



UNIVERSITY OF LEEDS

This is a repository copy of *Arabic learner corpus and its potential role in teaching Arabic to non-native speakers*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/83968/>

Version: Accepted Version

Proceedings Paper:

Alfaifi, AYG and Atwell, ES Arabic learner corpus and its potential role in teaching Arabic to non-native speakers. In: Proceedings of American Association for Corpus Linguistics Conference. American Association for Corpus Linguistics Conference, 26-28 Sep 2014, Northern Arizona University, Flagstaff, Arizona. Learner Corpus Association .

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Arabic Learner Corpus and Its Potential Role in Teaching Arabic to Non-Native Speakers

Abdullah Alfaifi¹
University of Leeds
scayga@leeds.ac.uk

Eric Atwell²
University of Leeds
e.s.atwell@leeds.ac.uk

Abstract

The literature on learner corpora (Al-Sulaiti, 2010; Granger & Dumont, 2012) shows that there is a need to compile an Arabic learner corpus, which can be used in research on Arabic language learning and teaching. This paper introduces the project to build the Arabic Learner Corpus (ALC) (Alfaifi & Atwell, 2013), and its potential uses in teaching Arabic as a second language (ASL). ALC includes 282,732 words of written and spoken materials produced by learners of Arabic in Saudi Arabia. The corpus covers two types of students, non-native Arabic speakers (NNAS) and native Arabic speaking students (NAS) who are learning to improve their written Arabic. The corpus also covers two general levels of study, pre-university and university. The corpus data is available online for download in TXT and XML format at <http://www.arabiclearnercorpus.com>. Further work is being performed to annotate the entire corpus for errors. The growing role of learner corpora in the last decade has provided valuable insights into the uses of this type of corpora, such as error analysis, contrastive interlanguage analysis, making learners' dictionaries, monitoring learners' improvements and designing language materials. These uses were considered when building ALC as an open-source corpus.

References:

- Al-Sulaiti, Latifa. (2010). *Arabic Corpora*. 10 August 2012, from: the University of Leeds, Latifa Al-Sulaiti's Homepage: http://www.comp.leeds.ac.uk/eric/latifa/arabic_corpora.htm
- Granger, Sylviane, and Dumont, Amandine. (2012). *Learner corpora around the world*. 17 July 2012, from: the Université Catholique de Louvain, Centre for English Corpus Linguistics Web site: <http://www.uclouvain.be/en-cecl-lcworld.html>
- Alfaifi, Abdullah and Atwell, Eric. (2013). Arabic Learner Corpus v1: A New Resource for Arabic Language Research. In proceedings of the *Second Workshop on Arabic Corpus Linguistics (WACL-2)*, Lancaster University, UK.
<http://www.comp.leeds.ac.uk/eric/wacl/wacl2proceedings.pdf>

¹ <http://www.comp.leeds.ac.uk/scayga>

² <http://www.comp.leeds.ac.uk/eric>