Compiling a Modern Corpus-based Arabic Collocation Dictionary

Sattar Izwaini, PhD
American University of Sharjah
sizwaini@aus.edu

Outline
- Traditional dictionaries
- New collocations
- Corpus
- Methodology
- Questions & Answers

Traditional Collocation Lexicons
- Collocation dictionaries:
  - fiqh al-lughah
  - al-mukhassas
  - naj'at al-ra'a'id

New Collocations
- group of airplanes "sirb taa'iraat"
- heavy fire "waabil min al rasas"a

Calque translations:
- money laundering (n. n.)
- search engine (n. n.)
- bilateral relations (n. adj.)
- highly inflammable (n. adj.)
- reserve the right (v. n.)
New Collocations

Other translations:
- Internet browsing (n. n.)
- run for elections (v. n.)
- abstain (v. n.)
- live transmission (adj. n.)

Corpus

- A corpus of classical collocation dictionaries & general use dictionaries (7 dictionaries)
- Three collocation dictionaries: fiqh al-lughah, al-mukhassas, naj’at al-raa’id
- Four general use dictionaries: lisan al-arab, al-sihah, al-kamil, al-qamoos

Methodology

- Lexical combinations identified
- Extracted in this corpus as well as in other corpora that are available online, e.g. ArabiCorpus: http://arabiccorpus.byu.edu/
- Leeds Corpus of Arabic
  http://corpus.leeds.ac.uk/query-ar.html
- Archaic combinations are removed.
- Combinations are recorded as entries

Corpus

- Two corpora of modern Arabic of about 3 million words
- 153 journalistic articles, part of the Essex Arabic Summaries Corpus (EASC) compiled by Mahmoud El-Haj
- 6000 news articles of alkhaleej & alwatan newspapers compiled by Mourad Abbas
- Diverse topics and areas: politics, economy, health, religion, sports, finance, education, news science & technology, and art & music
Methodology

- Entries: nouns, verbs, and adjectives
- Nouns are cited along with their noun, verb and adjective collocates
- Example:

Example:

- منح، اتخاذ، اعتماد، انتهاز، اقتباس، تحسين، استغلال، ضماع،
- اضاعة، توبيت
- عمل، العمر

Thank You
Any questions?