A Web and Software-Based Approach Blending Social Networks for Online Qur'anic Arabic Learning

Matin Abdullah[1], Al-Sakib Pathan[2], and Imad Al Shaikhli[2]
[1]Department of Computer Science and Engineering, BRAC University, Bangladesh
[2]Department of Computer Science, International Islamic University Malaysia, Malaysia

Abstract: *About 80 percent of the world's Muslim populations are non-native speakers of the Arabic language. Since it is obligatory for all Muslims to recite Qur'an in Arabic during prayers, an extraordinary social phenomenon has taken place in some parts of the Muslim world: Muslims are taught the complex phonological rules of the Arabic language in the context of Qur'an and they recite the "sounds" of Qur'an often understanding very little. This has given rise to a demographic segment of adult learners whose main learning goal is recalling a closed set of syntactic rules and vocabularies in the context of Qur'an while reciting or listening to it so that they can reconstruct a meaning in their native-language. Despite the availability of some resources for learning language for this specific purpose, according to our detailed investigation, no work has explored the possibilities of emerging adaptive and intelligent systems for collaborative learning to address this challenge. The goals of this work are: (a) To determine the applicability of learner corpus research, declarative memory modelling, and social learning motivation on the learners' specific pedagogical objectives; (b) To use the Design-Based Research methodology (DBR) to optimize the design of such a system in real-life setting to observe how the different variables and elements work out. We present here, a prototype to gather requirement analysis of such a system by bootstrapping a user community. The compiled data were used to design an initial architecture of an intelligent and adaptive Qur'anic Arabic learning system.*

Keywords: *Arabic, language, processing, qur'an, requirement, software, tools, user, web-based.*

Received January 29, 2014; accepted September 9, 2014

1. Introduction

About 80 percent of the world's Muslim populations are non-native speakers of Arabic; according to pew research centre, this constitutes almost one billion people [35]. Muslims consider the Qur'an as the divine revelation, compiled in a book which was recited and preserved in original Arabic language. According to Islamic rulings, it is obligatory for Muslims to read and listen to the Qur'an in Arabic during prayers. There is no other alternative to this as translated version of Qur'an in any other language or even with any kind of alternation in any part is not considered as the Qur'an. Because of this reason, Muslims are required to memorize or learn at least a part of Qur'an in Arabic language. Consequently, an extraordinary social phenomenon has taken place in some parts of the Muslim world-Muslims, men and women, are taught the complex phonological rules of the Arabic language in Qur'anic context. Many of them learn these rules and recite the "*sounds*" of it understanding very little of what they are reciting. Similarly, when listening to the Imam (the person who leads a formal prayer) reciting the Qur'an in prayer, they barely understand what they are listening to [40].

This has given rise to a demographic segment, who are consumers of Arabic language classes, language learning books and software to overcome this specific language barrier. Some resources have been developed to address this particular challenge which include; books, courses, and lexical resources focusing primarily on teaching Arabic language in the context of Qur'an. It should be noted that Qur'anic Arabic is considered as the highest form of Arabic eloquence and grammar; hence, often there are differences between spoken or locally used Arabic and Qur'anic Arabic. The same word in Qur'an may mean different things based on the context and period of revelation.

In the field of second language learning four basic language skills are distinguished. These are ordered along two dimensions [10] :

- Modality: Which is the difference between the auditory language mode versus the visual mode.
- Processing Activity: Which is the process of either encoding or decoding.

Together, these two binary dimensions define the so-called four skills: Speaking, listening, writing, and reading. The learning outcome of the Target Demography (TD) in question is to perform decoding of the visual and auditory modalities-in other words, reading and listening comprehension. Partial visual decoding skills, ability to correctly generate a phonological representation of each word, are present in our TD to the effortless level of automaticity. The learner has to comprehend a new lexical item by assigning meaning to it, thus establishing a new form-meaning connection. According to schema theory [3] the TD have been exposed to formal schema and content schema from their childhood by listening to Friday sermons and religious speeches. Therefore, the essential component missing is the linguistic schema which is the primary impediment to comprehension.

The listening comprehension is a bit more challenging than the reading comprehension because a characteristic of continuous speech is that speech contains no clear auditory equivalent of the inter-word white spaces that we find in written text [9]. The lack of explicit word boundary markers and the distortions of word sounds due to assimilation and reduction are the reasons why words that are known by the language user when presented visually, are often not recognized when they are part of a continuous speech. Although, there are 10 ways of reciting (called *Qira'at*) Qur'an, the most common is "*Hafs*" on the authority of 'Asim, which have been used all over the Muslim world [51]. This wide spread adaptation of Hafs eliminates the complexity of diverse dialects associated with second-language acquisition for the TD.

