

The Arabic Language Status in the Jordanian Social Networking and Mobile Phone Communications

Gheith A. Abandah
Computer Engineering Department
The University of Jordan
Amman, Jordan
abandah@ju.edu.jo

Waleed A. Anati
Arabic Language and Literature Department
Petra University
Amman, Jordan
anati_waleed@hotmail.com

Sami M. Ababneh
Arabic Language Department
The University of Jordan
Amman, Jordan
sami_ababneh@hotmail.com

Mohammed Zeki Khedher
Electrical Engineering Department
The University of Jordan
Amman, Jordan
khedher@ju.edu.jo

Ahmad A. Zghoul
Department of Mathematics
The University of Jordan
Amman, Jordan
a.zghoul@ju.edu.jo

Mamoun S. Hattab
Arabic Textware
Amman, Jordan
m.hattab@arabtext.ws

Abstract—The internet and smartphone penetrations continue to rising reaching large percentages of the world populations. Likewise, many Jordanians are actively communicating through the popular social networks and mobile phone messages. There are large questions and concerns related to the characteristics and quality of the language used in these forums and how to improve it. This study addresses these issues by collecting and analyzing a large sample of text from five sources: Facebook, Twitter, news sites, blogging sites, and mobile phone short messages. We analyzed the sample comprehensively including the sender, context, message, channel, and code. We present in this paper the results related to the used language, alphabet, dialect, text components, and style. The study concludes that the bilingualism problem is manifested in Twitter and Facebook with 24% and 14% of contributions in English, respectively. Moreover, 6.4% of the analyzed Arabic samples have English words and 13.2% are written in Arabizi (Arabic in English letters and numerals). The diglossia problem is manifested as 55.4% of the sample is in colloquial Arabic, 36.4% in the standard Arabic, and 8.2% in standard with some colloquial words.

Keywords—Arabic language; Jordan; social networks; Facebook; Twitter; blogging; electronic news sites; short messages

I. INTRODUCTION

The internet, social networks, mobile phones, and smartphone penetrations are increasing year after year globally [1-3]. The Arab World and Jordan are no exceptions. More and more people are accessing the internet and social networks through their computers and smartphones. In 2014, Jordan has reached internet and mobile phone penetrations of 74% and

147%, respectively [4]. Many Jordanians are actively communicating through social networks and mobile phones. The penetration of famous social networks in Jordan such as Facebook, LinkedIn, and Twitter has reached 47.9%, 5.0%, and 2.4%, respectively [5]. In fact, Facebook is the top internet site visited in Jordan [2] and Jordanians send more than 11 million tweets monthly and have exchanged over 2.5 billion short messages last year [4]. These rates are expected to continue rising due to the rising smartphone penetration and the increasing popularity of free messaging services such as WhatsApp, Skype, and Viber.

There are also many indicators that the number and percentage of internet contributions in the Arabic language through these forums in the Arab World are increasing [5]. However, there are concerns about the quality and type of the Arabic language used in these forums and how the internet affects the language and vice versa [6].

This paper summarizes our study of the status of the Arabic language that Jordanians use in social networks and mobile phone communications. The main objectives of this study are to find the main characteristics of the Arabic language used and to identify the main problems in the quality of the language used. Hopefully, this identification would lead to solutions to improve the quality and effectiveness of Arabic language communications in these forums.

This study incorporated Jakobson's effective communication model, including the sender, context, message, channel, code, and the receiver [7]. We collected many text samples and information about their sender, context, and

channel from five sources. The five sources are Facebook, Twitter, News sites comments, blogging sites, and mobile phone messaging.

There are several studies that have tackled the subjects of the Arabic language on the internet and mobile phone messaging in several Arabic countries [8-11, 17]. However, This study handles these issues more comprehensively in Jordan by collecting large sample from five sources and analyzing this sample on many aspects as detailed below.

The details of this study are published in a long technical report [18]. This paper summarizes the methodology used in collecting fair and representative sample and analyzing this sample. Moreover, we present the result analysis related to the used language, alphabet, dialect, text components, and style.

Section II summarizes the methodology used including the developed sample collection and analysis application and the sample collection methods from the five study sources. Section III presents the results of the used language, alphabet, dialect, text components, and style. Finally, Section IV summarizes and discusses the main results, identifies three main problems, and suggests some recommendations and future work.

II. METODOLOGY

In this section, we introduce the methodology used in this study. We describe the application developed to collect and analyze samples. We also describe how samples were collected from the five study sources.

