Collecting Arabic Dialect Variations using Games With A Purpose: A Case Study Targeting the Egyptian Dialect

Sara Nasser, Nada Sharaf, Mohamed Khamis, Slim Abdennadher and Caroline Sabty

Computer Science and Engineering Department,
The German University in Cairo
sara.zohny@student.guc.edu.eg, {mohammed.khamis, nada.hamed, slim.abdennadher, caroline.samy} @guc.edu.eg

Abstract. Arabs throughout the Arab world speak different dialects of Arabic in their daily conversations. We envision collecting a data set that maps different Arabic variations and dialects to Modern Standard Arabic (MSA). These mappings can be then used to facilitate the communication among Arabs from different regions. In this work, we developed a Game With A Purpose (GWAP) to collect mappings between MSA and the Egyptian dialect of Arabic as a case study. We show preliminary results that demonstrate that the game was successful in collecting considerable amount of data in a short time.

Keywords: Games With A Purpose, Arabic, Arabic dialects, Egyptian dialect

Introduction

In 2011, it was estimated that around 86 million Internet users are Arabic speakers [1]. Arabic is the 4^{th} most spoken language in the world [3], and the 7^{th} most used language on the Internet [2]. There is an agreed upon Modern Standard Arabic (MSA), which is usually used in formal talks, articles and shows. Nevertheless, the daily used forms of Arabic differ greatly from one region to another in the Arab world. Moreover, studies have shown that in each of the 22 countries of the Arab world, there are at least five different levels of Arabic [4].

This difference has made it harder for Arabs from different regions to communicate. Additionally, there is a lack of data sets that map these dialects to MSA, not to mention the lack of mappings of these dialects to each other. Collecting a data set of mappings of different dialects to MSA will pave the road to translating dialects, thus facilitating communication between Arabs from different regions. As an example, an application that makes use of such data sets can make it possible for an Egyptian, speaking in his own dialect, to address a Tunisian, who only speaks his own dialect of Arabic. The application would then be able to map the Egyptian's phrases to MSA, then from MSA to the corresponding Tunisian phrase.

In this paper, we propose using Games With A Purpose (GWAPs) [13] to collect different dialect variations. We start by collecting mappings between MSA phrases, that we believe to be important in conversations, and corresponding Egyptian dialect phrases. We envision several GWAPs to collect such mappings for several dialects, eventually building up a data set of the most commonly used phrases. Such data set can then be used to bridge the gaps between the different regions of the Arab world.

Related Work

Games With A Purpose (GWAPs) [13] are applications of human computation. The term "Human Computation" was coined by Luis von Ahn in 2005 [12]. The idea behind the field is to make use of the human brainpower on the Internet. Von Ahn argues that there are a lot of tasks that humans excel in doing, but computers still fail to do at the same speed and accuracy (e.g. translation and image recognition). Thus, hiding such tasks in activities that humans perform willingly seems plausible. A GWAP is an entertaining game in which players solve a problem as a side-effect of playing a game. In a relatively short time, GWAPs collect vast amounts of reliable and objective data that would otherwise be costly and tedious to collect.

Crowdsourcing and GWAPs are increasingly gaining popularity in the field of linguistics. They offer alternative methods for linguists in various applications such as collection and analysis of data. Several approaches use Amazon's Mechanical Turk (MTurk) to collect data for different purposes. For instance, in [5], turkers are asked to translate the sentences in Urdu to English. This helps create test sets for statistical machine translation engines. Another application of MTurk in the field was directed towards summarizing documents, by asking the turkers to select a maximum of half of a sentence from an article as its summary [7]. Moreover, in [17], MTurk is used to classify four different Arabic dialects based on their region and translates two of them (Egyptian and Levantine) into English. The platform Metropolitalia [10] focused on the dialect variations of the Italian language. The platform allows its users to collect and asses linguistic data. However, the engine is a crowdsourced-platform and not a GWAP. GWAPs have several advantages over MTurk platforms. GWAPs collect data for free, and can be improved in many ways by introducing new fun aspects, challenges and incentives. We acknowledge that there could be a tendency to cheat in GWAPs, nevertheless we think that the probability to cheat for a higher score is less that the probability to cheat for more money in the case of MTurk.

Some games were also introduced in the context of linguistics. For example, [8] proposed a multiplayer collaborative game to collect data in order to solve the morphological disambiguation problem of the Turkish language. Phrase Detectives is another game, that aims at identifying relationships between words and phrases in a short text [6].

There has been GWAPs that can accept and handle Arabic input [9]. However, to the best of our knowledge, this is one of the few GWAPs that address

Arabic-related problems, and the first attempt to collect mappings between Modern Standard Arabic and different Arabic dialects using GWAPs.

Case Study: Collecting a Map of Modern Standard Arabic to Egyptian Dialect

Although the Egyptian dialect is not the closest one to MSA, it remains the most commonly understood dialect across the Arab world, due to the popularity of Egyptian shows, songs and movies [16].

We developed a Game With A Purpose that provides players with phrases in MSA, and asks them to guess corresponding ones in the Egyptian dialect (see Figure 1). The more frequent a guess is matched, the higher the score the player gets.