The main learning goal of the TD is recalling a closed set of syntactic rules and vocabularies only in the context of Qur'an so that they can create a lexical form-meaning connection to reconstruct a meaning in their native-language while reading or listening to its verses. Hence, we hypothesize that the behavioural theories of language learning; as opposed to cognitive theories [27], of stimulus-response can be applied to achieve this learning goal. By virtue of this assertion, we can map the language learning problem into a stimulus-response or form-meaning recognition/recall problem of a closed optimal set of syntactic rules and lexical items. This presents the opportunity to define intelligent rules for our system relating to stimulus factors such as language input, reward magnitude, number, and timing of cues to the strengths of those intervening variables, and rules relating those variables to empirical response measures.

Despite availability of resources for this purpose, according to our investigation, no empirical research has yet identified the learning environments and the unique learning requirements of the TD. This work aims to explore the possibilities of using automated extraction of linguistic patterns from annotated Qur'anic corpora and gathering user stereotypes in the creation of task models to be used in the design and development of an adaptive and intelligent system for collaborative learning [37], that would help the TD of users recall a closed set of syntactic rules and vocabularies in the context of Qur'an. As Design-Based Research methodology (DBR) [8] seeks eventual adoption in real environments, it must be situated in real-life learning environments where there is no attempt to hold variables constant.

This kind of research work has not been attempted before. Hence, our initial goal was to gather a sample of the user-base, who would represent the subsection of our target demography and would become early adopters and co-researchers as a user community. Based on these thoughts, we present here an initial requirement analysis of such a system by utilizing a web-based prototype to bootstrap a user community. The gathered data were used to design an initial architecture of an intelligent and adaptive Qur'anic Arabic learning system.

The rest of the paper is organized as follows: In section 2, we analyze the theoretical background and multitudes of information related to the topic from the perspectives of different disciplines-all of these are required for our computational modelling, section 3 talks about our approach based on our thorough analysis, section 4 proposes our design of the system, section 5 presents the practical implementation-based results. Finally, section 6 concludes the paper delineating the achievements of this work along with some suggestions for future works.

2. Theoretical Background and Analysis

2.1. Role of Computational Corpus Linguistics in Language Pedagogy

Going back in time, the early 1990s saw an increasing interest in applying the findings of corpus-based research to language pedagogy [38]. These works cover a wide range of issues related to using corpora in language pedagogy, e.g., corpus-based language descriptions, corpus analysis, and learner corpus research. There have been indirect and direct usages of computers and corpora in language teaching.

2.1.1. Pattern Extraction from Language Corpora

One of the key contributions of learner corpus research is lexical approach to teaching language [18]. A lexical approach to teaching means the focus is on helping students acquire vocabulary. This movement away from a grammar-based syllabus, largely began in 1993 with the publication of "*The Lexical Approach*" by Lewis [34]. What it focuses on are structures made up of words, meaning that the actual paradigm shift was away from learning individual words to clusters of words, or lexical chunks as they are commonly. Hunston [29] describes an approach to lexical and grammatical description that uses large amount of corpus data to make discoveries about lexical items and the specific phraseological and grammatical patterns, otherwise known as "*collocation*". Thus, within lexical approach, two major types of collocations are defined: Lexical and grammatical collocations [20]. Grammatical collocations contain the combinations of a noun, a verb or an adjective, and a preposition or a grammatical structure [62]. In contrast, lexical collocations do not consist of grammatical elements, but are combinations of nouns, adjectives, verbs, and adverbs [33].

There have been some research works in defining collocation in Modern Standard Arabic (MSA) [17, 26] and use of collocations in teaching Arabic as a foreign language [2, 4, 24]. Researchers have also looked into automated extraction of collocations from Arabic corpora [25, 57]. Although, there is a dominance of the statistical paradigm in Natural Language Processing (NLP) in all areas including extraction of collocations, since we have tagged corpora of the Qur'an [14, 57], available hybrid techniques combining rule-based and statistical methods can be applied for better results. As we have investigated the area so far, the first automated morphological part-of-speech tagging of the Qur'an [57] was done at the university of Haifa, Israel. The annotation in the Haifa Qur'anic corpus was produced automatically using a rule-based morphological tagger, following a generative approach. Using a published concordance of Qur'an, a list of base-word forms was selected. An inflectional generator was then encoded into a Finite State Machine (FSM), consisting of 50 morphological rules for nouns, and 300 rules for verbs. The published dataset which resulted from the study is a list of possible analyses for each Arabic word in the Qur'an. The authors of the study report that about 70% of the words in the Qur'an received a unique morphological analysis, with the remaining words having several possible analyses each.