A. Sample Collection and Analysis Application

We have developed a web-based application to facilitate and speed up the processes of sample collection and analysis. This application supports two main roles: sample *collector* and sample *analyzer*. Fig. 1 shows the main page used in sample collection. For each sample, the collector uses this page to specify the following fields.

- The sample text and topic
- The URL of the sample source webpage
- Sample source (Facebook, Twitter, News, Blogs, or Messages)
- The author's age, sex, and education level (if available)

Fig. 2 shows the main page used by the sample analyzer to analyze the lingual characteristics of the collected samples. This page shows the sample text, URL, and number of words. The analyzer analyzes the sample text and uses the nine hyperlinks at the lower right part of this page to enter the analysis for the shown sample.

These nine hyperlinks allow the analyzer to access nine subpages and enter the following analysis information about each sample.

- 1) Language information (detailed below)
- 2) Foreign words present in the text
- 3) Common words used in the text

Fig. 1. Sample collection page.

Fig. 2. Sample analysis page.

- 4) Idiomatic expressions used in the text
- 5) Foreign prefixes and suffixes
- 6) Hybrid Arabic/English words

- The twitter is a person, not an organization
- The twitter's country is Jordan
- The tweet's language is Arabic, English, or mixed

Moreover, the sample collection process extended from Jul 18, 2013 to Sep 4, 2013 over all week and day times.

3) News Sites Comments

There are more than 118 electronic press sites in Jordan [19]. Most of these sites allow the visitors to comment on the posted news. For some sites, these comments reach hundreds of comments for some popular news items.

We have collected 1,504 samples of these comments over a two-month period over all week and day times. We collected samples from various news topics including politics, economics, sports, society, arts, and culture. The sites from where these samples were collected are the sites that are most visited in Jordan [2] and allow visitor comments. The sites that we have collected samples from are Jfra News, Khaberni, Ammon News, Alghad Newspaper, Assabeel, and Tasweer News.

4) Blogs

Blogging became popular in Jordan more than 10 years ago. Many bloggers use their blogging sites to express their views, ideas, and feelings. Many specialists think that some blogging sites such as the *Black Iris* have contributed in raising the ceiling of freedoms in Jordan [13]. However, traditional blogging is in decline as more and more bloggers are expressing themselves through Facebook and Twitter. Moreover, many Jordanian blogs are in English and reach selected segment of the Jordanian population.

Most blogging sites allow visitors to comment on the posted blogs. We have collected 52 original blogs from 52 blog sites and 459 comments on these blogs. These blogs come mainly from the most popular Jordanian blog sites according to *Jordan Blogs* and *Best Jordanian Bogs* [14, 15]. The details of these blogs are in [18].

5) Short Messages

Users of mobile phones often communicate through sending short text messages to each other. Recently, many smart phone users send such messages free of charge through specialized services such as WhatsApp, Skype, and Viber. As these messages are private from the sender to the receiver(s), we cannot collect samples of them through some open source venue. Therefore, we invited volunteers to give us samples they received on their mobile phones. To improve the fairness and representativeness of these samples, we asked each volunteer to submit 5-25 message samples that satisfy the following criteria.

- Arabic message or mixed (Arabic and English)
- Randomly selected without restriction on the message topic
- Not from some organization or some advertisement, but must be from a person

We have collected 2,502 from 141 volunteers most of them are male and female students from Jordanian universities. However, as the volunteers provided the messages they have

received (not sent), the sample represents a larger segment of the Jordanian population.

III. RESULTS

In this section, we present the language information analysis of this study. The following subsections present the results found about the used language, alphabet, dialect, components, and style on the five study sources. We also comment on these results and provide some explanations.

A. Language

The language of the text Jordanians use in the five study sources is Arabic, English, or mixed Arabic and English. This study concentrated on the samples that use either Arabic or mixed language. However, we have counted the number of samples encountered in this study that use pure English. In Facebook and Twitter, 14% and 24% of the users' contributions are in English, respectively.

Fig. 4 shows the distribution of samples that are not in English. The figure also shows the average of the distribution of the five sources. More than 95% of these samples are in Arabic and less than 5% are in Arabic with some English words or phrases. The lowest percentages of mixed language are observed in the News and Blogs (0.8% and 2.8%) and the largest percentages are in Twitter and Messages (7.3% and 6.4%).

