

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИННОВАЦИИ И ТРАДИЦИИ
В АРАБИСТИКЕ
И ИСЛАМОВЕДЕНИИ

*Сборник статей в честь профессора
Олега Ивановича Редькина*



ИЗДАТЕЛЬСТВО САНКТ-ПЕТЕРБУРГСКОГО УНИВЕРСИТЕТА

УДК 811.411.21,821.411.21,297,921,004,009,34

ББК 1.81.2,86.38,84,67,63.3,74

И57

Рецензенты:

д-р ист. наук *Е. А. Резван* (Музей антропологии и этнографии им. Петра Великого (Кунсткамера) РАН); канд. ист. наук *Д. В. Ермаков* (Федеральное государственное бюджетное учреждение культуры «Государственный академический Мариинский театр»)

Рекомендовано к публикации

Научной комиссией в области востоковедения

Санкт-Петербургского государственного университета

И57 **Инновации и традиции в арабистике и исламоведении:** сб. статей в честь профессора Олега Ивановича Редькина / под ред. О. А. Берниковой, А. Д. Кныша. — СПб.: Изд-во С.-Петерб. ун-та, 2019. — 340 с.
ISBN 978-5-288-05919-3

Сборник статей подготовлен в честь юбилея профессора, заведующего кафедрой арабской филологии Санкт-Петербургского государственного университета О. И. Редькина. Широкий круг интересов юбиляра определил тематику разделов сборника: «Востоковедение», «Исламоведение», «Лингвистика и арабский язык», «Междисциплинарные исследования» и др. Отдельные разделы посвящены общим вопросам востоковедения и педагогики. Среди авторов статей — друзья и коллеги О. И. Редькина, ведущие арабисты и исламоведы, а также молодые ученые.

УДК 811.411.21,821.411.21,297,921,004,009,34

ББК 1.81.2,86.38,84,67,63.3,74

ISBN 978-5-288-05919-3

© Санкт-Петербургский
государственный университет, 2019

© Авторы, 2019

Содержание

Предисловие	5
Вступительное слово	
<i>Пиотровский М. Б.</i> Почерк арабиста.....	13
Олег Иванович Редькин (к 60-летию со дня рождения)	
Творческая биография	15
Список основных научных и учебно-методических работ профессора О. И. Редькина	27
Арабистика и исламоведение в СПбГУ	
<i>Юзеев А. Н.</i> Продолжение традиций петербургской школы арабистики (в связи с юбилеем О. И. Редькина).....	39
<i>Берникова О. А.</i> Арабская филология и исламоведение в СПбГУ: современные тенденции	47
Востоковедение	
<i>Карабекян С. Б.</i> Структура востоковедного объекта и методологические аспекты современных востоковедных исследований	61
<i>Маточкина А. И.</i> Роль практического востоковедения в формировании образа ислама и мусульман на рубеже XIX–XX вв....	76
Лингвистика и арабский язык	
<i>Белова А. Г.</i> Имя действия и деление неделимого	91
<i>Ахматшина Э. К., Хана Яфия Юсиф Джамиль, Мокрушина А. А.</i> Функционально-прагматический подход в изучении терминов семейного родства в арабском языке (на примере египетского и иракского диалектов).....	105
<i>Бобровников В. О.</i> Опыт историко-топонимического словаря аварских <i>нисб</i> из горного Дагестана XVII–XX вв.	113
<i>Телицин Н. Н.</i> Несколько замечаний к интерпретации <i>-slyq~šlyq + er-</i> в древнеуйгурском языке	134
Исламоведение	
<i>Сюкияйнен Л. Р.</i> Правовые проблемы взаимодействия государства и ислама в современной России	139

