



Exploratory Workshop Scheme

Scientific Review Group for the Humanities

ESF Exploratory Workshop on

Breaking the Grounds for an Arabic Etymological Dictionary

Oslo (Norway), June 21-24, 2013

Convened by:

**Stephan Guth^①, Catherine Pennacchio^②
and Lutz Edzard^③**

① IKOS, University of Oslo, NO

② CERMOM, INALCO, FR

③ IKOS, University of Oslo, NO

SCIENTIFIC REPORT

Co-sponsored by

UiO : **Institutt for kulturstudier og orientalske
språk**

Dept. of Culture Studies and Oriental Languages

Det humanistiske fakultet

Faculty of Humanities



1. Executive summary

The workshop was held at Oslo/Norway over 4 days (Friday: arrival & get together; Saturday & Sunday: 2 core days; Monday: meeting of some participants with K. DMITRIEV, and departure). Participation numbered 21 people from 12 countries (incl. ESF representative). 2 invitees (ALBOUCHIKHI and AYOUB) had excused themselves shortly before the meeting (on grounds of work overload and bad health, respectively); 1 invitee (MARX) had to leave earlier on Sunday (but was 'doubled' by a colleague working on the same project), another (DMITRIEV) could only be met outside the core sessions on Monday morning at the hotel.

The Workshop had been prepared, following some initial e-mail circulars, by a Website (<<http://folk.uio.no/guthst/EtymArab/>>) to which relevant material and practical information had been uploaded. The participants had also been provided with the URL and log-in credentials for a Wikipedia used by Convenor1 (GUTH) for teaching and as a field of experimentation for entering etymological-cultural data.

The Workshop was designed mainly for discussion and brainstorming, with only 3 opening presentations and the remaining course of events only roughly structured, allowing for flexibility, deepening, and instant presentations by participants whenever this seemed to be appropriate. The final agenda was prepared/adjusted to new givens in a meeting of the 3 convenors on Friday morning, June 21, as well as in the breaks after the individual sessions.

The 'get together' meeting at Friday night was attended, though announced as 'informal', by most of those who by then already had arrived. During dinner, people could introduce themselves to each other, and practical information could be exchanged.

All participants from outside Norway were accommodated in one hotel in downtown Oslo. The hotel, as also the workshop venues, was easily reachable via public transport (tube). On arrival at the hotel, the participants received a pre-paid ticket, valid on all public transport for the period of their stay. 1 co-convenor stayed at the hotel herself and functioned as a guide for the group.

Being held over a weekend, with Norwegian summer holidays just having started, the workshop took place without any disturbances in a quiet atmosphere. The group was seated around a table so that everybody could see everybody else (as well as the screen). The venues— a meeting room on the 12th floor of the Faculty of Humanities building with view over Oslo (Saturday), and at the HL-senteret on Oslo's museum peninsula, with possibility of visiting the Viking Ship museum during lunch time or have a swim in the nearby fjord (Sunday)— were chosen to create an inspiring atmosphere and allow for variation. Movement from the hotel to the venues as well as to the places where dinner was held contributed to liveliness, prevented routine, allowed for sightseeing alongside the workshop and in this way provided for welcome distraction and motion after long and intense working sessions.

The general atmosphere during the event was characterized by the participants' engagement and the critically constructive and positive discussions that continued also into the coffee and lunch breaks and during dinner.

The presentation by the ESF representative gave many useful informations in a concise manner. The fact that the representative herself had a background in a related field made it possible for her to enrich discussions as if a regular participant.

The main objectives of the workshop—brainstorming and discussion of the project of an etymological dictionary of Arabic—can be said to have been fully met. Through the workshop, an informal network could be established that will be useful during the next steps (see below, section 3).

2. Scientific content of the event

The workshop's core sessions started on Saturday morning with a short welcome address by Convenor1 (GUTH) which also explained the objective of the programme's open structure (brain-storming, discussion) and encouraged participants to present own projects whenever they felt that would be appropriate (done so by EDZARD, BRAARVIG and LØKEN, JOCKERS, KILTZ, PIETRUSCHKA, THOMANN – see below). Convenor1 also pointed to the possible consequences of the unfortunate absence of the invitee who according to the preliminary programme should have been the first speaker (ALBOUCHIKHI) with a report on the newly launched project of a *Historical Dictionary of Arabic* in Qatar.

