

# UNIARAB: An Universal Machine Translator System for Arabic Based on Role and Reference Grammar

Yasser Salem and Brian Nolan

## Abstract

We use the RRG theory to motivate the architecture of the lexicon and the RRG bidirectional linking system to design and implement the parse and generate functions between the syntax-semantic interfaces. Our research has yielded significant results and in comparison with translations of (source) modern standard Arabic sentences in the native orthography from other software systems, we deliver more accurate and grammatical output in (target) English. This is, we believe, to the use of the RRG linguistic model in software.

## Challenges of Arabic MT

- ▶ Large set of morphological features.
- ▶ Features normally in the form of prefixes or suffixes that can completely change the meaning of the word.
- ▶ In Arabic there are some words that hold the meaning of a full sentence e.g., سنسافر, would translate to “we will travel” in English.
- ▶ Free word order

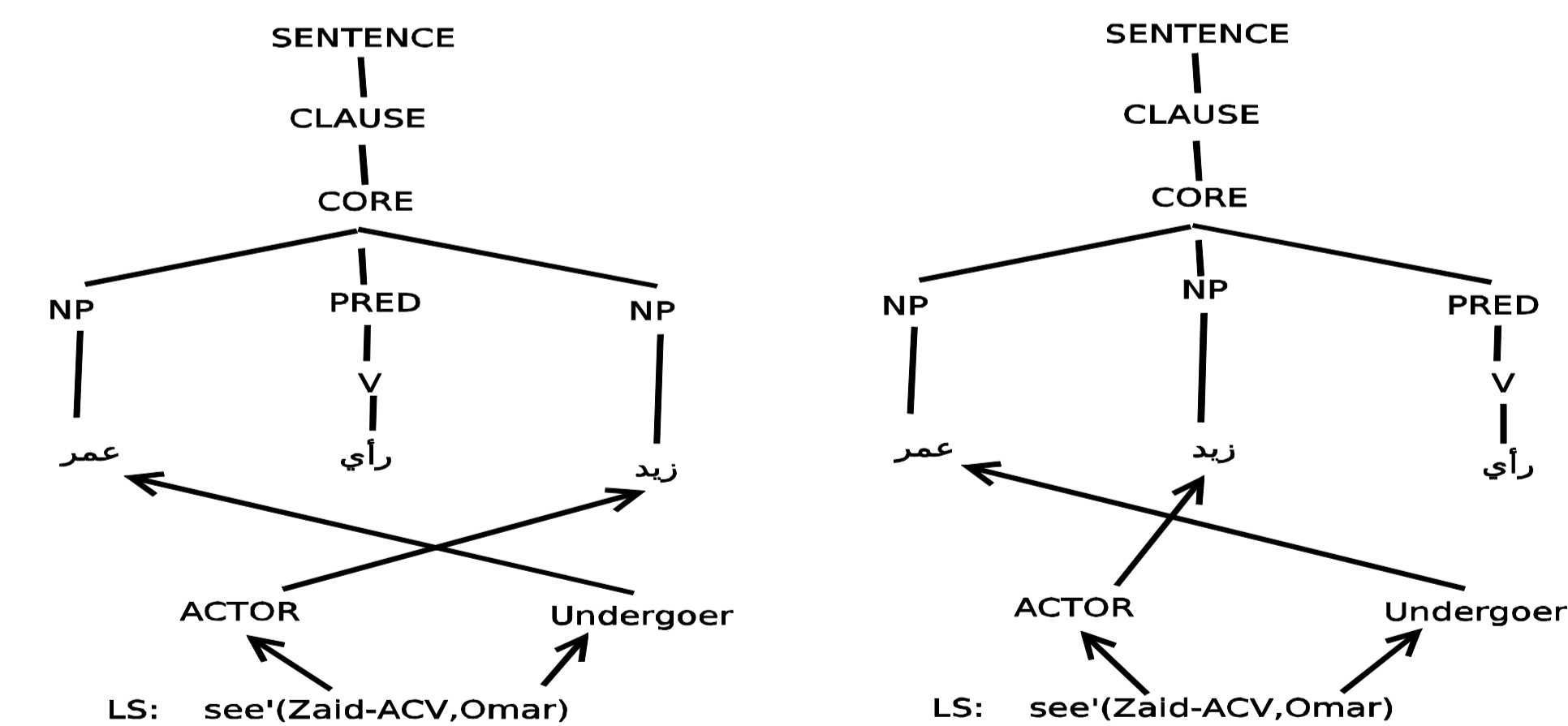


Figure 1: Arabic sentence types; verb subject object or subject verb object

- ▶ Verb Noun Noun
- ▶ Noun Verb Noun
- ▶ Arabic has SVO, VSO, VOS and OVS orders in sentences.
- ▶ No Copula verb in Arabic; “to be” or “to have”

