This is a repository copy of Computer-aided error annotation: a new tool for annotating Arabic error.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/83911/

---

**Conference or Workshop Item:**

---

**Reuse**
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher’s website.

**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.
Current Data

The Arabic Learner Corpus currently includes 1585 texts in a raw format.

Target Data

The ALC will include further information describing the language errors.

The Annotation Tool

Main functions

1. Text Tokenisation
   The Text Tokenisation function helps in segmenting the text into separate word-based tokens, this enables the annotator to attach tags to those tokens include errors.

2. Smart-Selection
   When an error exists more than once, the Smart-Selection function enables the annotator to find them all and tag them in a single step with no need to repeat the annotation process with each error.

3. Auto-Tagging
   The Auto-Tagging feature, which is similar to translation memories, recognises the tokens that have been manually annotated and stores them in a database. So, tokens with the same error can be detected and annotated automatically next time.

For further information about ALC please visit: www.arabiclearnercorpus.com