning

To explain a few key terms

- Corpus is a collection of texts compiled to analyse and find how language is used. The source can be both written and spoken.
- Corpus Analysis Technologies are software developed to analyse electronically saved corpus for language features.

https://doi.org/10.31357/ait.v2i4.5989

- Concordances (or concordance lines) is a list of language samples each containing the same keyword or phrase.
- Word frequency is the number of times a word occurs in a given text or corpus.
- Tags are labels made of a few letters and numbers which indicate a part of speech. Example- DT- determiner, JJ- adjective, NN- noun, NN2- plural noun, VVD- past tense verb

This paper reports on experimental English language teaching learning material design insights born through an endeavour undertaken for the Preliminary level Intensive English course for first-year undergraduates of the Faculty of Management & Finance, University of Ruhuna, Sri Lanka as part of the Accelerating Higher Education Expansion and Development (AHEAD) operation Project ICE/RIC/DOR, ELTA- ELSE. The software used was 'Free CLAWS web tagger', a free web-based software available at http://ucrel-api.lancaster.ac.uk/claws/free.html, and two corpus analysis software namely AntConc and TagAnt designed by Laurence Anthony available for free download on https://www.laurenceanthony.net/

The software allows the manipulation of editable digitally saved texts such as transcribed lectures, textbooks, handouts, book chapters, examination questions, and assignments and instructions in order to analyse and understand the language structures and vocabulary there in and if necessary edit these texts to create language learning activities as shown below. The AntConc programme can be used to locate words based on their frequency and morphological structure (e.g words ending with 'ing'). Figure 1 depicts two activities that incorporate data generated by AntConc on the lexical word 'companies' and the preposition 'in'. Figure 2 depicts instructions to introduce the 'Free CLAWS web tagger' to the students as a tool to identify grammar structures in texts and an activity based on such tagged data produced by the software TagAnt to get students to identify particular structures in whole texts. All activities depicted here are based on a passage taken from the students' textbook *Fundamentals of Financial Management* [6]. Thus, these activities have the potential to help 'learners to pay attention to features of authentic input'. [7][8][9]

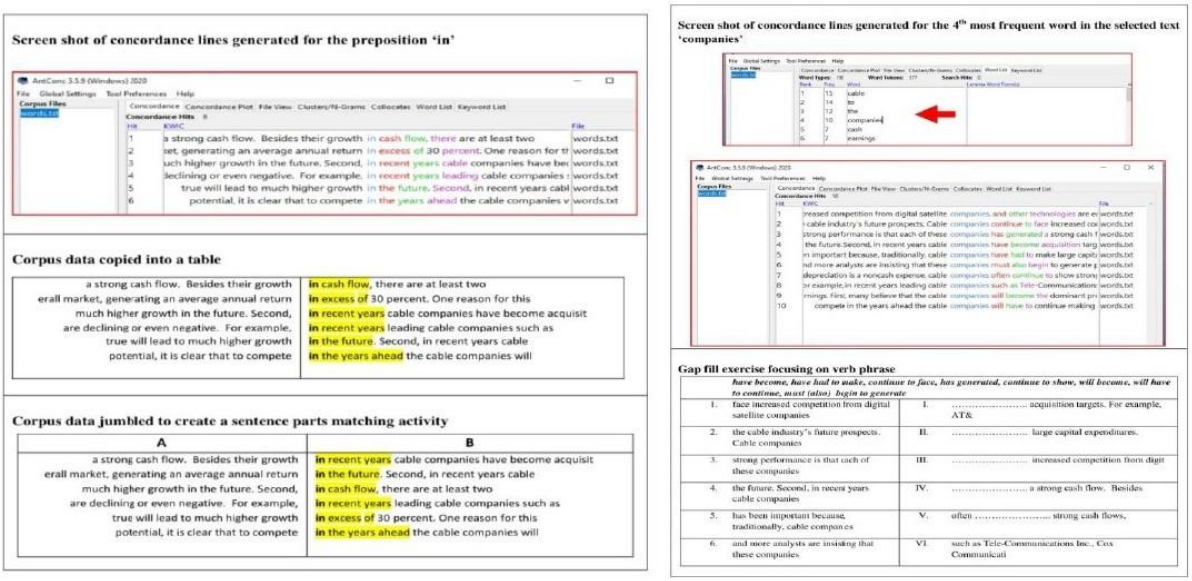

Figure 1- Using corpus analysis to isolate words and converting generated data into related language exercises