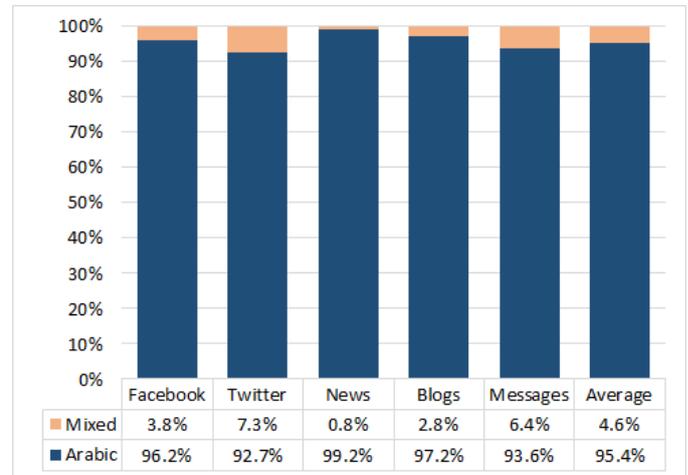


Fig. 4. Language used: Arabic or Arabic with some English words.

We explain the differences in using the mixed language among the five sources by the following points.

- The generally formal communication using news comments and blogs involves better attention to the language and using fewer foreign words.
- As tweets are limited to 140 characters each, the users strive to express their ideas with minimal characters and often use special Twitter features such as *@name* to draw the attention of some user and *#keyword* to hash tag their post with the intended subject's keyword.
- The entry difficulties in mobile phones (explained further in the next subsection) result in using higher percentage of English words in Messages.

B. Alphabet

Arabic is usually written using its alphabet that has 28 basic letters [20]. However, due to technical issues, some writers write Arabic using English letters. Currently, many people write Arabic using English letters and numerals. This writing style is called *Arabizi* [16]. Basically, The Arabic letters that have English counterparts are written using their English counterparts, e.g., ‘s’ for Arabic **Seen** (س) and ‘b’ for **Beh** (ب). The rest Arabic letters are written using English letter combinations, e.g., ‘th’ for **Thal** (ث) and ‘sh’ for **Sheen** (ش), or using numerals that are closest to them in shape, e.g., ‘3’ for **Ain** (ع) and ‘7’ for **Hah** (ح).

Fig. 5 shows the distributions of the samples according to the alphabet used. On average, 86.8% of the samples are written using the Arabic alphabet and 13.2% are written using the Arabizi alphabet. Again, News and Blogs have the lowest percentages of Arabizi (0.3% and 4.4%) and Messages has the largest percentage (31.0%). We explain these results as follows:

- The users involved in News, Blogs, and Twitter are usually more sophisticated and have better language skills that enable them to express themselves effectively in proper Arabic.
- Communication through Messages and Facebook is generally less formal and many users often pay less attention to their language especially during casual and daily communications.
- Moreover, entering text through mobile phones is harder than through computers. Many young Jordanians are skilled in entering fast English messages through their mobile phones. They also often enter their Arabic messages using the same skills, resulting in the high Arabizi percentage.

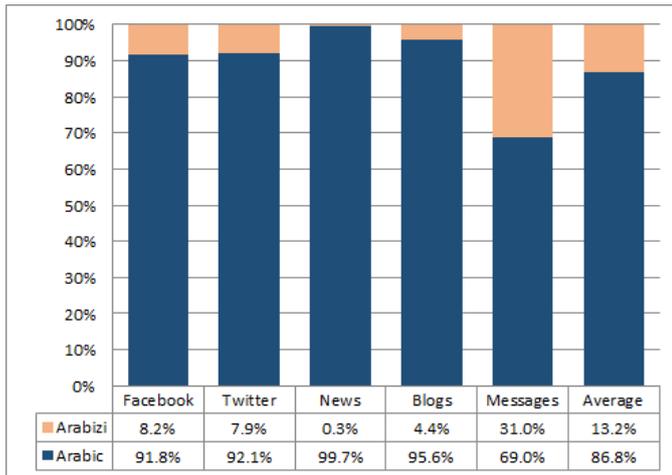


Fig. 5. Alphabet used: Arabic or Arabizi.

C. Dialect

Fig. 6 shows the distributions of the samples according to the dialect used. On average, more than one half the samples (55.4%) are using the colloquial Jordanian Arabic, above one third (36.4%) use the standard Arabic, and the rest (8.2%) use the standard Arabic with some colloquial words.