After the presentation of the ESF and its activities by the ESF representative (H. RUUS, Copenhagen), the morning session served the purpose of giving some examples of how entries in the *EtymArab* dictionary could or should look like, in order to initiate a broader discussion about the "ideal" *EtymArab*.

Convenor 2 (PENNACCHIO) first introduced the Qur'anic database she had prepared as her PhD, then brought three examples from Qur'anic words she had worked on, suggesting a typical *EtymArab* entry to be composed of the lemma in transliteration, grammatical information, a/the root, the word's semantic history with dated attestations and references, followed by etymological information (in the narrower sense of the word, i.e., Semitic root, foreign provenience, calquing, etc.).

The presentation functioned very well as an entrance into the subject matter, as could be seen from the fact that in the ensuing discussion a number of crucial practical questions were already raised: Should all etymological information that can be found in previous studies and seems to be relevant, be cross-checked before being entered into the dictionary? What kind of sources can/should be used? Should indigenous Arab lexicography also be considered? Which "Arabic" are we speaking about? Which transliteration system should be used? How is a common taxonomical framework to be established? Valuable suggestions were made, already at this point, for dealing with these challenges.

Given the fact that there is still no etymological dictionary of the Arabic language that could serve as a point of reference for the present *EtymArab* project, Steven FASSBERG's report about an endeavour that started in 1959 and has since then gone through a number of processes and developments—the *Historical Dictionary of the Hebrew Language*—came with expertise that helped to estimate the relation between possible/desirable options and

their feasibility, depending on the amount of time and manpower needed to achieve comparable results. Other important questions raised during the discussion that followed FASSBERG's presentation included the arrangement of the material and the nature, composition and size of the corpus of source texts that are used for attestation.

Among many other issues and details, Peyman MIKAILI's presentation raised the question of user-orientation (conciseness vs. comprehensiveness of entries) and, based on his own theory of 'root clusters' (bi- and tri-consonantal roots + pre- and suffixes), triggered a (controversial) discussion on (whether, and if so) how to incorporate theories of glottogony into the dictionary. MIKAILI also suggested an entry structure similar to PENNACCHIO's (Lemma, followed by Derivatives • [if applicable:] the word's proto-Semitic ancestor as reconstructed from • the Semitic cognates • Afro-Asiatic cognates • Broad etymological discussion, accompanied by the corresponding references and suggestions for further reading). In connection with MIKAILI's detailed demonstration it was also suggested that the ideal *EtymArab* dictionary should differentiate between various levels of certainty about the correctness of a given etymology ('secured', 'highly probable', 'doubtful', etc.), a suggestion that opened the way for a discussion of the kind and number of categories that should be marked/tagged and in this way be searchable. As explained by M. BÜCHLER, one will have to find one's way between a possible underspecification and, on the other hand, overspecification. The latter might be a choice nevertheless, bearing in mind possible extension(s) of the database in the future (manpower etc provided...). Regarding the nature of mark-up categories, a marking of the source languages, tribe and place names etc. was also discussed.

During the discussions, a number of issues already began to crystallize as majority consensus: Not roots, but individual *words* should be the lemmata; words should be given in *pausa*; for other languages than Arabic only *transliteration* should be used; the suggestion of a "1000 lemmata" version as a starting point seemed to be sensible; the entries should definitely reflect *historical* and *cultural conditions* and processes. In this connection, it was also suggested to provide links from place names and, as the case may be, tribal genealogy, given through the textual attestations, to maps (interface allowing for connection with geography software); it would, among other things, also be desirable to include geographical information based on the movement of the various tribes in order to be able to create maps showing the usage of specific, vernacular words (where appropriate and feasible).

After lunch, a paper presented by Convenor3 (EDZARD) opened the first afternoon session, widening the scope of the morning session's discussions by a contribution on "Inner-Semitic Loans/lexical doublets vs. genetically related cognates". In the discussion also the problem of pseudo-borrowings was raised.

