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:

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

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

Abdullah Alfaifi[1]

University of Leeds

scayga@leeds.ac.uk

Eric Atwell[2]

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