<i>Добронравин Н. А.</i> Квазирукописная исламская литература в Чаде: репертуар и графическое оформление.....	154
<i>Розов В. А.</i> Архаический <i>садж</i> в контексте иных форм экстатической речи.....	162
<i>Герасимов И. В.</i> Мир суданской <i>халуа</i> : терминология, смысл, назначение	172
<i>Базиленко И. В.</i> Лев Толстой и мусульманская гетеродоксия	180
История стран Ближнего Востока	
<i>Дьяков Н. Н.</i> Из опыта французской историографии истории арабов: Доминик Шеваллье (1928–2008)	196
<i>Ulanova D. V.</i> Legal Aspects of Marriage in the Ottoman Empire in the 15 th —17 th Centuries	207
Литературоведение, арабская литература и театр	
<i>Суворов М. Н.</i> Изучение современной арабской литературы в России: достижения и перспективы.....	216
<i>Соколов О. А.</i> Образ Крестовых походов в арабском театре конца XIX — начала XX в.....	227
<i>Улангина А. О.</i> Фантазия и реальность в романе Самуэля Шимона «Иракец в Париже»	238
Междисциплинарные исследования	
<i>Граничин О. Н.</i> Математические модели при анализе культурно-исторических и социальных явлений	246
<i>Lutkov V. V.</i> Physical Education and Sports: Examining a Modern Islamic Perspective.....	254
<i>Romanov M. G.</i> 1001 Morphological Patterns: Algorithmically Generated Materials for Teaching Classical Arabic	262
Вопросы образования	
<i>Knysh A. D.</i> Secularization or De-Secularization in Kazakhstan: A Case for Religious Literacy	275
<i>Маточкина А. И., Уланова Д. В.</i> Дистанционные образовательные технологии в программах подготовки специалистов с углубленным знанием истории и культуры ислама	309
<i>Ибрагимов И. Д.</i> К вопросу об обучении арабскому языку теологов....	316
<i>Ибрагимова М. М.</i> Трудности обучения арабскому языку в дистанционном формате.....	326
Об авторах	333

M. G. Romanov

University of Vienna

1001 Morphological Patterns: Algorithmically Generated Materials for Teaching Classical Arabic

Together with Anna Arkad'evna Dolinina (*raḥima-hā Allāh wa-a'ṭā-hā al-jannat*) who taught us classical Arabic and Yāfi'at Yūsif Jamīl Ḥanná who taught us modern standard Arabic, Oleg Ivanovich Red'kin gave us a perspective on Arabic through the thousand and one morphological patterns and drilled into us the way of thinking about the Arabic word through its morphology. Ever since, looking up a word in a dictionary is never quite complete without probing for additional meanings that may be implied by different morphological patterns. Writing an article to honor Oleg Ivanovich, it will only be logical to focus on the Arabic language and how we learn and teach it. This article presents a method in which a computer is used to generate thematic readers from a corpus of original texts. While developed primarily with classical Arabic in mind (more specifically, Ḥadīṭ), the approach is in fact universal and can be used for any genre and for any language.

Keywords: Arabic, Classical Arabic, computationally generated teaching materials, readers, Ḥadīṭ, corpus-based research.

Introduction¹

Even if we leave its dialects aside, Arabic is a difficult language to learn. One of my professors at the University of Michigan remarked that Arabic is hard only for the first ten years — after that it gets worse. Joke or not, this nonetheless is a profound observation. This is not to say, of course,

¹ The article uses a somewhat unconventional transliteration system that was developed to facilitate computational analysis and interoperability by allowing for automatic conversion between Arabic script and transliteration — something that none of the existing academic transliteration systems can do. It relies on one-to-one character representation, where every Arabic letter and short vowel is transcribed distinctively. Combining one-letter representations from American/British, German and French transliteration systems, the one that I propose should be easily recognizable to Arabists. There is only one new character: *t* for *tā' marbūṭat* — the least consistently transliterated letter, even among specialists. Two other somewhat unconventional uses include *ā* for dagger *alif*; and *á* for *alif maqṣūrat*. All attached conjunctions, prepositions, pronominal suffixes are separated with “-”.

that one's knowledge of Arabic diminishes after ten years of learning, yet what does happen is the dawning realization of how massive the Arabic written legacy is — geographically, chronologically, as well as in terms of the variety of subjects that it covers.

To give a taste of the scale, we can turn to the bio-bibliographical collection *Hadiyyat al-ʿarifīn* [1] that was written by Ismāʿīl Bāšā al-Baġdādī (d. 1338/1919 CE) [2] in the beginning of the 20th century. Continuing the famous *Kašf al-ẓunūn* of Ḥājjī Ḥalīfat (d. 1067/1656 CE) [3], Ismāʿīl Bāšā collected records for about 8,800 authors and listed about 40,000 works² (Carl Brockelmann in his *Geschichte der arabischen Litteratur* provides a similar number of records for both authors and works [5]). While written predominantly in Arabic (over 90 % of all works), geographically these works were produced throughout the Islamic world — from Spain in the West to India in the East. While we may call the bulk of this tradition “classical Arabic”, one might still wonder: would the Arabic of a Spaniard writing in al-Andalus be the same as the Arabic of a Persian writing somewhere in Ḥurāsān? Many senior colleagues shared their observations that if one works for a long time with texts written in the eastern Islamic lands and then has to read something that was written in its western counterparts, that “western” Arabic would look and feel quite different — and the other way around. And why would that not be the case? Figure 1 shows that a significant number of books were produced where Arabic was not the main spoken language, and the dominant languages of those lands also often belonged to different language families³. So far, unfortunately, our field has not produced studies that would analyze linguistic differences within *classical* Arabic texts written in different provinces of the Islamic world, but we can look for guidance and inspiration among corpus

² For my analysis of the *Hadiyyat al-ʿarifīn*, see: [4].