The dynamic of the discussion then suggested a direct continuation with what in the preliminary programme was planned as Session no. 3 ("Choice and grouping of entries"). Discussions here followed Convenor1 (S. GUTH)'s presentation ("A wikipedia-like online dictionary?"), which also introduced the group, as far as information was available to himself, to the work of the Qatar team on a *Historical Dictionary of the Arabic Language* and made Muḥ. al-Dabbāgh's suggestion for the arrangement of lexicographical material available to the partic-

ipants. Discussions in this session focussed mainly the question of the criteria that should inform the choice of entries in a "1000 lemmata" pilot project. It was agreed that the choice should be representative of Arab cultural history and appealing to a larger general public. The choice could be a blending that takes wordlists like the Swadesh list, Jeffery's (1938) list, material provided in Ali/Leaman's "Islam—the Key Concepts", the thematic lists in Buckwalter/Parkinson's *Frequency Dictionary*, and others (G. Bohas, E. Benveniste, Newman's *Arabic English thematic lexicon*, Arabic word net, *al-Mawrid*, ...) as its starting point. A certain percentage could also be selected according to frequency considerations (Buckwalter/Parkinson) and with the help of a random generator.

The two sessions on Sunday were dedicated to the discussion of the question of corpora that had remained open from the day before, then proceeded to technical issues as well as fund raising.

With regard to the corpus, the group agreed that, for the earliest attestations, one should refer to the Qur'ân (material easily available, cf. PENNACCHIO's database, the *Glossarium Coranicum* project, etc.) and pre-Islamic poetry (attestations expected to be accessible via data exchange with K. DMITRIEV's project of a Dictionary of Early Islamic Poetry). The most convenient way of providing further attestations will probably be through linking up to the Historical Dictionary of the Qatar group. The latter project being in its upstart phase only, however, *EtymArab* may start with a minimal corpus in its trial stage first, and try to connect to the Historical Dictionary later (as soon as it begins to take a shape that promises good results for *EtymArab* purposes). In the meantime, *EtymArab* could restrict itself to a limited corpus of (an average of) 1-2 key texts per century. It was suggested that the choice also include letters (easy to localize in time and place!) and scientific texts (interesting from the perspective of cultural transfer). Reliable editions (or editions that only need to be cross-checked with standard ones) are available already in large numbers in digitalized and searchable form, both on CD and at various places in the Internet (e.g., *Alpheios* <alpheios.org> [presented by U. PIETRUSCHKA], *Wikisource* <wikisource.org>, *al-Warrâq* <www.alwaraq.net>, *arabiCorpus* <arabicorpus.byu.edu/>, *al-Miškât* <almeshkat.net/>, *al-Waqfiyya* <waqfeya.com/>, to a limited extent also in the *Perseus Digital Library* <www.perseus.tufts.edu>, in the Digital Averroes Research Environment [DARE] <http://dare.uni-koeln.de/>, as well as in the *Bibliotheca Polyglotta* <www2.hf.uio.no/polyglotta>). The latter was presented by J. BRAARVIG and H. LØKEN also as a tool that originally was developed for a sentence-by-sentence and side-by-side display of parallel corpora texts but might serve the *EtymArab* project, too. It was discussed also as a tool used in teaching where students can get credit points for entering relevant data—a point that seemed particularly interesting with regard to the question of manpower and an efficient use of existing structures that at the same time has a high pedagogical value. For this purpose, also the tools provided by *E-portfolio* could be used since it would allow to issue certificates/attestations about the work done by students entering data into the Dictionary.

Technical issues such as a possible exchange with, or linking-up to, other databases were raised especially after J. THOMANN's presentation of his "donation" to the *EtymArab* project: an updated concordance of lemma IDs (that has already proven to work well in facilitating data exchange between the *Arabic Papyrology Database*, APD <www.ori.uzh.ch/apd> and the *Trismegistos* platform <www.trismegistos.org/>). The discussion also included the question of compatibility of fonts and software, tagging in XML (cf. the Text Encoding Initiative <www.tei-c.org/>, latest version P5) vs. the use of "simple" tables, the organisation of the material in an XML-database (e.g. EXIST) if XML is to be used, addressing/managing relational databases via MySQL, PostgreSQL (geographical information).

