Arabic Profile: CEFR for Arabic
— a learner corpus approach

Björn Norrbom

Department of Language Testing, m National Center for Assessment in Higher Education (NCA), Saudi Arabia
• Established in 2000, Semi-governmental

• Mission: To provide comprehensive and integrated solutions that scientifically measure and evaluate knowledge, skills and aptitude with the purpose of achieving fairness, maintaining quality and satisfying development needs.

• Approximately 1.5 million test takers/year
Background 1/3

General

• CEFR has moved beyond Europe and European languages
  - available in 39 languages, including five non-European
    (http://www.coe.int/t/dg4/linguistic/CADRE1_EN.asp)

• Arabic translation of CEFR in 2008 (Dar Elias 2008)

• CEFR gaining ground in Arab world, particularly in the Gulf, and particularly for English (HCT/Cambridge ESOL 2011, Brennan 2013)

• General Progress for L1 & L2 Arabic research in past two decades, including corpora (Buckwalter & Parkinson 2011) although still under-researched

• However, limited progress for CEFR and L2 Arabic for both teaching and testing purposes.
The need for Reference Level Descriptors (RLDs)

• Very difficult to produce tests directly from the CEFR descriptors (Alderson et. al. 2004)

• “... an inventory of the linguistic material necessary to implement the [CEFR] competences” (Council of Europe 2005a:5)
Background 3/3

Ways of Developing Reference Level Descriptors (RLDs) – Examples from English

**Expert Judgment**

“T-Series” (van Ek & Trim)

**Combined approach**

Core Inventory (BC/EAQUALS)

**Learner-based corpus**

English Profile (UCLES/CUP)

---

A1-B1

A1-C1

A1-C2
Purposes and End-Products

Research Topics

• Order of Acquisition *(Bachman, personal communication, Carlsen 2012)*

• Different types of Arabic (MSA, colloquial, Quranic) and the CEFR. *(ACTFL 2012; Barnasek 2008; Buckwalter & Parkinson 2011; Younes 2013)*

• Comparing methods of RLD creation *(Alderson 2007)*

• CEFR and supporting materials in Arabic translation *(Takala, personal communication)*

• Arabic morphology and corpora *(Buckwalter & Parkinson 2011)*
Methodology:
- The Three Phases

“The critical component in a learner corpus is the validated leveling of learner data” (B. Knight, personal communication, August 8, 2013).

**Phase 1:** Provisional Reference Level Descriptors (RLDs)

**Phase 2:** Calibrating tests to the CEFR

**Phase 3:** Test application, learner data collection and analysis, start building corpus
Methodology: Phase 1
Provisional Reference Level Descriptors (RLDs)
- What students should know

• RLDs created as outlined in Council of Europe (2005a, 2005b)
  - Literature review (course books, readers, Arabic L2 syllabi, tests related to the CEFR, ACTFL etc.) coupled with expert judgment.
• Few materials directly linked – Initial classification:

“beginner/elementary/pre-intermediate” (A1&A2),
“intermediate” (B1&B2),
“advanced” (C1&C2)
(North et. al. 2011)
Methodology: Phase 2
Calibrating tests to the CEFR

• No known L2 Arabic test clearly and explicitly linked to the CEFR covering all four language skills and with full documentation.
• Carefully linking the NCA:s Arabic test STAPSOL to the CEFR using RLDs from Phase 1.
• STAPSOL linkage in tandem with CEFR linkage of our English exam STEP; identify materials for translation.
• Data for the corpus mainly from written and spoken performances. However, learner profiles are uneven; a test covering all four skills needed for data validation.
Methodology: Phase 3

Test application, learner data collection and analysis, start building corpus - What students do know

- Data collection in- and outside Saudi Arabia
- Participant profile:
  - Multiple countries and first languages
  - Students of MSA, colloquial, and Quranic Arabic
- Data is analyzed through both corpus and expert judgment approaches.
- Beta-version of online profile
- All welcome to participate!
Methodology
– continuous development

Phase 1:
Expert judgments

Phase 2:
Test calibration

Phase 3:
Corpus building
Research topics by phase

Phase 1: Provisional Reference Level Descriptors (RLDs) and

Phase 2: Calibrating tests to the CEFR
  - Order of Acquisition
  - Different types of Arabic (MSA, colloquial, Quranic) and the CEFR.
  - CEFR and supporting materials in Arabic translation

Phase 3: Test application, learner data collection and analysis, start building corpus
  - Order of Acquisition
  - Different types of Arabic (MSA, colloquial, Quranic) and the CEFR.
  - Comparing methods of RLD creation
  - Arabic morphology and corpora
Possible items for discussion

• Universal nature of CEFR (or any framework)
• CEFR and diglossia/dialects

شكراً لحسن متابعتكم

b.norrbom@qiyas.org