³ A fascinating example of coping with the difficulties of the Arabic language among the speakers of non-Semitic languages can be found in the North Caucasus. Already by the 17th century, local Muslims have had a highly-developed system of explanatory signs which were used to clarify ambiguous syntactic structures in Arabic. These explanatory signs were used, for example, to connect the verb, the subject, and the pronominal suffixes referring to the subject. It appears that these signs were used both in books and in correspondence, and were [re]copied meticulously together with the Arabic text. For a detailed study of this phenomenon, see: [6].

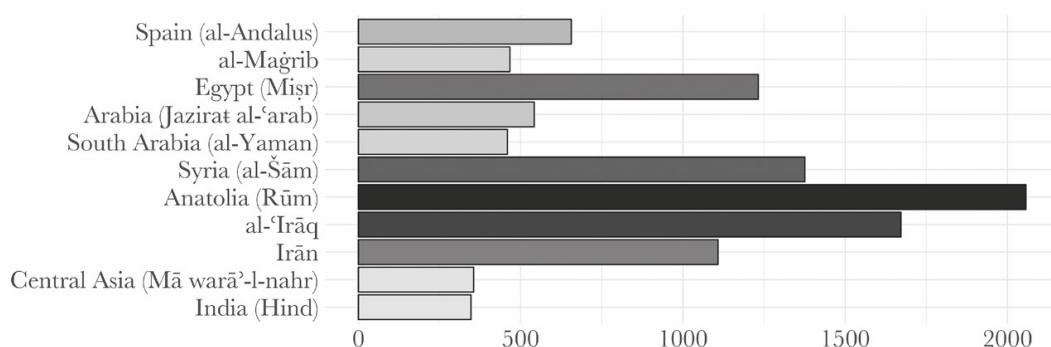


Figure 1. Regional contribution to the Arabic written tradition. Based on the *Hadiyyat al-ʿarifin* of Ismāʿīl Bāšā al-Bağdādī (d. 1338/1919 CE).

linguists who use different linguistic features — word choices, cohesion, complexity, error categories, etc. — for native language identification (NLI)⁴.

In addition to — or, as a consequence of — the chronological and spatial complexity, it is claimed that Arabic has one of the largest vocabularies. While doing a comparative count is methodologically extremely difficult (if not impossible), a corpus-based study by my colleagues argues that the average lifespan of Arabic words is significantly longer than, for example, the average lifespan of English words. The most conservative evaluation gives the average of 161 years for Arabic words versus the average of 88 years for English words⁵.

All of these issues inevitably find their way into the classroom. While most of us took great pleasure in advanced reading classes with our professors, we also often struggled with the overwhelming volume of new vocabulary; switching to a new text — or a new genre — made things even more complicated. Figure 2 shows a matrix of lexical overlap across the frequency lists that represent large thematic specimens of Arabic language. The most dramatic lexical difference is between *al-Kutub al-Sittat*, the Six Sunnī Collections of *ḥadīṭs*, and *al-Šark al-awsaṭ*, the modern newspaper: the frequency lists of these two sources share only 20 % of the most frequent 3,000 word forms (tokens). Even among the “classical” works the lexical distance is quite significant, with the percentage of shared vocabulary fluctuating between 38 % and 58 % (for

⁴ For the overview of the native language identification field, see: [7].

⁵ For more details, see (in print): [8].

100	54	54	38	44	50	37	41	20	<i>al-Kutub al-Sittat</i>
54	100	48	48	56	43	41	62	21	<i>Tafsīr al-Ṭabarī</i>
54	48	100	38	39	50	43	39	21	<i>Kitāb al-Aġānī</i>
38	48	38	100	58	36	49	57	25	<i>al-Futūḥāt al-Makkiyyat</i>
44	56	39	58	100	39	46	62	24	<i>Fatāwā Ibn Taymiyyat</i>
50	43	50	36	39	100	40	37	21	<i>Taʾrīḥ al-Islām</i>
37	41	43	49	46	40	100	48	40	<i>Mağallat al-Risālat</i>
41	62	39	57	62	37	48	100	26	<i>Tafsīr al-Mizān</i>
20	21	21	25	24	21	40	26	100	<i>al-Šarḥ al-Awsat</i>
<i>al-Kutub al-Sittat</i>	<i>Tafsīr al-Ṭabarī</i>	<i>Kitāb al-Aġānī</i>	<i>al-Futūḥāt al-Makkiyyat</i>	<i>Fatāwā Ibn Taymiyyat</i>	<i>Taʾrīḥ al-Islām</i>	<i>Mağallat al-Risālat</i>	<i>Tafsīr al-Mizān</i>	<i>al-Šarḥ al-Awsat</i>	

