Humanities and Mathematical Approaches in the Case of Arabic and Islamic Studies

Olga Bernikova[1] and Oleg Redkin[2][2]

Research Laboratory for Analysis and Modeling of Social Processes,
St Petersburg State University,
11, Universitetskaya emb.
St Petersburg, 199034, Russia

[1]o.bernikova@spbu.ru, [2]o.redkin@spbu.ru

Abstract

Interdisciplinarity in the modern world is no longer considered a novelty. The joint work of representatives of various scientific fields of study allows achieving the maximum result. Examples of cooperation between computer specialists and humanitarians over several decades make it possible to analyze the peculiarities of interdisciplinary activities in diachrony and to outline possible prospects for the development of science and practice.
This study demonstrates how Arabic and Islamic studies have changed since the beginning of computer era, and the consequences it has brought. The authors also consider the implementation of Information and Communication Technologies along with traditional approaches in humanities as well as use of digital solution in education. Using the examples of some innovative projects the researchers demonstrate outcomes of interdisciplinary communication in the case of Arabic and Islamic Studies and review the development perspectives of this trend.

Keywords*: Arabic, Islamic Studies, ICT, interdisciplinary communication, Digital Humanities, computer scientists.*

1. Introduction

This article expands the research topic, the results of which were presented during the *20th World Multi-Conference on Systemics, Cybernetics and Informatics* (Bernikova & Redkin, 2016). In 2016 on the example of Arabic Studies, we discussed the peculiarities of the collaboration between computer

[1] Corresponding Author.
[2] We would like to express our gratefulness to Dr. Valery Lutkov for his detailed peer-editing of this document.

specialists and humanities in terms of goals, objectives, research methods, and their implementation and effectiveness. The conclusions presented four years ago highlighted the benefits of interdisciplinary collaboration in these areas, emphasizing the importance of the use of computer data processing techniques in Arabic Philology research. As a *case study*, the main trends in the development of Arabic philology are considered, which were partially presented in our article *Dialectal Atlas of the Arab World - Between Intention and Reality* published in the *Journal of Systemics, Cybernetics And Informatics* in 2017 (Bernikova, Redkin, 2017).

The effectiveness of interdisciplinary research methodologies have been proved by their rapid development and spread of their implementation in recent years. Besides that, Information and Communication Technologies (ICT) are increasingly penetrating our life. While development and spread of artificial intelligence in general and machine learning in particular have affected social processes, they have also changed the character and the role of the humanities in the development of interdisciplinary projects.

A new set of challenges has emerged in the above mentioned area, including forecasting the prospects of interdisciplinary dialogue. In this regard, this study analyzes modern approaches to the essence of interdisciplinary dialogue in comparison to the data presented by the authors in the 2016 and 2017.

The current results are based on the authors' participation in experimental projects and years long monitoring of the application of computer solutions in science and education on the example of Arabic and Islamic Studies.

Collaborative interdisciplinary work has become the essential feature of the present day research. In fact, various branches of humanities have always been overlapping and collaborating with each other (linguistics, cultural studies, ethnography, etc.) and in this sense research in this field has had multi- and interdisciplinary character. For example, no literature analysis can be made without historical context; mastering a foreign language is supposed to be achieved by reading and translating different kinds of texts.

The 'digital era' of time, i.e. introduction of mathematical approaches in humanities that occurred in last decades, has opened new possibilities for

scholars and enhanced their research activity. As a result we witness appearance of two major trends in the present day humanities.

The first trend is a traditional one, which relies first and foremost upon empiric data, while the second one is based on mathematical analysis which provides quantitative and objective characteristics and measurements that previously were almost completely neglected or based upon researcher's personal attitudes and evaluations.

This tendency needs to be demonstrated on the example of a definite field of study. The Arabic language is a good object which can help trace all the consequences of this new stage in the world science development. First of all, it is the language which is spoken by over 300 million people. Secondly, it is the language which is regarded to be a symbol of culture which has a long history.

Among ever existed idioms and dialects E. Sapir mentioned five languages, which played the most significant role in history as guides of culture - Chinese, Sanskrit, Arabic, Greek and Latin (Sapir, 1949). Besides that, Arabic is one of the most widespread languages of the Internet. In this regard, there is an urgent need for processing of large volumes of data in Arabic that are available on the World Wide Web.