In the remaining part of the Sunday afternoon session, the participants were asked about their own availability for, and possible engagement in, the project (see below, section 3), and the question of manpower and financial resources was discussed. The brainstorming about fund-raising yielded suggestions like the *Qatar Foundation*, names of colleagues who have been able to gather expertise about fundraising from Gulf countries, Norwegian/Scandinavian institutions, the *Swiss Academy of Humanities and Social Sciences* (SAGW/ASSH) or the German academies of sciences, the *German Research Foundation* (DFG), the *Gerda Henkel* and the *Volkswagen* Foundations, *HERA* (joint research programmes involving participants from at least 3 European countries), and others. Key words that seem to promise successful applications these days include 'social interaction', 'world heritage', 'multidirectionality', 'launching', 'Islam', 'digitalisation'.

In an informal session on Monday morning, K. DMITRIEV presented his project of an *Analytical Database of Early and Classical Arabic Poetry* (ADECAP) to a smaller group of the workshop participants. The project, which is still in its upstart phase, has received good funding and can therefore serve as an example of what can be achieved with a team consisting of 1 project coordinator, 1 IT specialist, 3 PhDs and 3 postdoctoral candidates working together over a limited period of time. As in the *APD* presented by THOMANN, in *ADECAP* too each lemma has its own ID number, which facilitates easy addressing and exchange with other data bases. *ADECAP* mentions an item's root (or virtual root), assigns it (via tagging in XML) to one or several semantic categories (e.g., Geometry, Astronomy, etc.), and lists textual attestations (which can be hidden on mouse click) as well as bibliographical references (the latter being managed via *zotero* <www.zotero.org>). In the discussion, additional aspects of transliteration (e.g., the use of *ǎ ĭ ů*), software, user-friendliness, data exchange, sources (incl. existing dictionaries), text corpora (e.g., V. Polosin's *Slovar' počtov plemeni 'Abs*, or the Library of Arabic Literature <www.libraryofarabicliterature.org>), and funding were treated.

3. Assessment of the results, contribution to the future direction of the field, outcome

Perhaps the major achievement of the workshop was to have confirmed the previous assumption that the *EtymArab* project, as a "1000 items" pilot version, appears to be feasible within a period of a few years. Availability and accessibility of the necessary sources and references; the possibility of linking up to related ongoing projects; the certainty of getting assistance for solving technical problems; the establishment of an informal network; the existence of an interested general public as well as a high number of specialists who are eager to contribute—all these facts are encouraging enough to make the Convenors draw the project farther up to the next level(s).

These will include

- the establishment of an entry template;
- the selection of the 1000 lemmata for the *EtymArab* pilot version (Convenor1 will suggest a choice of entries, assembled from various thematic word lists, and ask the group, i.e. the informal network, for their opinion);
- (as soon as lemma list is established) the selection of a basic corpus of source texts to be used for attestation;
- (meanwhile) the continued entering of existing etymological data from earlier research, ‘translating’ own previous work (GUTH, PENNACCHIO, MIKAILI, ...) into the template structure, incorporate already received/incoming sample entries and material (e.g., EHRET, Afro-Asiatic root mergers);
- (if possible) enforced collaboration with the Qatar project in order to distribute research tasks and energy and prepare interlinking—the head of the project’s executive board, prof. Albouchikhi (who had been invited to the workshop but unfortunately was not able to attend) has already signaled the group’s readiness to do so (e-mail of July 15);
- the elaboration of a standard for transliteration (Convenor3);
- (in due time, i.e. as soon as entry template, lemma list, basic text corpus, and transliteration standard are ready) ask the network to produce some sample entries themselves;
- (meanwhile) try to raise funding for the employment of specialized full-time collaborators (PhD, postdoc, ...)
- (perhaps, for the purpose of testing) send out a worldwide call for entries via *ArabList*;
- (while working on the 1000 entries) ensure connectivity (via concordances of lemma IDs) with related projects wherever possible (already available: concordance for connecting to *APD* [THOMANN] and *Frequency Dictionary* [BUCKWALTER/PARKINSON]).

Within the informal network, BÜCHLER and LØKEN agreed to assist with IT related problems, GHERSETTI will provide expertise on pre-modern Arabic literary sources (mostly prose) and the literary tradition in general, JOCKERS will ensure collaboration with the *Arabic-Latin Glossary*, KILTZ have a look at anything Aramaic and provide access to/exchange with the *Glossarium Coranicum*, RETSÖ provide general assistance, WENINGER help wherever Ethiopic is

involved, and ZAMMIT contribute with his expertise in Arabic dialects. The network will of course also contribute to "spread the message" and invite other colleague to contribute.