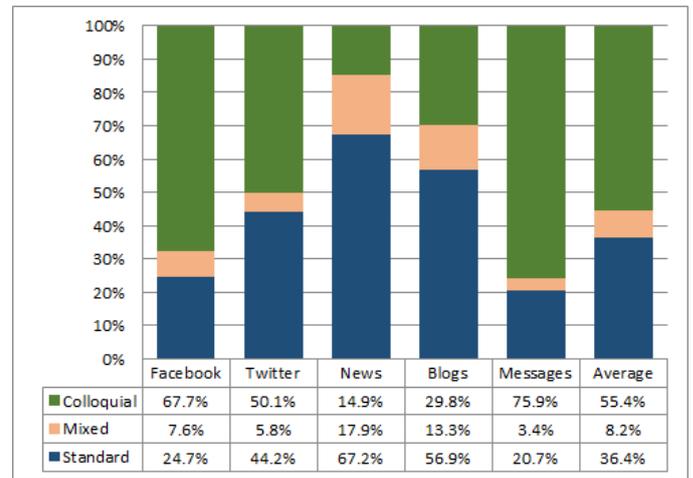


Fig. 6. Dialect used: Standard Arabic, mixed standard and colloquial, or colloquial Arabic.

For the reasons mentioned above, Messages and Facebook have the highest percentages of colloquial and mixed dialects (79.3% and 75.3%) because they are closer to casual communication. Whereas, News and Blogs have the highest percentages of standard Arabic (67.2% and 56.9%). Recall that the results of the Blogs include the blogs and their comments. When concentrating on the blogs without their comments, we notice that the percentages of standard, mixed, and colloquial are improved at 62%, 27%, and 11%, respectively.

The relatively high percentage of the mixed language in News reflects that the people involved in these comments generally prefer using standard Arabic. However, they often resort to inserting some colloquial words that they feel will enhance conveying their ideas and feelings.

D. Text Components

Fig. 7 shows that an average of 79.2% of the samples uses standard letters and characters and 20.8% additionally uses special symbols and suggestive visual writing. The highest percentages of this mixed writing are in Twitter, Facebook, and Messages at 37.3%, 22.1%, and 19.2%, respectively.

This mixed writing includes using special character sequences to express feelings such as :) for happy face, :(for sad face, and <3 for love heart. Some sites, like Facebook, automatically convert such sequences to nice images like ☺, ☹, and ♥. Moreover, some applications, like WhatsApp, have special menus and keyboard layouts to facilitate entering these popular drawings. This mixed writing also includes repeating some letters to give some suggestive expression such as هههههههه for long laugh, خخخ for cynical laugh, and ممم for wonder.

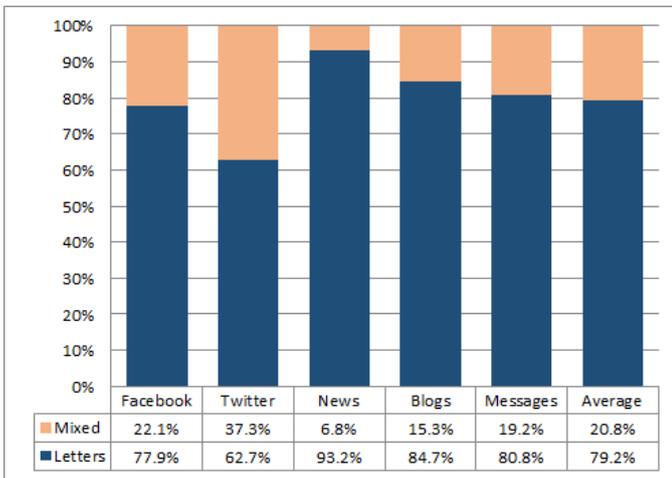


Fig. 7. Text components used: Letters or letters and symbols.

This mixed writing allows the users of Twitter, Facebook, and Messages to express their feelings efficiently with small number of characters. For example, instead of writing “I feel happy” one can enter :).

As the figure suggests, this mixed text is more common in personal communication. It complements the text messages directed to others to include feelings that the body language usually reflects in face-to-face communication. Therefore, this mixed text is less common in the formal communication of News and Blogs.

E. Language Style

Fig. 8 shows the distributions of the samples according to the rhetorical style used. On average, 82.6% of the samples use the normal or plain style. This is expected as most of these samples are related to direct day-to-day communications. However, Twitter users that are usually educated and pay good attention to their tweets often use the metaphor and cynical styles at 17.5% and 17.1%, respectively.