It is the development and spread of the Internet that greatly contributed to the emergence of new interdisciplinary research areas. New times bring new challenges. Today Big Data processing of the Internet content is becoming relevant not only from the point of view of linguistic analysis, but also as study of public opinion on a particular issue, which in turn affects a number of specialties - from linguists and sociologists to psychologists and political scientists.

The present study aims to demonstrate how Arabic and Islamic Studies have been changed since the dawn of computer era and the consequences which it has brought. Are there any negative effects? How is it combined with the traditional approaches in humanities? What is its impact on education? What is the degree of implementation of such technologies as well as research and teaching attitudes in the institutions of higher education worldwide? On the examples of some innovative projects we demonstrate outcomes of

interdisciplinary approaches and consider future perspectives and developments in the case of Arabic and Islamic Studies.

2. Digital Humanities in the Case of Arabic

The role of Digital Humanities in present day scientific world was masterfully marked in the paper *Digital Humanities in Ancient Jewish, Christian and Arabic Tradition: Introduction to the Special Issue*: "…when data are digitally listed and mixed up, categorization used in the printed culture are overcome and new form of knowledge appear" (Clivaz, 2016, p.9).

As a matter of fact, methods of digital research could be implemented for any field of study. Thus, Digital Humanities involve all of the scientific paradigm, knowledge and skills accumulated by each of the relevant disciplines, using at the same time tools and perspectives opened up by Digital Technologies (Dacos, 2011). All of the aforementioned is also relevant in the case of the Arabic and Islamic studies.

Although Digital Humanities came into being in the late 1980s (Jänicke et al. 2015), systematic formalization of Arabic had had almost a decade before that with the paramount interest focused on the development of tools for automatic conjugation system of this language (McCarthy, 1981).

In the next years, the development of morphological analyzers for Arabic had remained one of the main topics of interdisciplinary research in the case of Arabic. In order to define the topics of the Arabic language research, which are among the most widespread nowadays, we have analyzed several data bases (including *SCOPUS* and *Web of Science*) and came to the following conclusions.

- The number of publications, dedicated to interdisciplinary issues in the case of Arabic greatly increased in comparison with 2005. Probably the situation may be explained by the fact that data bases the above mentioned focus on articles and books on issues related to engineering and natural sciences, which are far from being 'humanitarian' by their character.

Nevertheless, strong indicators which show the rising trend in the use of ICT in Arabic studies nowadays are quite obvious.

- In 2005 the most popular interdisciplinary topics in Arabic processing were the development of solutions for morphological analyzer and the optical character recognition.
- There was a shift from an automatic structure analysis of Arabic which was obvious in 1980-2005 to the implementation of the methods of visualization which became widespread nowadays.
- Among the most popular topics today are the speech processing and the dialect processing. The latter may be explained by the fact that dialects are used as the mean of oral communication. That is why the development of tools for speech processing and dialect processing are the two tasks closely related to each other. Among other issues we may mention handwriting recognition and manuscript digitalization. Significant progress over the past five years has been also achieved in the field of the development of technology for image recognition and automatic translation of texts from and into Arabic.

Political and social changes in the world in general and in the Middle East in particular that took place over the past decade, posed new challenges, in regard to Big Data processing of Internet content in Arabic. Not surprisingly, this has led to the emergence of new studies on the development of technologies of sentiment analysis of social networks and, primarily, the Tweeter. (See, for example: Assiri et al., 2016; Abdul-Mageed et al., 2014; Refaee & Rieser, 2014).

ICT also made it possible to reveal some important information for historians and sociologists which is hidden in an array of media in Arabic. In 2015-2016 the authors of the present paper were among the international group of linguists and mathematicians that carried out the project titled *Modeling and Visualization of Media in Arabic* (Volkovich et al., 2014).

The content of the newspapers was considered in diachrony as a mirror reflecting changes in the life of the society. For this purpose, a novel method for analyzing media in Arabic using new quantitative characteristics was proposed. The research was based on a sequence of daily issues of an Egyptian newspaper *Al-Ahraam* and a Lebanese newspaper *Al-Akhbaar*. The results

demonstrated a high ability of the proposed approach to expose changes in the linguistic content and to link them to structural changes in the society. This method can be applied in different spheres including analysis of media in other languages and mediaeval manuscripts authorship identification. In 2020 this approach has evolved to a new method to detect the correlation between the changes in social life and corresponding transformations of the linguistic content in Arabic newspapers (Bernikova et al., 2020).