Results of the Workshop and further proceedings will be communicated to a larger public first at the 32nd Meeting of Oriental Studies in Germany (*Deutscher Orientalistentag*), 23-27 Sept., 2013, then in regular intervals via Internet (*ArabList* etc.).

Exchange of students and faculty via Erasmus between Leipzig (PIETRUSCHKA), Venice (GHERSETTI), and Oslo (GUTH) will further contribute to making the project public and make the younger generations interested in it.

4. Final programme

By and large, the final programme was identical with the preliminary one. However, due to BOUCHIKHI's absence, his presentation was replaced through one by PENNACCHIO (Convenor2). In order to ensure continuity of discussion and a smoother and more logical transition, as well as the incorporation of spontaneous presentations, what was planned as "Session II" and "Session III" changed place in the final programme. Discussion with DMITRIEV took place outside the two core days, on Monday morning, in an informal group at the hotel.

In addition to the short presentations/suggestions that had been announced in the preliminary programme, 8 participants gave introductions into problems related to the Arabic/Semitic vocabulary and/or explained projects they currently are involved in as well as these projects' relevance with regard to the *EtymArab* project: (on Saturday afternoon) EDZARD about *Inner-Semitic loans...* ; (on Sunday) PIETRUSCHKA about *Alpheios.net*, BRAARVIG and LØKEN about *Bibliotheca Polyglotta*, KILTZ about *Glossarium Coranicum*, JOCKERS on *Latin-Arabic Glossary*, THOMANN on *Arabic Papyrology Database*; (on Monday morning) DMITRIEV on *Analytical Database of Early and Classical Arabic Poetry*

Friday, 21 June 2013

- 10am preparatory meeting of the convenors and 1 participant who had already arrived (informal)
- 7pm get together dinner at *Café Mistral*

Saturday, 22 June 2013 (at the University of Oslo)

- 9:00am Welcome by Convenor1 (Stephan Guth)
- 9:20am Presentation of the ESF (Hanne Ruus, Scientific Review Group for the Humanities)

Morning sessions (9:40am-12:30): Ongoing projects of Etymological Dictionaries

- 9:40am The Qur'ân Vocabulary Software: some Examples of Etymological notices (presentation by Catherine Pennacchio)
- 10:15am A Historical Dictionary of the Hebrew Language (Steven Fassberg)

- 11:00am An Etymological Dictionary of Arabic (Peyman Mikaili)
All of these immediately triggered discussions which continued into the
Afternoon sessions (2pm-6pm),
where the scope was widened by Convenor3's contribution on
- 2:00pm Inner-Semitic Loans/lexical doublets vs. genetically related cognates (Lutz Edzard).
In order not to interrupt ongoing fruitful discussion before lunch, Convenor1's presentation was postponed until after the afternoon coffee-break:
- 3:30pm A Wikipedia-Like Online Dictionary? (S. Guth)
The dynamic of the discussion then suggested a direct continuation with what in the preliminary programme was planned as Session no. 3 (*Choice and grouping of entries*).
Discussion: *How should an Arabic Etymological Dictionary look like?*

Sunday, 23 June 2013 (at HL-senteret, Bygdøy, Oslo)

Morning sessions (10:00am-12.30): The ideal and the feasible, incl. technical aspects

Open discussion, supplemented in-between by short presentations
Alpheios.net (U. Pietruschka)
Bibliotheca Polyglotta (J. Braarvig and H. Løken)
Glossarium Coranicum (D. Kiltz)

Afternoon sessions (2:00pm-4.30pm): The question of text corpora; Collaboration and follow-up

Assessment of sub-tasks, incl. M.A. / Ph.D. and other projects; assessment of manpower; where to go for funding?; forms of collaboration and contribution
Open discussion, supplemented by interventions on
Latin-Arabic Glossary (B. Jockers)
Arabic Papyrology Database (J. Thomann)

Monday, 24 June 2013

Morning session (at the hotel, 10am-11.30am)

- 10am Analytical Database of Early and Classical Arabic Poetry (K. Dmitriev)
Open discussion.