The second automated morphological part-of-speech tagging of Qur'an was done at the university of Leeds, UK [14]. In this Qur'anic Arabic corpus, initial automatic tagging was carried out using a modified version of the Buckwalter Arabic Morphological Analyzer (BAMA) [36], adapted to the unique language of Qur'an. This was then followed by several stages of manual corrections. After this initial process, the Qur'anic Arabic corpus was improved through a model of collaborative annotation, where volunteers proofread part-of-speech tagging and syntactic analysis, and then suggested corrections through an online message-board forum [13]. Finally, syntactic dependencies between words and phrases were annotated using Arabic syntactic dependency labels. The annotated corpus also included further additional tagging, including named entity references-an ontology of semantic concepts and an automatically generated phonetic transcription.

2.1.2. Interactive Corpus Interface for Learning Lexical Items

While indirect usages, such as syllabus design and materials development (by computationally extracting patterns from a large body of corpora) are closely associated with what to teach, sophisticated interfaces for learner interaction with corpora have also provided valuable insights into how to teach. Input flood treatment includes the artificially increased incidence of the target items in the audio and visual texts that learners are exposed to, without any explicit instruction or feedback [43]. Thus, learning the target items in such a treatment is a by-product of the reading/listening comprehension activity. Wong [61] points out that the basic idea in this kind of treatment is that flooding the input (language data which the learner is exposed to) with the target items increases the chance of noticing those items since it is highly likely that learners attend to something that occurs frequently in the input.

One of the biggest challenges in Language for Specific Purposes (LSP) is helping students acquire the vocabulary and collocations that they need to begin reading in a subject area. Students typically need to know words and their different morphological forms that are measured in the thousands, not hundreds, but receive language instruction measured in months, not years. In this time-squeeze, vocabulary course developers choose between breadth (explicit learning of words on lists) and depth (implicit learning of words through extensive reading). But list-learning creates superficial knowledge, and acquisition through reading is too slow for the time available. The literature of vocabulary acquisition is virtually unanimous on the value of learning words through several contextual encounters [42]. This paradox has been viewed as being un-resolvable using traditional learning technologies, but advances in Human Computer Interaction (HCI) suggest new possibilities.

Computerized concordances can help resolve the breadth-depth paradox. This method combines the benefits of list coverage with at least some of the benefits of lexical acquisition through natural reading. Cobb's [11] work is the only one which has empirically tested the effectiveness of corpus-based techniques to teach vocabulary. Cobb compared the vocabulary learning outcomes of his students when new words were learned by viewing multiple concordance lines vs. a single sentence accompanied by a short definition of the word, and found that viewing concordance lines leads to small but consistent gains in his students' vocabulary knowledge. Cobb, in PET-2000 algorithm, pioneered motivation to use corpus-based concordance software extensively to resolve the breadth-depth paradox of language learning [31].

2.2. Computational Modelling of Declarative Memory

By mapping the language learning problem into a recognition/recall problem, we can harness the advances in computational modelling of declarative memory to optimize the lexical item acquisition of our TD. Research on lexical item acquisition demonstrates that systematic rehearsal is essential for successful learning. Scheduling of rehearsing opportunities is usually divided into two types: Spaced learning and massed learning. In 1886, Hermann Ebbinghaus, a German psychologist who first pioneered the experimental study of memory, demonstrated that there is a predictable and measurable way by which people forget what they have learned and there is an optimal time to review what one has learned. Over 100 years of longitudinal experimental studies have confirmed this basic hypothesis of "*an optimal moment of review*". For a given amount of study time, spaced learning yields significantly higher recall rates than massed learning-a phenomenon referred to as a spacing effect [5, 6, 22, 39, 42].

Many studies have examined computer-controlled sequencing algorithms for optimized scheduling of rehearsal opportunities by keeping a record of the learner's performance on individual lexical items and presenting the item that needs to be rehearsed most at the time [16, 21, 42]. Computers can easily be programmed to keep track of a learner's performance and control the sequencing of items to make sure that unknown or hard items are studied more frequently than known or easy items [41, 49].