Figure 2. The matrix shows lexical overlap across the frequency lists (top 3,000 items, in %) that represent large thematic specimens of Arabic language. The specimens are arranged chronologically, starting with the earliest (right-top corner, 9th century) to the latest (20th century). **Texts compared:** *al-Kutub al-Sittat* (2,8 mln words), the 6 Sunnī collections of Ḥadīṭ (~9th century CE); *Tafsīr al-Ṭabarī* (or *Jāmiʿ al-bayān*, 3 mln words), a Qurʾān commentary by al-Ṭabarī (d. 310/922 CE); *Kitāb al-Aġānī* (1,5 mln words), a poetic anthology by Abū-l-Faraj al-Iṣbahānī (d. 356/967 CE); *al-Futūḥāt al-Makkiyyat* (1,7 mln words), an extensive Šūfī text by Ibn al-ʿArabī (d. 638/1240 CE); *Fatāwā Ibn Taymiyyat* (2,9 mln words), a collection of legal decisions and epistles of Ibn Taymiyyat (d. 728/1327 CE); *Taʾrīḥ al-Islām* (3,2 mln words), a biographical collection and chronicle by al-Ḍahabī (d. 748/1347 CE); *Majallat al-Risālat* (16 mln words), an early 20th-century Egyptian literary journal; *Tafsīr al-Mizān* (2,3 mln words), a modern Šīʿī Qurʾān commentary by al-Sayyid al-Ṭabāṭabāʾī (d. 1981 CE); and *al-Šarḥ al-Awsat* (2,5 mln words), a modern Arabic newspaper (collected by Tariq Yousef, Leipzig University, from <http://aawsat.com/>).

the interquartile range)⁶. Committed students usually overcome this difficulty by their sheer passion for the subject, but the introduction of excessive vocabulary may create an insurmountable obstacle to many truly capable students.

The increasing availability of electronic texts and computational methods of text analysis allow us to rethink how we teach difficult languages.⁷ We can identify the most frequent features within a corpus and focus our attention on them. For example, the top 100 most frequent lexical items constitute about 56% of the entire “vocabulary” of over 34,000 Prophetic sayings (*ḥadīṭ*) from the Six Sunnī Collections (approximately 2.8 mln words). Relying on such data, we can generate a frequency-based reader that will allow us to ensure that students focus on the most essential vocabulary items and features of texts expressed through them. With a paced increase in difficulty of texts and incremental expansion of vocabulary, students will be able to digest much larger volumes of text both in class and at home, and such an extended exposure will enable students to internalize the authentic language more efficiently.

Description of the method

What follows is the description of my computational method for generating frequency-based readers (here using a Ḥadīṭ reader as an example). The overall procedure is not particularly complicated and includes five main steps.

Step I

The Six Sunnī Ḥadīṭ collections were downloaded from <http://sunnah.com/> using the Wget tool [11]. Then, initial texts were converted from their initial HTML format into plain text table format (CSV), and normalized with scripts written in Python (<https://www.python.org/>).

⁶ The ratio of unique vocabulary items to the overall length of the text (in words) can serve a useful measure for identifying “easy” and “difficult” texts.

⁷ Linguists working in the area of “natural language processing” (NLP) developed methods that rely on wordnets, vector space models, and gazetteers of named entities to reduce the complexity of Arabic texts, making them more accessible to learners. In a nutshell, these methods automatically replace rare words and phrases with their more frequent counterparts used in similar contexts. For interesting examples of such studies, see: [9; 10].