## An Interlingua model for Arabic MT

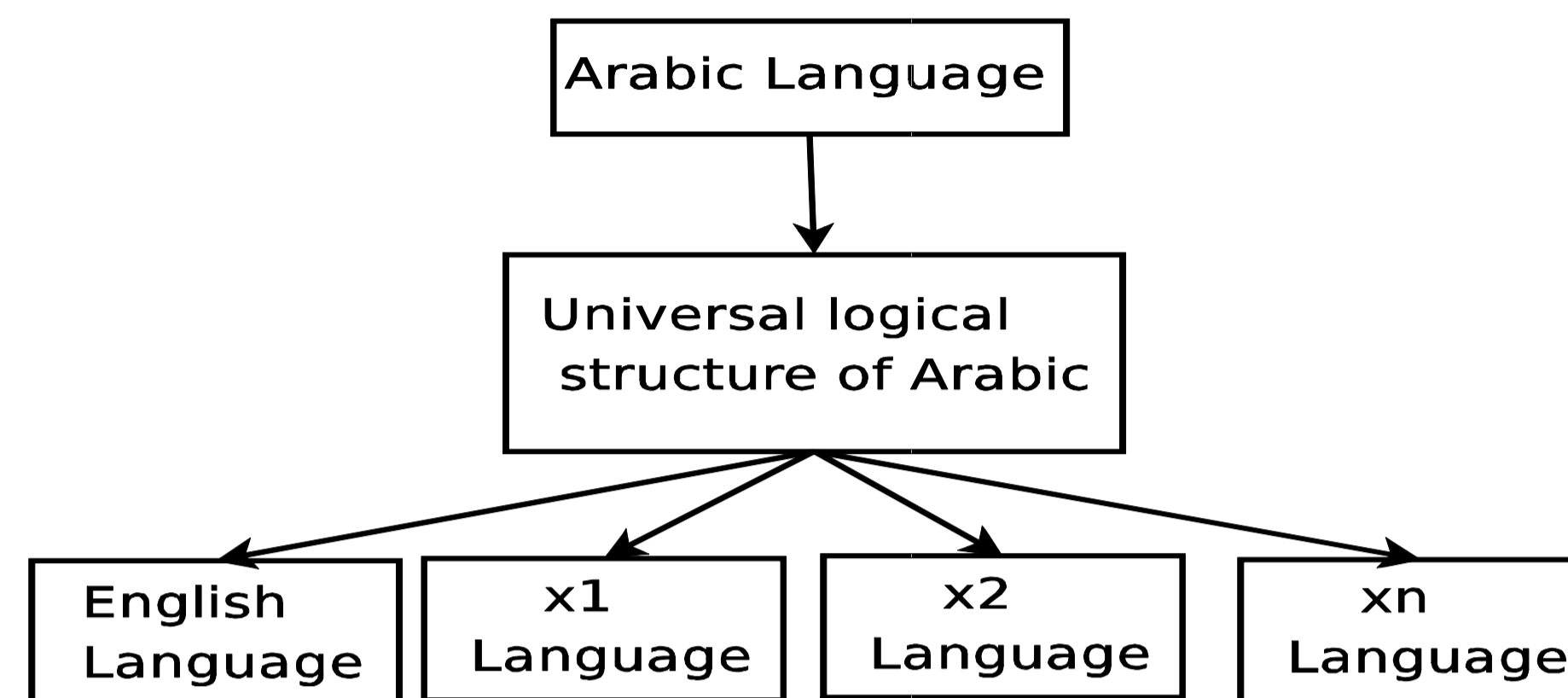


Figure 2: An Interlingua model of Arabic MT

- ▶ This model represents a more flexible way for multiple language generation.

## Role and Reference Grammar

- ▶ RRG describes mainly a sentence of a specific language in terms of:
  - ▶ logical structure
  - ▶ grammatical procedures
- ▶ We adopt RRG as a base for multi-languages translator system.

## Syntactic Representation

Linking Algorithm

## Semantic Representation

Figure 3: Layout of Role and Reference Grammar

## The UniArab System

- ▶ UniArab is a proof-of-concept system supporting the fundamental aspects of Arabic, such as the parts of speech, agreement and tenses.
- ▶ UniArab is based on the linking algorithm of RRG (syntax to semantics and vice versa).