Another example concerns the field of Arabic dialectology. The study of modern Arabic dialects presupposes field investigations 'on the place, i.e. in the dialectal environment while analysis of the collected data should be carried out mostly 'at the desk'. Although modern technologies have expanded the sphere of dialects use which today play a role as a mean of communication on the Internet, for example, in social networks, this does not exclude the need to carry out field research directly among the dialect speakers. There is an urgent need for developing of the atlas of the Arab-speaking world that may be presented as a traditional paper publication to be periodically updated and improved, but above all it must be a computer model showing both the evolution of idioms and the real state of the dialects during certain period of time. In recent years significant dialectal material has been collected and research methodologies for field and desk studies were developed in addition to the apparatus that allow to process research data with the help of digital technologies. The proposed project does not require significant material resources and lies in the area of scholar interests of researchers engaged in the Arabic dialectology. In this regard, it is important that since 1993, the international school for the study of the Arabic dialects - International Association of Arabic Dialectology, AIDA, (International Association of Arabic Dialectology, 2020) has become a panel for exchanging ideas for international scholars, which may be considered the prototype of a research team working on this project. To a large extent, it will represent the implementation and synthesis of their recent achievements and results of research work. It will also sum up the work of the dialectological community which has been carried on for over the past 150 years. The proposed project should be based on principles of international cooperation, interdisciplinary studies as well as usage of the cutting-edge technologies and methods of linguistic research. Implementation of these principles and methodologies will step up collective efforts of the project team members and combine skills of a large pool of scholars who may carry on joint research. The proposed project of

the dialectal atlas of the Arab world requires implementation of tools and attitudes similar to the *Rosetta Project*[3], bringing together international scholars and using digital technologies and online resources, combining results of field investigations and desk research, along with the real-time monitoring of the current state of the linguistic picture of the Arab world. In the future, it is necessary to make prompt upgrades to the map of the distribution of modern Arabic dialects and to check and, if necessary, modify existing isoglosses of phonetic, morphological and lexical peculiarities of these vernaculars (Redkin & Bernikova, 2017).

The above mentioned examples clearly demonstrate main tendencies in the field of the Arabic language studies. These trends are not only interdisciplinary, but multidisciplinary in their nature, and include not only the methodology of the Arabic linguistics and computer science, but also research methods of sociology, history, Islamic Studies and other humanities. In other words, in order to receive the proper result in any field of humanities, it is necessary to involve a vast range of specialists including mathematicians who can develop methods aimed at language processing which may provide quantitative and qualitative characteristics for such kind of research.

3. Digital Humanities and Education

Development of collaborative activity between IT engineers and specialists in humanities can be considered on the example of St. Petersburg State University (Russia).

Classical universities which have their long standing academic history provide a good opportunity to trace all aspects of implementation of ICT in teaching and research in order to demonstrate the process of integration of innovations into school of higher education.

The history of the St. Petersburg State University dates back to 1724 and the study of the Arabic language there started in 1818. Since the very beginning Arabic teaching and research at the University were combined with studing of history, geography, literature of the Middle East, and Islamic Studies.

[3] http://rosettaproject.org

Foreign language acquisition based on the study of original classical texts (including manuscripts) along with the study of history, religion and culture are among the characteristics that helped the Department of Asian and African Studies to achieve high international recognition.

At the same time, traditionalism which is considered an essential feature of any classical university makes it very complicated to introduce of innovations.

One of the examples of such innovations was the project *Oriental Languages Distant Learning System* launched at the Department in 2007. The project was aimed at the development of e–learning materials designed for Arabic and some other oriental languages. As a result, e-learning courses were created for IBM collaborative learning system and a set of good quality teaching books was developed and used on the initial stage. However, every e-content requires constant support and upgrade provided by computer specialists from one side and teachers from another. Besides that, the use of ICT in teaching on the initial stage is much more labor-intensive and requires more time and efforts in comparison to traditional methods.

Development of e-learning materials requires extra skills and qualifications of teachers who should be aware of the peculiarities of oriental language formalization, meanwhile the faculty of this kind is very rare.

Another technical problem was that the developers dealt with the IBM workplace collaborative learning platform which did not allow teachers to make modifications to their courses without involving IT specialists. The core of it was the complexity of IBM software solutions and difficulties of the customization of the Arabic script on this platform.

Since 2007, both e-learning technologies and the scale of their distribution in the world have been significantly improved.