In terms of computational linguistics and natural language processing (NLP), “normalization” refers to the disambiguation of spelling, formatting and encoding irregularities of Arabic text. The most important tasks include fixing encoding issues which arise when a typist used a Persian or Urdu keyboard layout to enter Arabic text.⁸ For example, several Arabic letters in most cases look exactly the same in all three languages, however, in the Unicode table they occupy different locations, and therefore, computationally, are different symbols. Having “Persian” letters in an Arabic text (and the other way around) leads to a number of problems during any form of computational analysis. “Complicated” letters which require normalization include: *kāf*; *yā’* (plus, inconsistencies in the usage of *yā’* and *alif maqsūrat* — both are usually replaced with *yā’*); *alif* (five different forms of *alif* are replaced with the “bare” *alif*); and medial and final *hamzat* (usually replaced with the standalone *hamzat*).

It is important to note that there are multiple ways of how specimens of other genres can be obtained and then pre-processed for a similar reader. Together with my colleagues Sarah Bowen Savant (Aga Khan University — ISMC [UK]) and Matthew Thomas Miller (University of Maryland — College Park [USA]), we have been building a machine-readable corpus of texts. Named *Open Islamicate*⁹ *Text Initiative* (Open ITI), our project strives to provide the essential textual infrastructure in Persian and Arabic for new forms of macro textual analysis and digital scholarship. OpenITI includes about 4,300 unique book titles (at the moment, predominantly in Arabic) by over 1,800 authors, which amounts to approximately 750 mln words (1, 46 bln words with all versions). Most texts have been collected from open-access online collections of high-quality digital reproductions of premodern texts. The corpus is available on Github[14] at: <https://github.com/OpenITI/>; a detailed description of the corpus can be found at <https://maximromanov.github.io/OpenITI/>. The corpus is still in the process of development, but is already used in several large-scale European projects (For our Open ITI manifesto, see: [15]).

⁸ For more details on normalization, see: [12].

⁹ Introduced by the University of Chicago historian Marshall Hodgson, the term “Islamicate” refers to phenomena Islamic and non-Islamic, religious and non-religious, that were produced or existed in the vast geographical region we refer to as the “Islamic world”. See, [13], pp. 3–71.

Step II

All vocabulary items from the corpus were collected and converted into a frequency list. This list was then inverted into a ranking list, where the most frequent item received rank 1, the second most frequent one — rank 2, the third — 3, and so on; items with the same frequency were assigned the same rank.

It should be noted that vocabulary items have not been parsed with a morphological analyser, so different forms (tokens) of the same word (lexeme) are treated separately (i. e., *qāla*, *qīla*, *qālat*, *fa-qāla*, etc. have their own frequencies and ranked separately). The main reason for not using the results of automatic morphological analysis is largely technical, since existing morphological analyzers are meant to work with modern standard Arabic and do not perform reliably well on classical Arabic. For example, Buckwalter Morphological Analyser, which has been tested with this corpus (using Perseus morphological services at Tufts University), returned no results for about 25 % of tokens, single results for another 25 %, and more than one result for the rest 50 %. Needless to say, such outcome is hardly useable for our purposes¹⁰. Although the lack of a proper morphological analyzer may seem to be a disadvantage, in fact there is a lot of value in using frequencies of word forms (tokens) rather than dictionary forms (lexemes). Most importantly, frequent forms of words will be given more often in the generated reading materials — such as, for example, very frequent singular masculine *qāla* vs. quite rare dual masculine *qālā*.

Step III

The average mean of ranking values was then calculated for each *ḥadīṭ* (the sum of all ranks of words in a *ḥadīṭ* divided by the number of words in the same *ḥadīṭ*). The resultant values then served as difficulty indices, where texts with the most frequent vocabulary would have the lowest average means, and vice versa. These indices were then used as sorting values that allowed rearranging all 34,000 *ḥadīṭs* by the difficulty of their vocabulary. The advantage of the average mean here is that even a single low-frequency lexical item increases the difficulty index of an

¹⁰ There have been new developments since 2015: existing tools have been updated and several new tools have appeared, but they are yet to be thoroughly tested for their usability for classical Arabic [16].

entire text, which is pushed down the list. In practical terms this means that students will be exposed first to the easiest texts.

This computational approach turned up several unplanned, but valuable effects. First, as the length of a text increases so does the probability of rarer lexical items — as a result, the easiest texts are also the shortest ones. This convenient outcome allows students to begin with the shortest texts and move gradually to the longer ones. Second, the most frequent vocabulary also tend to appear in the most frequent grammatical and syntactic structures. Third, the most frequent vocabulary also tend to appear in “massively transmitted” *ḥadīṭ*s¹¹ — so students also start reading *ḥadīṭ*s which are, one can argue, more important historically.

