

## ARABIC LEXICON

Arabic Lexicon is a set of software tools able to generate in an accurate and quasi-exhaustive manner all the grammatical forms that common nouns, adjectives and verbs can take in Arabic. The breakthrough lies in the reversal of the traditional root-and-pattern Semitic model into pattern-and-root.

## PRESENTATION

In most languages, common nouns, adjectives and verbs can take very various forms in sentences, depending on the grammatical rules of the language. This is especially true in the Arabic language, where a single root of three consonants can generate hundreds of different forms.

While traditional dictionaries cover only a small fraction of the whole range of forms found in texts, our technology has been used to generate a database of 65 000 entries with their 6 millions of forms, covering more than 98 % of the forms found in any sort of text (literature, newspaper articles etc.), the remaining 2% including proper names.

Arabic Lexicon interfaces with Unitex, which is an open source corpus processing system for language processing, developed by Gaspard Monge Laboratory (LIGM UPEM).

Unitex Arabic has been presented to prestigious organizations, like Al-Ghazali Institute of La Grande Mosquée de Paris and L'Institut du Monde Arabe. It now can be used in a wide range of domains, like text editors, digitalization of printed documents, data mining in Arabic web contents and e-learning of Arabic.

Root	ك ت ب [k t b]	ك ت ب [k t b]
Pattern	ف ع ل [fa 'a la]	م ف ل [maf 'alatun]
Form	كتب [kataba]	مكتبة [maktabatun]
	He has written	Library

*Morphology of words in Arabic*  
© Alexis Neme 2014

Semitic languages - Arabic - Orthography - Grammar - Unitex

## COMPETITIVE ADVANTAGES

- Accuracy
- Exhaustivity
- Responsiveness

## DEVELOPMENT PHASE

- Development of Unitex software tools completed.
- Database of 65 000 entries with their 6 millions of forms available.

## INTELLECTUAL PROPERTY

Copyright

## PUBLICATIONS

Neme, Alexis, Laporte Éric (2013). "Pattern-and-root inflectional morphology: the Arabic broken plural". In Language Sciences Volume 40, November 2013, pages 221–250.

Neme, Alexis (2011). "A lexicon of Arabic verbs constructed on the basis of Semitic taxonomy and using finite-state transducers". In Proceedings of the International Workshop on Lexical Resources (WoLeR) at ESSLLI.

Neme, Alexis. Why Microsoft Arabic Spell checker is ineffective ?

## APPLICATIONS

- Orthographic correction
- Automatic typing word completion
- E-reputation analysis on web sites
- E-learning of the Arabic language
- Digitalization of documents

## CONTACT

+33 (0)1 44 23 21 50

industriels@erganeo.com

Ref. project : 138