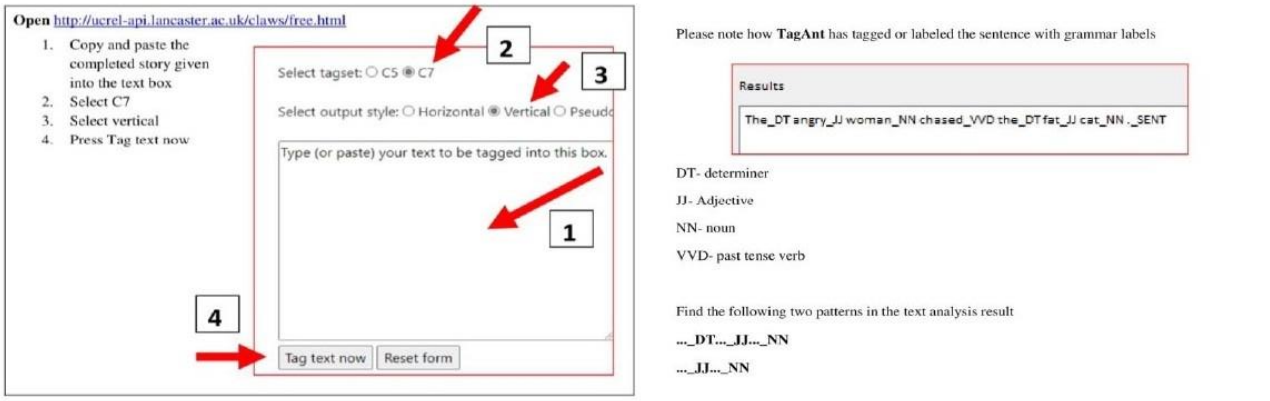

Figure 2- Using corpus analysis to tag texts for syntactic features and converting generated data into related language exercises

References

[1] T. Cobb, A. Boulton. Classroom applications of corpus analysis in *Cambridge Handbook of English Corpus Linguistics*, Cambridge, Cambridge University Press, **2015**, pp. 478-497. Available from https://www.researchgate.net/publication/280856533_Classroom_applications_of_corpus_analysis. [01 December 2021]

[2] Z. D. Hussain, Gathering Transition Perspectives: Views of Undergraduates with Low English Proficiency (LEP) in the First Year of English Medium Instruction (EMI) In Proceedings of the 2nd *SLIIT International Conference On Advancements In Sciences And Humanities*, Sri Lanka. Institute of Information Technology, Sri Lanka, 03 & 04 December **2021.** Available from: https://static.sliit.lk/wp-content/uploads/downloads/SICASH-2021-Conference-Proceedings.pdf [17 December 2021]

[3] E.E. Khairas, Using AntConc Software As English Learning Media: The Students' Perception, *Epigram*,

2019, 16 (2), 189–194. Available from: https://jurnal.pnj.ac.id/index.php/epigram/article/view/2234. [01 December 2021]

[4] N. Stojković, , Corpus Analysis for Language Studies at the University Level in *Theoretical Considerations For Teaching Foreign Languages In University Education Settings*, Newcastle upon Tyne, UK, Cambridge Scholars Publishing, **2021**, pp. 8-20. Available from: https://www.cambridgescholars.com/resources/pdfs/978-1-5275-6470-1-sample.pdf [27 December 2021]

[5] L. Anthony. AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit, In Proceedings of IWLeL 2004, International Conference Center, Waseda University, Japan, 10 December, **2004**, Available from: http://www.laurenceanthony.net/research/iwlel_2004_anthony_antconc.pdf. [01 December 2021]

[6] E.F. Brigham, & J.F. Houston, Fundamentals of Financial Management, South-Western College, United States of America, **2003,** p. 49

[7] B. Tomlinson, Developing Materials for Language Teaching, Cambridge, Cambridge University Press, **2013**, p. 7

[8] K. Z. Li, The Use of Concordance Programs in English Lexical Teaching in High School, *Higher Education of Social Science*, **2015**, 8 (1), 60-65..DOI: http://dx.doi.org/10.3968/626. [24 December 2021]

[9] S. Marinov, Training ESP Students in Corpus Use - Challenges Of Using Corpus-Based Exercises With Students of Non-Philological Studies, *Teaching English with Technology*, **2013**, 13 (4), 49-76. [24 December 2021]