Step IV

Automatically ranked *ḥadīṭ*s were not immediately useable since the method grouped together items that are almost identical. Here manual revision was required to cull repetitions. However, this outcome — i.e. similar *ḥadīṭ*s were grouped together — also has educational value because students can be exposed to a variety of different ways of phrasing the same ideas.

4. *Ḥadīt Rank 7*

ID: muslim44.365670

حَدَّثَنَا مُحَمَّدُ بْنُ بَشَّارٍ حَدَّثَنَا مُحَمَّدُ بْنُ جَعْفَرٍ حَدَّثَنَا شُعْبَةُ عَنْ هِشَامِ بْنِ
زَيْدٍ سَمِعْتُ أَنَسَ بْنَ مَالِكٍ يَقُولُ مِثْلَ ذَلِكَ
صحيح مسلم، كتاب فضائل الصحابة رضى الله تعالى عنهم

Frequency report

2) بن (109,869);	20) محمد (15,288);	51) مالك (4,835);	67) شعبه (3,631);	121) جعفر (2,054);
3) عن (104,417);	38) يقول (6,047);	55) ذلك (4,519);	81) هشام (2,999);	161) بشار (1,550);
5) حدثنا (65,785);	46) سمعت (5,402);	61) انس (3,969);	83) زيد (2,791);	180) مثل (1,370);

Figure 3. An example of a *ḥadīt* from a frequency-based reader. From top to bottom: 1) the rank of the *ḥadīt*; 2) the unique ID of the *ḥadīt* in the corpus; 3) the text of the *ḥadīt*; 4) the source of the *ḥadīt* (from Muslim’s al-Ṣaḥīḥ, from The Book of the Virtues of the Companions); 5) Frequency report.

¹¹ The Arabic term for a “massively transmitted” *ḥadīt* is *mutawātir*, as opposed to *aḥādī*, “transmitted by single transmitters”. *Mutawātir* implies that a *ḥadīt* is so widely transmitted that it cannot possibly be forged. For more details, see [17].

Step V

At last, the selection of *ḥadīṭs* was converted into a format for *LaTeX*, a document preparation system which allows to automatically typeset high-quality PDFs for printing and publishing [18]. Quite a few *ḥadīṭs* that made it to the beginning of the reader feature only *isnāds*, “the chains of transmitters”, and do not have *matns*, the actual texts of *ḥadīṭs* (Figure 3). The lack of *matns* actually gives an instructor a great opportunity to focus on the concept of *isnād* and provide some essential explanations of the culture of knowledge transmission in the Islamic world, to discuss transmission terminology (*ʿan*, *ḥaddaṭa-nā*, *aḥbara-nā*, etc.), as well as to explain the complexities of the traditional Arab name and its essential elements (*ism*, *nasab*, *kunyaṭ*, *laqab*, *nisbaṭ*, and *ṣuhraṭ*).

In the classroom

The reader was designed for my course “Classical Arabic through the Words of the Prophet” (Tufts University, Spring 2015). In the course of one semester we managed to cover about 400 *ḥadīṭs*, while at the same time reviewing the grammar of classical Arabic [19] and having regular discussions of thematic readings that helped students to understand the cultural importance of *Ḥadīṭ* across almost fourteen centuries of Islamic history [17].

Encouraged by my mentor at Tufts University, Gregory R. Crane, I used this reader in combination with ‘micropublications’, which provided each student with a thorough practice of foundational skills necessary for mastering the language: for each *ḥadīṭ* students provided full vocalization, morphological stemming, and translation aligned with its Arabic original. Such ‘micropublications’ help an instructor to monitor students’ progress, and, potentially, can be used to automatically grade such assignments, thus freeing up time for in-class discussions. Producing these micropublications, students were making academic contributions as they were producing valuable training data that can be used for various teaching and research purposes (e.g., automated syntactic and morphological analysis, machine translation, etc.). Additionally, students were required to complete written practice assignments where they had to practice verbal morphological patterns, conjugation of perfect and imperfect verb forms, conjugation of weak verbs, and handwriting (by tracing different texts in *ruqʿat* script).

Biography, Rank # 2

Ta'riḥ al-islām (000000002-023270-000.txt)

الحسن بن محمد بن موسى بن إسحاق بن موسى، أبو علي الأنصاري. [الوفاة: ٣٤٢ هـ] سمع جده موسى، وابن أبي الدنيا، والمبرد، وغيرهم. وعنه القاضي أبو القاسم بن أبي عمرو، ومحمد بن أحمد بن أبي عون شيخا الخطيب. وثقه الخطيب، وقال: توفي في ذي الحجة.

