

Compiling a Modern Corpus-based Arabic Collocation Dictionary

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

Outline

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

Traditional Collocation Lexicons

- *Collocation dictionaries:*
- *fiqh al-lughah* فقه اللغة
- *al-mukhassas* المخصص
- *naj'at al-raa'id*
نجعة الرائد وشرعة الوارد في المترادف والمتوارد
- General classical Arabic dictionaries

New Collocations

- group of airplanes *sirb taa'iraat* سرب طائرات
- heavy fire *waabil min alrasaas* وابل من الرصاص
- noise of a tank *hadeer al-dabbabah* هدير دبابة
- to conduct negotiations/investigation
يجري مفاوضات/تحقيقاً

Calque translations:

- *money laundering* غسل الأموال (n. n.)
- *search engine* ماكينة بحث (n. n.)
- *bilateral relations* علاقات ثنائية (n. adj.)
- *highly inflammable* سريع الاشتعال (n. adj.)
- *reserve the right* يحتفظ بالحق (v. n.)

New Collocations

Other translations:

- *Internet browsing* تصفح الانترنت (n. n.)
- *run for elections* يخوض الانتخابات (v. n.)
- *abstain* يمتنع عن التصويت (v. n.)
- *live transmission* نقل مباشر (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*

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

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

Methodology

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

- تسنح، تحين، تواتي، تضيع، تفوت؛
- يمنح، يتيح، يغتنم، ينتهز، يقتنص، يتحين، يستغل، يضيع، يفوت؛
- سانحة، موثية، متاحة، ممكنة، مناسبة، ضئيلة، ثمينة، عظيمة، ذهبية، طيبة، تاريخية؛ أقرب، أفضل، أحسن؛
- منح، اناحة، اغتنام، انتهاز، اقتناص، تحين، استغلال، ضياع، اضاءة، تفويت؛
- عمل، العمر.

Thank You
Any questions?