5. Final list of participants

Convenors: Stephan **GUTH**, Dept. of Culture Studies and Oriental Languages (IKOS), University of Oslo • Catherine **PENNACCHIO**, CERMOM INALCO, Paris • Lutz **EDZARD**, Dept. of Culture Studies and Oriental Languages (IKOS), University of Oslo

Participants: Jens Erland **BRAARVIG**, Bibliotheca polyglotta, IKOS, University of Oslo • Marco **BÜCHLER**, Institut für Informatik, Universität Leipzig • Kirill **DMITRIEV**, School of Modern Languages, University of St Andrews • Christopher **EHRET**, Department of History, UCLA • Orhan **ELMAZ**, School of Modern Languages, University of St Andrews • Steven E. **FASSBERG**, Caspar Levias Chair in Ancient Semitic Languages, Dpt. of Hebrew Language, HUJI, Jerusalem • Antonella **GHERSETTI**, Dipartimento di Studi sull' Asia e sull' Africa Mediterranea, Università Ca' Foscari, Venezia • Barbara **JOCKERS**, Dpt. of Philosophy, University of Würzburg • David **KILTZ**, Corpus Coranicum c/o Berlin-Brandenburgische Akademie der Wissenschaften, Berlin • Heidi **LØKEN**, IT services, Digital Media in Teaching and Research, Fac. of Humanities, University of Oslo • Michael **MARX**, Corpus Coranicum c/o Berlin-Brandenburgische Akademie der Wissenschaften, Berlin • Peyman **MIKAILI**, Iranian Academy of Sciences / Urmia University of Medical Sciences • Ute **PIETRUSCHKA**, Corpus der arabischen und syrischen Gnomologien / Orientalisches Institut, Martin-Luther-Universität Halle-Wittenberg • Jan **RETSÖ**, Institutionen för orientaliska och afrikanska språk, University of Göteborg • Johannes **THOMANN**, Orientalisches Seminar, University of Zürich • Stefan **WENINGER**, Centrum für Nah- und Mittelost-Studien (CNMS), Fachgebiet Semitistik, University of Marburg • Martin R. **ZAMMIT**, Oriental Studies/Inst. of Linguistics, University of Malta

ESF representative: Hanne **RUUS**, Department of Scandinavian Studies and Linguistics, Copenhagen University

6. Statistical information on participants

Name	First name	male female	age bracket junior mid-way senior ¹	country of origin
BRAARVIG	Jens Erland	m	sen	NO
BÜCHLER	Marco	m	jun	DE
DMITRIEV	Kirill	m	jun	UK
EDZARD	Lutz	m	mid	NO
EHRET	Christopher	m	sen	USA
ELMAZ	Orhan	m	jun	UK
FASSBERG	Steven E.	m	sen	IL
GHERSETTI	Antonella	f	mid	IT
GUTH	Stephan	m	mid	NO
JOCKERS	Barbara	f	jun	DE
KILTZ	David	m	mid	DE
LØKEN	Heidi	f	jun	NO
MARX	Michael	m	mid	DE
MIKAILI	Peyman	m	jun	Iran
PENNACCHIO	Catherine	f	jun	FR
PIETRUSCHKA	Ute	f	mid	DE
RETSÖ	Jan	m	sen	SE
THOMANN	Johannes	m	mid	CH
WENINGER	Stefan	m	mid	DE
ZAMMIT	Martin R.	m	mid	MT Malta
TOTALS (without ESF representative)		5 f 15 m	7 jun 9 mid 4 sen	1 CH • 6 DE • 1 FR • 1 IL • 1 Iran • 1 IT • 1 MT • 4 NO • 1 SE • 2 UK • 1 US (11 countries)

with ESF representative

RUUS	Hanne	f	sen	DK
TOTALS		6 f 15 m	7 jun 9 mid 5 sen	1 CH • 6 DE • 1 DK • 1 FR • 1 IL • 1 Iran • 1 IT • 1 MT • 4 NO • 1 SE • 2 UK • 1 US (12 countries)

¹ marks where in his/her *professional* life (imagined as consisting of three roughly equally long phases), a researcher has arrived.