Proceedings Paper:
Alfaifi, AYG and Atwell, E Comparative Evaluation of Tools for Arabic Corpora Search and Analysis. In: UNSPECIFIED BAAL / Cambridge University Press Applied Linguistics Seminar, 14th June 2014, Leeds Metropolitan University, Leeds. . (Unpublished)

Comparative Evaluation of Tools for Arabic Corpora Search and Analysis

Abdullah Alfaifi[1]
University of Leeds
scayga@leeds.ac.uk

Eric Atwell[2]
University of Leeds
e.s.atwell@leeds.ac.uk

Abstract

Learner corpora are considered one of the resources used in language learning and teaching (Nesselhauf, 2004) and second language acquisition (Granger, 2003). The Arabic Learner Corpus is a recently developed project that includes 282,732 words of written and spoken materials, collected from 942 learners of Arabic in Saudi Arabia. Different tools have been investigated in order to evaluate their functionalities and suitability for searching and analysing Arabic corpora, such as the KACST Arabic Corpora Processing Tool (Al-thubaity, 2014), aConCorde (Roberts, 2014), AntConc (Anthony, 2005), WordSmith Tools (Scott, 2012), The Sketch Engine (Kilgarriff et al., 2004), IntelliText Corpus Queries (Sharoff, 2011), CQPweb (Hardie, 2008) and others. This paper aims to cover three aspects: (1) an overview on learner corpora uses in language research, (2) a brief summary of the second version of the Arabic Learner Corpus, (3) and a particular focus on evaluating search and analysis tools that can be used on the Arabic corpora. This evaluation is based on specific criteria in order to conclude whether the tools provide functions suitable for Arabic, a right-to-left reading language, in order to be effectively searched and analysed. The criteria include, for example, the ability to read Arabic texts, showing right-to-left texts in the correct order and the availability of uploading a personal corpus.

References

Al-thubaity, Abdulmohsen. (2014). ACP Tool ver3.0. Retrieved 6 April 2014, from: https://www.researchgate.net/publication/256305105_ACP_Tool_ver3.0.

Anthony, Laurence. (2005). *AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom.* In the proceedings of the Professional Communication Conference, 10-13 July 2005, Limerick, Ireland.

Granger, Sylviane. (2003). The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research. *TESOL Quarterly, 37*(3): 538-546.

Hardie, Andrew. (2008). CQPweb tool: Corpus Query Processor. Retrieved 6 April 2014, from: https://cqpweb.lancs.ac.uk/.

Kilgarriff, Adam, Rychly, Pavel, Smrz, Pavel, and Tugwell, David. (2004). *The Sketch Engine.* In the proceedings of the Euralex, 6-10 July 2004, Lorient, France.

---

[1] http://www.comp.leeds.ac.uk/scayga
[2] http://www.comp.leeds.ac.uk/eric

Nesselhauf, Nadja. (2004). Learner Corpora and Their Potential in Language Teaching. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 125-152). Amsterdam & Philadelphia: Benjamins

Roberts, Andrew. (2014). aConCorde.   Retrieved 6 April 2014, from: http://www.andy-roberts.net/coding/aconcorde.

Scott, Mike. (2012). WordSmith Tools version 6, Liverpool: Lexical Analysis Software. from: http://www.lexically.net/wordsmith.

Sharoff, Serge. (2011). IntelliText Corpus Queries.   Retrieved 6 April 2014, from: http://smlc09.leeds.ac.uk/itweb/htdocs/Query.html#.