Vocabulary (by frequencies)

1) بن — son;	48) إسحاق — Ishāq (Isaac),	161) الأنصاري — al-Anṣārī,
2) في — in;	name;	nisbat denoting descent
3) أبو — father;	64) عمرو — ‘Amr, name;	from the Anṣār, “the Helpers
6) محمد — Muḥammad, name;	90) القاضي — [the] judge, qāḍī;	of the Prophet”;
9) الوفاة — death;	90) موسى — Mūsá (Moses),	259) الحجة — here, a part of
9) ه — abbr. of <i>hiǧrī</i> , Islamic	name;	<i>dū-l-ḥiǧǧat</i> , the 12th month of
lunar calendar;	116) الخطيب — [the] orator,	the Islamic lunar calendar;
11) أحمد — Aḥmad, name;	Friday preacher; here, name of	301) الدنيا — “this world”,
12) وقال — and [he] said;	a prominent Ḥadīṭ scholar—	“this life”; here, a part of the
13) علي — ‘Alī, name;	al-Ḥaṭīb al-Baǧdādī;	name of Ibn Abī-l-Dunyā, a
18) وعنه — and from him;	117) ذي — gen. of <i>dū</i> , owner,	prominent early scholar;
19) الحسن — al-Ḥasan, name;	possessor, “that of ...”;	301) جده — his grandfather;
19) توفي — (passive) [he] died,	129) وغيرهم — and others, lit.	454) شيخا — two <i>ṣayḥs</i> , two
lit. “he was taken [by God]”;	”and other than them”;	teachers;
19) سمع — [he] heard, listened	152) وثقه — [he] considered	952) والمبرد — al-Mubarrad, a
(i.e., studied);	him trustworthy (technical	prominent Ḥadīṭ scholar;
35) القاسم — al-Qāsim, name;	term);	

Grammar: For شيخا الخطيب, review dual and its behavior in *idāfat* (Thackston, IKCA, L4-§8; for *idāfat*, see Thackston, IKCA, L4-§8). For وثقه, see Factitive verbs of Form II: Thackston, IKCA, L27-§64. For توفي, see Reflexive/Medio-passive verbs, Form V: Thackston, IKCA, L29-§67.

Culture: Traditional Arab name is quite complex and includes up to six different parts. Here we find person’s “first name” (*ism*), which is al-Ḥasan. Then we find his “genealogy” (*nasab*)—the male names connected with *bn*—the name of his father, his grandfather, his great grandfather and his great great grandfather; *nasab* can go back as far as to the time of the Prophet. Then we find *kunyat*, “the father of ...”—Abū ‘Alī. And the last one is *nisbat*—al-Anṣārī: this type of names asserts relationship between a person and some kind of entity. In this case, the person traces his lineage back to the Anṣār, “the Helpers [of the Prophet],” who accepted Muḥammad in Yaṭrib as their leader. Yaṭrib later became known as Madīnat al-Nabī, or simply al-Madīnat (Medina).

Figure 4. An example section from a Biographical Reader based on the massive *Ta’riḥ al-islām* of al-Ḍahabī (d. 748/1347 CE): each biography can be supplied with translation notes, grammatical commentary (with references to textbooks), and relevant cultural notes.

The version of the reader that was used for the course can be accessed here: https://maximromanov.github.io/files/Romanov_CATWOP.pdf (an example of the written practice assignment is also included). This version was a test run and it can be improved in a variety of ways. Figure 4 shows a record from the working version of the biographical reader that is constructed with the same method from 30,000 biographies from the *Ta'rīḥ al-Islām* of al-Ḍahabī (d. 748/1347 CE): each reading item can be supplied with translation notes, grammatical commentary (with references to textbooks), and relevant cultural notes. Scripts for generating frequency-based readers are also openly available [20].

A Last Thought

While developed primarily with classical Arabic in mind (for Ḥadīṭ, more specifically), the approach is in fact universal and can be used for any genre and for any language. It works best with serialized texts — that is collections of relatively short texts of the same type. In the case of Arabic that would be *ḥadīṭ* collections, chronicles, biographical dictionaries, compendia of legal decisions (*fatāwá*), poetic anthologies, contemporary newspapers, etc. Because the vocabulary of various literary forms and genres may differ quite significantly (see, again, Figure 2), this method can be used for designing thematic reading courses that will help to ease students into the language of particular genres in the most efficient manner. Courses based on such readers can be a valuable addition to any language program and will be particularly welcomed by graduate students who often face the need to develop their readings skills as quickly and effectively as possible.