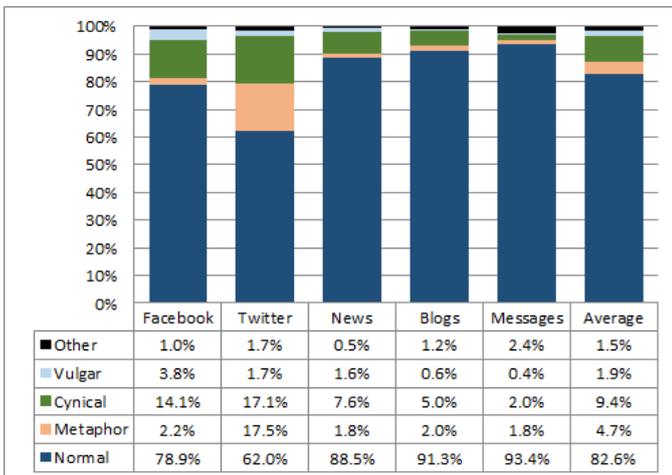


Fig. 8. Language style: normal, metaphor, cynical, vulgar, or other.

The cynical style is also common in Facebook and News at 14.1% and 7.6%, respectively. We think that this is an interesting phenomenon worth of further investigation. We

wonder how this phenomenon is related to the hard Jordanian political, economic, and social situations.

Finally, the percentage of the vulgar style is low at an average of 1.9%. This indicates that most contributions observe politeness and good manners. However, Facebook has the highest percentage of this style at 3.8%.

IV. DISCUSSION AND CONCLUSIONS

In this paper, we have described the methodology used in a comprehensive study of the status of the Arabic language in the social networking forums and mobile phone messaging. This study collected a large sample of 8,538 Jordanians’ contributions in the five study sources: Facebook, Twitter, News, Blogs, and Messages. These samples were analyzed taking into consideration Jakobson’s communication model.

This paper presents the study results related to the code and some message aspects, specifically the results of the language information analysis. We can summarize the results for each source as follows.

- **Facebook** has significant percentages of messages in Arabizi and colloquial Arabic at 8.2% and 67.7%, respectively. Users often include symbols in their text (22.1%) and mainly use the normal and cynical styles (78.9% and 14.1%).
- **Twitter** has about 24% of the tweets in English and 7.3% of the remaining tweets are in Arabic with English words. Arabizi and colloquial percentages are also significant at 7.9% and 50.1%, respectively. Twitter users show the highest use of symbols (37.3%) and the highest use the metaphor and cynical styles (17.5% and 17.1%).
- **News** has the lowest percentages of message in Arabic with English words, using Arabizi, and with symbols at 0.8%, 0.3%, and 6.8%, respectively. However, it has highest percentages of messages in standard Arabic and standard Arabic with some colloquial words at 67.2% and 17.9%, respectively.
- **Blogs** also has low percentages of message in Arabic with English words, using Arabizi, and with symbols at 2.8%, 4.4%, and 15.3%, respectively. And high percentages of messages in standard Arabic and standard Arabic with colloquial words at 56.9% and 13.3%, respectively.
- **Messages** has the highest percentages of messages using Arabizi, in colloquial Arabic, and in the normal style at 31.0%, 75.9%, and 93.4%, respectively. The use of symbols is also high at 19.2%.

These results show that there are the following three problems related to the status of the Arabic language on these forums.

- **Bilingualism Problem:** In addition to Arabic, English is highly present in these forums. The contributions of Jordanians in Twitter and Facebook are 24% and 14% in English, respectively. Moreover, the analyzed Arabic contributions from the five sources show that 6.4% of

the messages have English words and 13.2% are in Arabizi.

- **Diglossia Problem:** The colloquial Arabic is common in conversation and casual communications. The standard Arabic, on the other hand, is used in formal communications. These dual dialects were observed in the five study sources at averages of 55.4% in colloquial Arabic, 36.4% in standard Arabic, and 8.2% in standard with colloquial Arabic.
- **Linguistic Weakness Problem:** This problem is not presented in this paper, but the study shows that there is high percentage of contributions that have weak Arabic language. This weakness is manifested in large rates of spelling, lexical, morphological, and grammatical errors.

We think that these problems can be mitigated by technical and nontechnical solutions, including legislative, informational, and educational solutions. The technical solutions should concentrate on improving how Arabic text is efficiently entered, especially on mobile phones. Moreover, there is a great need to improve the operating systems and applications' support of the Arabic language. Arabic spell and grammar checkers, for example, are not available or expensive. Developing and freely providing such support would definitely improve the Arabic language on such communication forums.