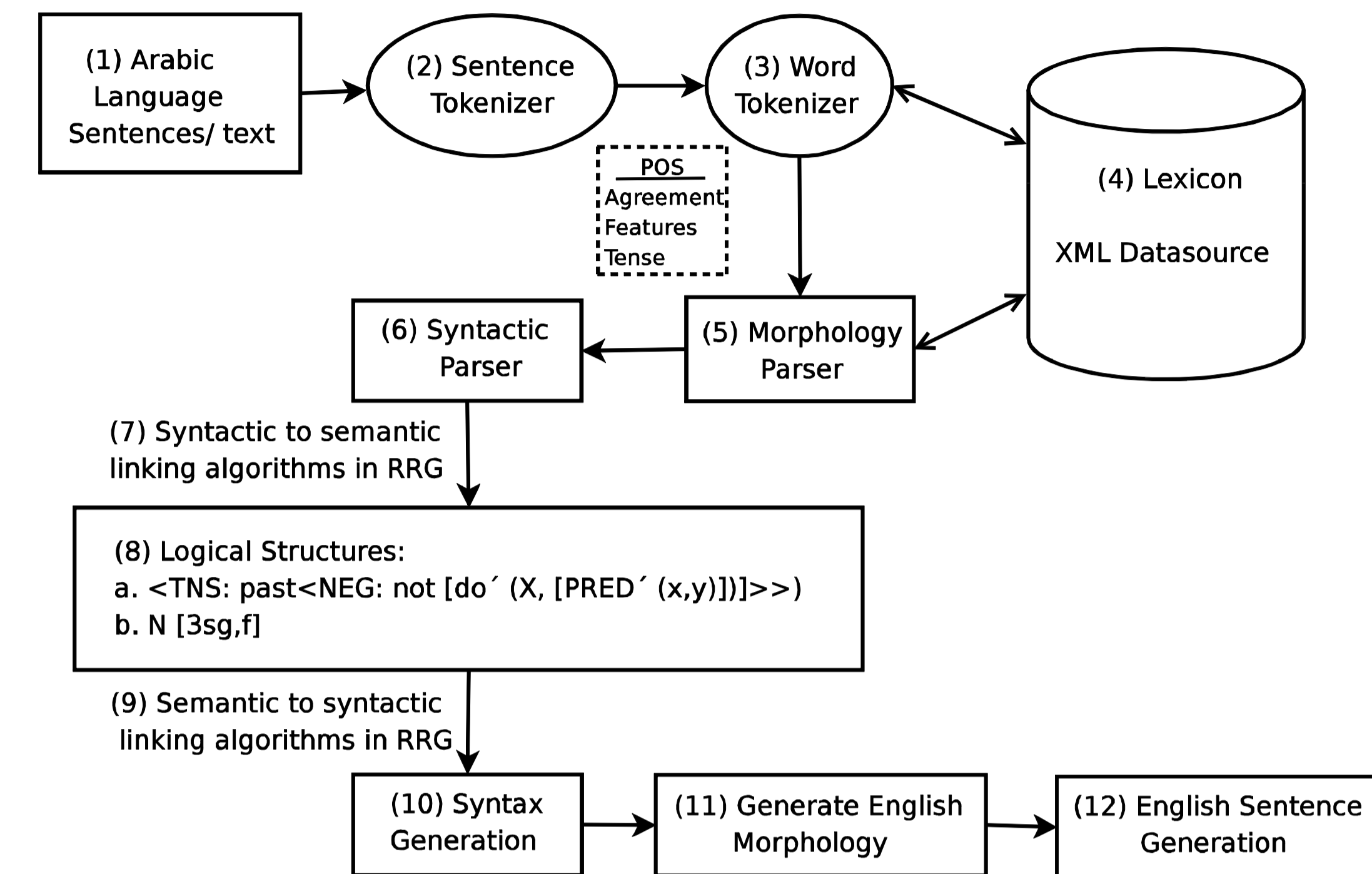


Figure 4: The conceptual architecture of the UniArab system

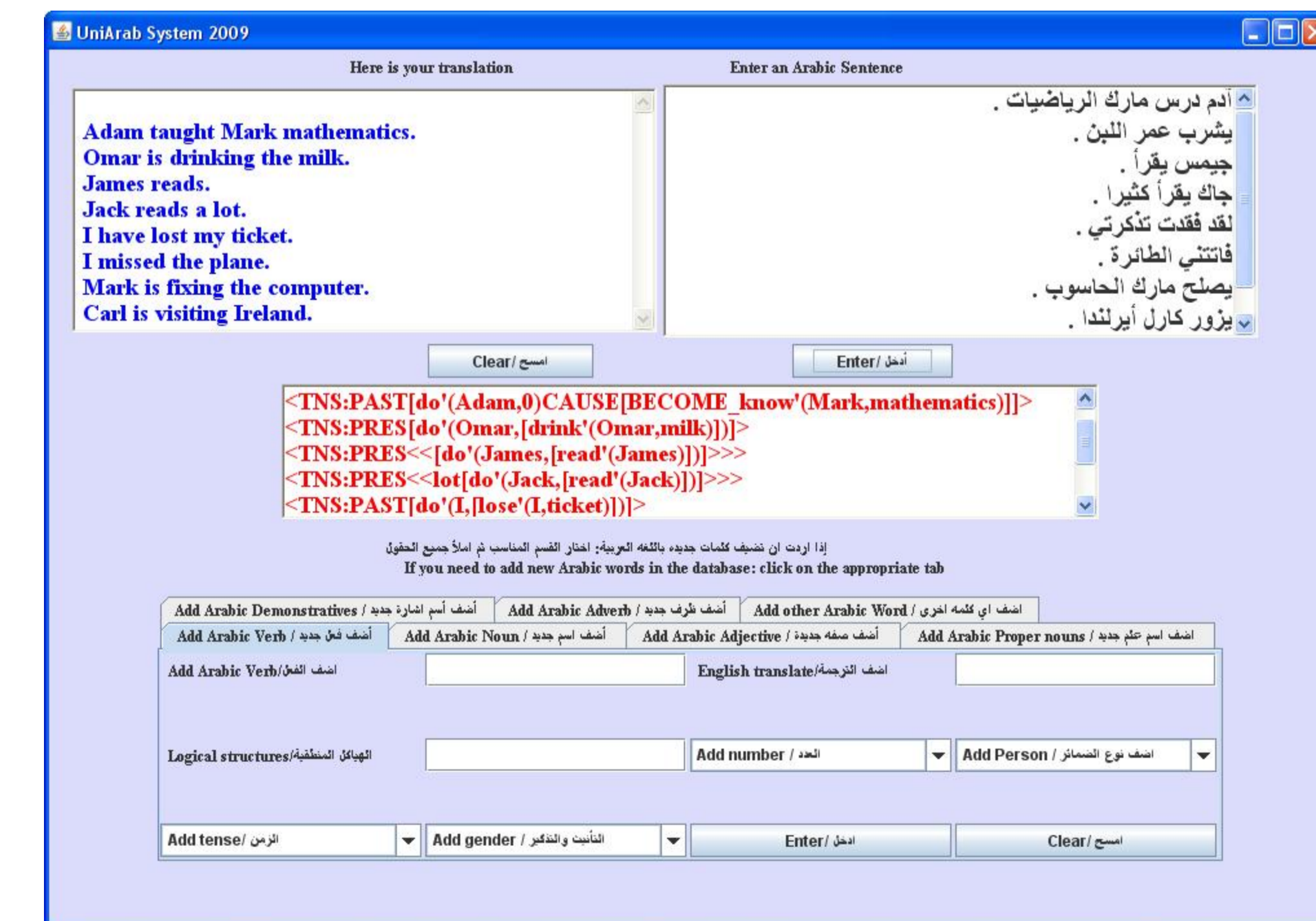


Figure 5: UniArab's GUI

## References

Hutchins, W. J. 2003. *Machine translation: General overview*. Oxford: Oxford University Press.

Izwaini, S. 2006. *Problems of Arabic machine translation: evaluation of three systems*. The British Computer Society (BSC), London, pages 118–148.

Ryding, K. C. 2005. *A Reference Grammar of Modern Standard Arabic*. Cambridge University Press.

Salem, Y., Hensman, A., and Nolan, B., 2008. Implementing Arabic-to-English Machine Translation using the Role and Reference Grammar Linguistic Model, in *Proceedings of the Eighth Annual International Conference on Information Technology and Telecommunication (IT&T 2008)*, Galway, Ireland, October 2008.

Salem, Y., Hensman, A., and Nolan, B., 2008. *Towards Arabic to English machine translation*. ITB Journal May 2008 Issue 17: accessible at: <http://www.itb.ie/site/researchinnovation/itbjournal.htm>

Van Valin, R. 2007. *The Role and Reference Grammar analysis of three place predicates*. *CEEOL Contemporary Linguistics*, 63:31–63.

Van Valin, R. and LaPolla, R., 1997. *Syntax: Structure, Meaning, and Function*. Cambridge University Press.

## For Further Information

More information on Yasser's MSc project and his research can be obtained at <http://informatics.itbresearch.ie/~ysalem/> Or email [yasser.salem@itb.ie](mailto:yasser.salem@itb.ie)