Fig. 1: The player's perspective after providing one guess

Game Idea

According to von Ahn's classification of GWAP templates, our game falls under the category of output-agreement games[15]. Like other output-agreement games, players in our game are picked randomly and have no means of communication. In order to win, the players are given the same input, which in our case is the MSA phrase, and are expected to *output* the same phrase in the Egyptian dialect. We refer to the player's output as a guess. Since the paired players cannot communicate with each other, we can argue that the collected data come from independent sources. Thus, the higher the agreement upon a mapping is, the more likely that it is correct.

Walkthrough

Players are first asked to register/login in order to keep track of their scores. Once logged in, players can choose to read the game instructions, or directly start playing. Upon choosing to play, the game picks a random pair of the online players, but makes sure that each player sees a phrase that he/she hasn't played before. This was done to avoid influence from previous games. If there were no enough players online, the unassigned player is directed to a waiting page, till new players login.

Most of the traditional output-agreement GWAPs [14,15] take players to the following level once they output a match. Our game on the other hand, permits players to further provide guesses till the 30 seconds round time is over, in exchange for higher score points. This enables us to better utilize the players' intellectual ability.

Leader boards and scoring were proven by von Ahn to be strong incentives to GWAP players [13]. Players are awarded points based on the following criteria: ten points if both players output a match for the same MSA phrase in the same game, five points if the output guess was mentioned in a previous game that had the same MSA phrase, and one point if this output guess is a new one. This encourages players to provide guesses that others will agree upon, thus more likely to be correct. Additionally, as a further incentive, the records of the top seven players are kept in a hall of fame page.

During the game, players are allowed to pass a round if they find the phrase too difficult or not interesting. The player should wait until their opponent agrees. Only when the opponent confirms, a new phrase is displayed to both players.

At the end of every round, both players are directed to a recap page displaying their answers. The answers are classified into three groups according to the score given for each answer. The recap phase remains in display for ten seconds, then the players are directed to another round, with a different MSA phrase.

Feedback from Players

A questionnaire was administered to 12 players who had tried the game. The players were asked to mention what they liked about the game, what they disliked about the game, and what they would improve. Following the open questions, they were asked to evaluate the time given for each round (30 seconds), and the fun aspect of the game, both using a five-point likert scale.

The results show that many of the players found the game interesting and enjoyed playing: "I enjoyed thinking of different ways of saying the same thing". Players disliked that the rounds were limited to two players only and that a guess might be given a single point only even though it is a correct one. Four players suggested increasing the number of players per round. Two suggested adding other dialects to the game. Some players commented that allowing players to provide many guesses could unjustly earn them points. Eight players found the game fun, 2 found it boring, and 2 found it neither fun nor boring. Ten out of

12 players found 30 seconds time per round was neither too long, nor too short, while 2 of them found it short.

Collected Data

We tested the game using 60 phrases in MSA. A total of 10 Egyptian players provided 621 phrases in the Egyptian dialect. Out of these guesses, only 139 were unique MSA to Egyptian dialect mappings, meaning that the rest (482) were repeated.

It is interesting to find that after counting the number of player outputs (see Figure 2), 21 different guesses (Egyptian dialect phrases) matched across the 10 players. This makes up $21 \times 10 = 210$ guesses, which is roughly 33% of all the collected guesses. It is noteworthy that as much as 40 guesses were only mentioned one time. While the number of guesses that matched across 2, 3, 4 and 5 players gradually decreased till it reached 7 matched guesses. Then increased again as it approached 10 players

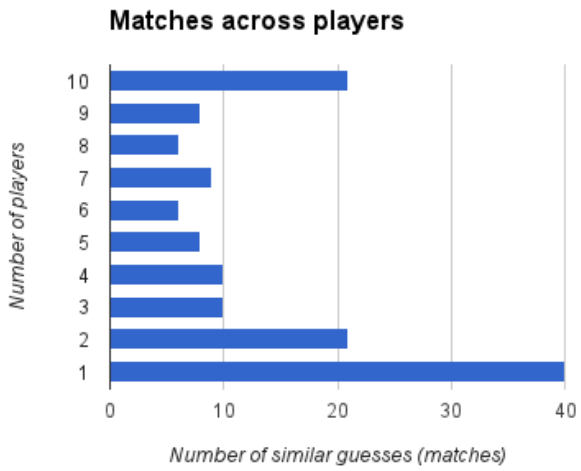

Fig. 2: The number of repeated outputs across players

Looking at a sample of the collected data, we find that for the MSA phrase: "أنا في الغرفة" which means "I am in the room", the most collected Egyptian dialect phrase is: "أنا في الأوضة", which has the same meaning, and uses the same words for "I am in". However, as a result of the Turkish influence on Egypt during the Ottomon rule, Egyptians tend to use the Turkish "أوضة" instead of the MSA "غرفة" for the word "room".

Conclusion & Future Work

In this work, we demonstrated that GWAPs can be used to collect mappings between MSA and regional dialects. A prototype GWAP was implemented. It has indeed collected a big number of results that would have otherwise been expensive and time-consuming to collect.