Finally, and as ideas for future work, we are interested in studying these issues in other Arabic countries and even for Arab communities in foreign countries. Moreover, we are interested in monitoring these issues over time and studying the effect of technological advancements in smartphones and communications on these issues.

ACKNOWLEDGMENT

This research was organized and financed by the Jordan Academy of Arabic and the Jordanian National Committee for the Advancement of the Arabic Language towards the Knowledge Society. The authors are grateful for the financial support of this research.

REFERENCES

- [1] Internet Live Stats, Internet Users by Country (2014), <http://www.internetlivestats.com/internet-users-by-country/>, Jul 1, 2014.

- [2] Alexa, Top Sites in Jordan, <http://www.alexa.com/topsites/countries/JO>, last visited Apr 23, 2014.
- [3] M. Kakihara, "Grasping a global view of smartphone diffusion: An analysis from a global smartphone study," Int'l Conf. on Mobile Business, 2014.
- [4] Telecommunications Regulatory Commission of Jordan, Telecommunications Indicators: 2014/Q3, <http://www.trc.gov.jo>, last visited Jan 22, 2015.
- [5] Arab Social Media Report, Citizen Engagement and Public Services in the Arab World: The Potential of Social Media, 6th ed., The Governance and Innovation Program, Mohammed Bin Rashid School of Government, Dubai, available on: www.ArabSocialMediaReport.com, June 2014.
- [6] D. Crystal, Language and the Internet. Cambridge University Press, 2001.
- [7] R. Jakobson, "Closing statement: Linguistics and poetics," in Style in Language, T. A. Scbcok, Ed. New York: Wiley, 1960, pp. 350-373.
- [8] M. Wrschauer, G.R. ELSaid, and A. Zohry, "Language choice online: Globalization and identity in Egypt," J. of Computer-Mediated Communication, vol. 7, no. 4, Jul 2002.
- [9] R.A. Abdulla, "Arabic language: Use and content on the internet," Bibliotheca Alexandrina Access to Knowledge Toolkit I, pp.124-140, 2009.
- [10] R. Peel, "The internet and language use: A case study in the United Arab Emirates," Int'l J. on Multicultural Societies, vol. 6, no. 1, pp. 146-158, 2004.
- [11] M.A. Al-Khatib and E.H. Sabbah, "Language choice in mobile text messages among Jordanian university students," SKY J. of Linguistics, vol. 21, pp. 37-65, 2008.
- [12] Social Bakers, Jordan Facebook Page Statistics, <http://www.socialbakers.com/facebook-statistics/jordan>, last visited Jan 15, 2015.
- [13] The Black Iris of Jordan, <http://black-iris.com>, last visited Jan 18, 2015.
- [14] Araboo, Jordan Blogs, <http://www.araboo.com/dir/jordan-blogs>, last visited Apr 23, 2014.
- [15] Squidoo, Best Jordanian Blogs, <http://www.squidoo.com/jordan-blogs>, last visited Apr 23, 2014.
- [16] M. A. Yaghan, "Arabizi: A contemporary style of Arabic slang," Design Issues, vol. 24, no. 2, pp. 39-52, 2008.

REFERENCES IN ARABIC

- [١٧] و. المنصور، "من استعمالات اللغة المحذثة (العربيزي)"، وقائع مؤتمر اللغة العربية ومواكبة العصر، الجامعة الإسلامية بالمدينة المنورة، ٢٠١٣.
- [١٨] اللجنة الوطنية الأردنية للتهوض باللغة العربية للتوجه نحو مجتمع المعرفة، رصد واقع اللغة العربية في ميدان التواصل على شبكة الإنترنت والهاتف المحمول - دراسة علمية ميدانية تحليلية - الواقع والمأمول، مجمع اللغة العربية، أيلول ٢٠١٤.
- [١٩] ويكيبيديا، ملحق: قائمة الصحف الإلكترونية الأردنية، <http://ar.wikipedia.org>، آخر زيارة ١٨ كانون ثاني ٢٠١٥.
- [٢٠] غ. عبندة وف. خندقجي، "قضايا في رمز الحروف العربية"، العلوم الهندسية، مجلة دراسات، مجلد ٣١، عدد ١، ص. ١٦٥-١٧٧، ٢٠٠٤.