References

1. Al-Baġdādī, Ismā'īl Bāšā. 1992. *Hadīyat al-‘arīfīn asmā’ al-mu’allifīn wa-aṭār al-muṣannifīn* [*The Gift to the Knowledgeable: The Names of Authors and the Legacy of Compilers*]. Bayrūt: Dār al-kutub al-‘ilmīyat (in Arabic).
2. Witkam Jan Just. 2012. Ismā'īl Paṣḥa Baġhdādī. *Encyclopaedia of Islam* (EI2-Online). Eds P. Bearman, Th. Bianquis, c. E. Bosworth, E. van Donzel, and W. P. Heinrichs. Leiden: Brill.
3. Şaik Gökyay, Orhan. 2012. Kātib Čelebi. *Encyclopaedia of Islam* (EI2-Online). Eds P. Bearman, Th. Bianquis, c. E. Bosworth, E. van Donzel, and W. P. Heinrichs. Leiden: Brill.

4. Romanov Maxim G. 2017. Algorithmic Analysis of Medieval Arabic Biographical Collections. *Speculum* 92: S226–S246. doi:10.1086/693970.
5. Brockelmann Carl. 1943. *Geschichte der arabischen Litteratur*. In 5 vols. Weimer; Leiden: Felber; Brill.
6. Barabanov A.M. 1945. Explanatory Notes in the Arabic Manuscripts and Documents of the North Caucasus. *Soviet Oriental Studies* III: 183–214.
7. Jarvis Scott, Magali Paquot. 2015. Learner Corpora and Native Language Identification. *The Cambridge Handbook of Learner Corpus Research*. Eds Sylviane Granger, Gaetanelle Gilquin, Fanny Meunier. Cambridge: Cambridge University Press: 605–628. doi:10.1017/CBO9781139649414.027.
8. Belinkov Yonatan, Alexander Magidow, Alberto Barrón-Cedeño, Avi Shmidman, Maxim Romanov. 2019. Studying the History of the Arabic Language: Language Technology and a Large-Scale Historical Corpus. *Language Resources and Evaluation* (in print).
9. Al Khalil Muhamed, Nizar Habash, and Hind Saddiki. 2017. Simplification of Arabic Masterpieces for Extensive Reading: A Project Overview. *Procedia Computer Science* 117: 192–198. doi:10.1016/j.procs.2017.10.109.
10. Awajan Arafat. 2016. Semantic Similarity Based Approach for Reducing Arabic Texts Dimensionality. *International Journal of Speech Technology* 19: 191–201. doi:10.1007/s10772-015-9284-6.
11. Kurschinski Kellen. 2013. Applied Archival Downloading with Wget. *Programming Historian*. <https://programminghistorian.org/lessons/applied-archival-downloading-with-wget> (accessed on 27.02.2019).
12. Habash Nizar Y. 2010. *Introduction to Arabic Natural Language Processing*. Synthesis Lectures on Human Language Technologies 10. San Rafael, Calif.: Morgan & Claypool Publ.
13. Hodgson Marshall G.S. 1974. *The Venture of Islam: The Classical Age of Islam*. In 3 vols. Vol.1. Chicago: University of Chicago Press.
14. *GitHub: The World's Leading Software Development Platform*. 2008. <https://github.com> (accessed on 27.02.2019).
15. Miller Matthew Thomas, Maxim G. Romanov, and Sarah Bowen Savant. 2018. Digitizing the Textual Heritage of the Premodern Islamicate World: Principles and Plans. *International Journal of Middle East Studies* 50: 103–109. doi:10.1017/S0020743817000964.
16. Buckwalter Tim. 2004. Buckwalter Arabic Morphological Analyzer Version 2.0 LDC2004L02. *Linguistic Data Consortium*. <https://catalog.ldc.upenn.edu/LDC2004L02> (accessed on 27.02.2019).
17. Brown Jonathan A.C. 2011. *Hadith: Muhammad's legacy in the medieval and modern world*. Foundations of Islam. Oxford: Oneworld Publ.

18. *LaTeX: A Document Preparation System*. 2019. <https://www.latex-project.org/> (accessed on 27.02.2019).
19. Thackston W.M. 1994. *An Introduction to Koranic and Classical Arabic: An Elementary Grammar of the Language*. Bethesda, Md: Iranbooks.
20. Romanov M. 2018. *Scripts for Generating Frequency-Based Readers for Historical Languages: MaximRomanov / Frequency Based Readers*. <https://github.com/maximromanov/FrequencyBasedReaders> (accessed on 27.02.2019).