Recent developments caused by the global pandemic demonstrated the exceptional importance of e-learning systems digital remote work format. The obtained results of quality control of the educational process confirm the fact that the availability of educational and methodological materials in digital form

and the information skills of pedagogical staff contributed to the effectiveness of distance learning.

The events of 2020 have changed the paradigm of key teaching requirements, and highlighted the importance of computer and telecommuting competences. For example, teaching Arabic remotely at St. Petersburg University has been facilitated thanks to the development in 2018-2019 of two online courses[4] that may be used at different levels of study. The development of these courses also contributed to the improvement of the methodological and informational skills of teachers, which made transition to a remote form quite natural, smooth and easy.

It should be noted that scholars in humanities who apply ICT in their research as a rule also implement them in pedagogical sphere. For example, in St Petersburg University different technologies for Arabic language processing were developed, which found their implementation 'in the classroom'. Table 1 demonstrates how results of interdisciplinary research reflect the methods of pedagogy in the case of Arabic.

The list of such kind of examples may be continued. Results of interdisciplinary activity 'in-classroom' provide training of a new generation of qualified specialists who will be able to implement different interdisciplinary approaches in their future work.

4. The Role of Scholars in the Humanities and Computer Specialists in Interdisciplinary Researches

The importance and effectiveness of interdisciplinary collaboration has various aspects of its manifestation and it raises a number of questions about fundamental differences in the roles assigned to the humanities on one hand and engineering and natural sciences from another.

[4] Arabic language. Introductory course. Available at: https://openedu.ru/course/spbu/ARBLNG/ ("National Platform of the open education") (accessed: 26.07.2020). (In Russian). Arabic language. Part 2. Available at: https://openedu.ru/course/spbu/ARBLNG2/ ("National Platform of the open education") (accessed: 26.07.2020). (In Russian).

Table 1. Effect of Digital Humanities approach on education[5]

Research	Education
Development of the technology of word frequency definition (Redkin & Bernikova, 2015).	Development of the vocabularies for text books, based upon a word frequency.
Development of morphological analyzer for Arabic (Redkin & Bernikova, 2015).	Improvement of courses on the Arabic grammar.
Development of technology of manuscript digitalization (Bernikova & Redkin, 2015).	Improvement of the discipline "Description of Arabic manuscript".
Development of mathematical modeling and visualization methods for media in Arabic (Volkovich, 2016).	Improvement of set of courses by adding quantity characteristics to the humanitarian disciplines ("History of the Middle East", "Media in Arabic", etc.).

Are their roles always considered equally important in such projects, do they have the same tasks and what are the prospects of such collaboration?

In fact, engineering and natural sciences specialists have always had an advantage since the subject of their research has been always associated with stability, although any computer solution has a limitation of its applicability, as well as its specific time and space (Granichin, 2019). Classical works of art and literature never lose their value being the foundations of the humanities while any technical and software methods of data processing develop rapidly and are relevant here and now, i.e. they are of interest exclusively from the point of view of the history of science.

It is unlikely to process large amounts of data using the first versions of computer hard and software, while classical works of art constitute the basis and foundations for the humanities.

[5] The data are based upon teaching and research activity of the authors of this paper.

The humanities traditionally balance between New and Old, trying to preserve the so-called classical approach in research and education. Mathematicians provide humanities with tools for fast and high-quality data processing, whereas the role of philologists in the area of natural research is much smaller. Only a couple of decades ago development of linguistic software products required a big amount of pure linguistic work, whereas in current projects the role of linguists has been changed and is now limited to testing of the application of certain solutions. The principle of operation of the latest machine translation systems could confirm this fact. It is quite likely that similar trends will be further developed along with the evolution of machine learning technologies and their large-scale application.

5. Conclusions

Implementation of mathematical approaches in humanities opens new opportunities for getting better and quicker results of scholar research, facilitates development of new scientific areas, especially in the sphere of interdisciplinary studies, provides practical application of results of any research. At present, a purely theoretical approach receives less and less support from scientific and educational organizations now.

Mathematical methods in humanities gave birth to a new multidisciplinary field of teaching and research better known as “Digital Humanities”. The case of the Arabic language processing proved effectiveness of this new attitude, as well as importance of its use in education.

The most popular topics in the Arabic and Islamic Studies at the present time are interdisciplinary in nature and assume the use of mathematical methods, which are implemented in speech and dialect processing, handwriting recognition, manuscript digitalization, sentiment analysis of media and social networks.