By analyzing the collected data, we concluded that many of the provided Egyptian dialect phrases were very relevant to the given MSA phrases. However, further analysis should take place to better evaluate the quality of the results of the game. Keeping track of the time spent by every player will enable us to evaluate the GWAP using von Ahn's GWAP evaluation metrics [15], thus estimating the throughput, average lifetime play (ALP) and expected contribution of our players. In our experiments, we did not face problems due to typos and spelling mistakes. However, as our data grows, the probability of facing such problems will be higher. Especially that Arabic is full of visually similar characters such as د and ذ. This will make it harder to match the players' guesses. A possible counter measure is to use a suggestions-box, that would show the players suggestions from previous players' guesses based on the current player's guess, and the MSA phrase in question. This will make it easier to match the guesses, while imposing the threat of influencing the players.

In the future, the same platform could be used to collect similar mappings for different Arabic dialects. Recently, Arabic-speaking Internet users tend to type using what is colloquially known as Franco-Arabic [11]. Franco-Arabic is a transliteration of Arabic using English characters, which also uses some numbers to represent letters that do not exist in the English alphabet. The presented game could be also used with the emerging typing text.

Moreover, such data can help Arabic websites which are traditionally presented in the standard Arabic, to change their interface into the users' dialects to be more user-friendly.

The collected data can be analyzed to extract rules that can automate the process of converting texts using a dialect to/from MSA. For example, Egyptians tend to use the letter (أ) as a replacement for the letter (ق) in MSA, such a rule can be learned to be later applied automatically in translations.

Furthermore, the collected mappings can be used to give a major sprint to translation engines. In many social networks, people of different nationalities interact and each one of them occasionally uses their mother tongue or spoken dialect to pass ideas through their profiles. The collected database can be used to infer rules about dialect-translations. Most of the automated translation techniques available fail to translate the colloquial Arabic dialects to other languages. Having a mapping for each dialect to/from MSA will facilitate the process of translating phrases in different dialects to different languages.

Last but not least, we plan to use the feedback we got from the players to make the game more interesting and popular in order to be able to collect large-enough data sets.

References

1. Arabic Speaking Internet Users Statistics. http://www.internetworldstats.com/stats19.htm. Accessed: 2013-09-27.
2. Internet World Users by Language. http://www.internetworldstats.com/stats7.htm, 2013. Accessed: 2013-09-27.
3. Central Intelligence Agency. The cia world factbook. https://www.cia.gov/library/publications/the-world-factbook/geos/countrytemplate_xx.html, 2009. Accessed: 2013-09-27.
4. Reem Bassiouney. *Arabic sociolinguistics*. Edinburgh University Press, 2009.
5. Michael Bloodgood and Chris Callison-Burch. Using Mechanical Turk to build machine translation evaluation sets. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 208–211, Los Angeles, June 2010. Association for Computational Linguistics.
6. Jon Chamberlain, Massimo Poesio, and Udo Kruschwitz. Phrase detectives - a web-based collaborative annotation game. In *In Proceedings of I-Semantics*, 2008.
7. Mahmoud El-Haj, Udo Kruschwitz, and Chris Fox. Using mechanical turk to create a corpus of arabic summaries. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Valletta, Malta, 2010. In the Language Resources (LRs) and Human Language Technologies (HLT) for Semitic Languages workshop held in conjunction with the 7th International Language Resources and Evaluation Conference (LREC 2010).
8. Onur Gngr and Tunga Gngr. Morphological annotation of a corpus with a collaborative multiplayer game. In *LECTURE NOTES IN COMPUTER SCIENCE*, pages 74–85. Springer, 2010.
9. Jörn Hees, Mohamed Khamis, Ralf Biedert, Slim Abdennadher, and Andreas Dengel. Collecting links between entities ranked by human association strengths. In *ESWC*, pages 517–531. Springer, 2013.
10. Fabian Kneissl and François Bry. Metropolitalia: A crowdsourcing platform for linguistic field research. In *Proceedings of the IADIS International Conference WWW/Internet 2012, Madrid, Spain (18th–21st October 2012)*, 2012.
11. David Palfreyman and Muhamed Al Khalil. a funky language for teenzz to use: representing gulf arabic in instant messaging. *Journal of Computer-Mediated Communication*, 2003.
12. Luis von Ahn. Human computation, 2005.
13. Luis von Ahn. Games with a purpose. pages 92–94, 2006.
14. Luis von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326. ACM, 2004.
15. Luis von Ahn and Laura Dabbish. Designing games with a purpose. *Communications of the ACM*, pages 58–67, 2008.
16. Mark Warschauer, Ghada R El Said, and Ayman G Zohry. Language choice online: Globalization and identity in egypt. *Journal of Computer-Mediated Communication*, 7(4), 2002.
17. Rabih Zbib, Erika Malchiodi, Jacob Devlin, David Stallard, Spyros Matsoukas, Richard Schwartz, John Makhoul, Omar F. Zaidan, and Chris Callison-Burch. Machine translation of arabic dialects. In *The 2012 Conference of the North American Chapter of the Association for Computational Linguistics*, Montreal, June 2012. Association for Computational Linguistics.