In order to receive the proper results in any field of study, it is necessary to provide cooperation between linguists and IT specialists who can develop methods specially designed for language processing and provide instruments

and criteria for quantitative and qualitative characteristics and measurements which may be applied in any kind of humanities.

In the nearest future combination of different approaches (i.e. mathematical, philological, etc.) will become imminent. 'Pure' humanitarian scientific research, isolated and limited by its nature will hardly bring solid results. On the contrary, team work, especially an interdisciplinary one will ensure the scholar's success.

Present-day humanities involve the implementation of Digital Technologies to a greater or lesser extent. It has also raised a lot of new topics as well as brought new challenges for computer specialists. As a result, interdisciplinary approaches and methodologies are considered to be an integral part of the scholar world today.

5. Acknowledgments

The funding no. 15.61.2219.2013 from Saint-Petersburg State University (Russia) to support this research project is gratefully acknowledged.

6. References

Abdul-Mageed, M., Kubler, S., Diab, M. (2014). SAMAR: Subjectivity and sentiment analysis for Arabic social media. *Computer Speech and Media, 28 (1)*, 20-37.

Assiri, A., Emam, A., Al-Dossari, H. (2016). Saudi Twitter Corpus for Sentiment Analysis. *Journal of Computer, Electrical, Automation, Control and Information Engineering, 10 (2)*, 233-236.

Bernikova, O., Granichin, O., Lemberg, D., Redkin, O., & Volkovich, Z. (2020). Entropy-based approach for the detection of changes in Arabic newspapers' content. *Entropy, 22 (4)*.

Bernikova, O., Redkin, O. (2015). The Arabic Language Processing: Peculiarities of Interdisciplinary Research. *Proceedings of the IMSCI 2015 - 9th International Multi-Conference on Society, Cybernetics and Informatics, July 12-15* (pp. 81-86). Orlando; United States: IIIS.

Bernikova, O., Redkin, O. (2016). Humanities and Mathematical Approaches in the Case of Arabic. *Post-conference proceedings of the 20th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2016, 5-8 July*. Orlando, USA: IIIS.

Clivaz, C. (2016). Digital Humanities in Ancient Jewish, Christian and Arabic Traditions: Introduction to the Special Issue. *Journal of Religion, Media and Digital Culture, 5 (1)*, 1-20.

Dacos, M. (2012, 1 25). *Manifeste des Digital Humanities.* Retrieved from Hypotheses: http://tcp.hypotheses.org/318

Granichin, O. (2019). Mathematical Methods in the Analysis of Cultural, Historical and Social Phenomena. *Innovations and Traditions in Arabic and Islamic Studies: Festschrift in Honor of Professor Oleg Redkin*, 246-254. Saint-Petersburg: St. Petersburg State University.

International Association of Arabic Dialectology. (2020). Retrieved from http://aidabucharest2015.lls.unibuc.ro/

Jänicke, S., Franzini, G., Cheema, M.F. & Scheuermann, G. (2015). On Close and Distant Reading in Digital Humanities: a Survey and Future Challenges. *Proceedings of EuroVis — STARs*, (pp. 83-103).

McCarthy, J. (1981). A Prosodic Theory of Nonconcatenative Morphology, 12. *Linguistic Inquiry*, 373-418.

Redkin, O., Bernikova, O. (2015) Interdisciplinary Studies of Arabic and Islamic Manuscripts. *Proceedings of the 9th International Multi-Conference on Society, Cybernetics and Informatics, July 12 - 15 July* (pp. 87-91). Orlando; United States: IIIS.

Redkin, O., Bernikova, O. (2017). Dialectal Atlas of the Arab World - Between Intention and Reality. *Journal of Systemics, Cybernetics and Informatics, 15 (5)*, 44-47.

Refaee, E., Rieser, V. (2014). Subjectivity and Sentiment Analysis of Arabic Twitter Feeds with Limited Resources. *Proceedings of the Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools, 24 May* , (pp. 16-21). Reykjavik, Iceland.

Sapir, E. (1949). *Selected writings in language, culture and personality*. Berkley: University of California Press.

Volkovich, Z., Granichin, O., Redkin, O., Bernikova O. (2016). Modeling and visualization of media in Arabic. *Journal of Informetrics, 10 (2)*, 439-453.