The most sophisticated psychological modelling of this process is derived from the Adaptive Control of Thought-Rational (ACT-R) activation-based model of declarative memory; where each test of an item introduces a new memory trace whose decay rate is a power law function of all traces for the item at the time of the test [45, 56]. Here, the power law function is the mathematical equation modelling the rate of forgetting as a function of time and memory activation.

The Fact and Concept Training (FaCT) system for [46, 47] uses the ACT-R [64] cognitive model to predict the best item to test at any point in time, introducing new items when no existing items are near their optimum point for review. This aims to maximize the long-term recall gains of each review, rather than keeping the recall likelihood of all items above a threshold level. Other algorithms like SuperMemo [63] and Anki [21] target a fixed retention rate. The advantages of these adaptive spaced repetition systems are that they evolve according to the learner's performance over time. MemReflex is a mobile-based, adaptive and context-aware flashcards system for mobile micro-learning [15]. It takes into account learner history as well as the immediate cues of time, location, and motion. The algorithm exploits both text and audio interaction modalities to engage learners with different learning styles as well as learners who move between contexts where different modalities are most appropriate.

2.3. The Role of Social Networking Platforms in Learner Motivation

According to Houle [28], apart from the intrinsic goal of learning new knowledge or achieving a certain goal, learners may be motivated by social reasons-to seek social contacts. The wish for peer or teacher-recognition, or achieving high reputation in the group, falls under "*social*" motivation. Incentive mechanisms can capture social and extrinsic motivations. In designing such mechanisms, one has to create a payoff matrix that defines the rewards for particular actions that are aligned with the learner's goal, but also with the teaching goal and certain social/community goals. For example, many social sites work based on a reputation economy [19]. Therefore, a social learning environment needs to:

- Create a feelings of achievement/self-actualization.
- Tie learning more explicitly to social achievement related to status/reputation in the peer group.
- Tie learning more explicitly to social rewards in terms of marks and credentials.

The notion of peer ratings in education is supported by educational research in the area of peer assessment. Johnston and Miles [32] found that students took peer assessment seriously and Pope [48] found that both peer and self-assessment contribute positively to a student's course performance. Research also suggests that peer assessment can help motivate student participation and foster student initiative to learn [50]. The task of designing a mechanism in a learning/educational setting needs to consider the utility or the personal goals of learners: Social presence, social interaction, and such.

A number of theories look at the role that people play in an online learning environment, including connectivism, social constructivism, behaviourism, social learning, situated learning, and social presence. Social presence theory also considers the degree to which an individual's perception of an online community, in its entirety, affects his or her participation in that community [54]. In other words, social presence refers to a communicator's sense of awareness of the presence of an interaction partner. Within human-computer interaction, social presence theory considers how "*sense of community*" is shaped and affected by technological interactions [59].

In this work, we propose to design, construct, and implement an online peer assessment system based on current social networking systems as a source of motivation for our target demography.

2.4. Context-Aware Mobile Learning

There have been significant number of research works in measuring cognitive load of a learner using sensors to detect changes in blink rate, blink duration, skin conductivity, peripheral temperature (finger, wrist, ambient), heart rate, blood oxygenation and mouse pressure in different e-learning systems [23, 30]. Cognitive load can be used as a metric to determine when or whether should an educational system intervene and adapt to the user. Although, the goal of context-aware adaptive mobile learning is to help the mobile learners to increase his satisfaction and learning, the aforementioned sensors have mainly been studied in mobile devices in the domain of medical applications [44].

The diversity of mobile and wireless technologies and the nature of dynamics in various mobile environments make context detection a challenging task. Context-aware mobile learning puts emphasis on mobile learners who are carrying portable devices, such as smart phones, that have been augmented with hardware sensors. Apart from psycho-physiological context, computing context (e.g., network connectivity, communication costs, communication bandwidth, nearby resources such as printers, displays, and workstations), user context (e.g., user's profile, location, people nearby, and current social situation) and physical context (e.g., lighting, noise levels, traffic conditions, and temperature) have also been proposed as relevant context categories for learners [53]. For instance, the trade-off between walking speed and interaction performance has been investigated [7] and as an alternative to visual feedback, audio feedback while walking has been shown to result in higher interaction accuracy as well as lower mental and physical demands [60].

3. Our Research Approach

One of the key objectives of this work is to contribute to the existing body of research in the field of intelligent and adaptive system for collaborative online learning. To achieve this goal, we adopt the DBR methodology [8]. In DBR, a researcher is mainly concerned about the way things should be in order to attain specific goals. To achieve the goals, a researcher devises artifacts (in our case, artifact refers to an object in a digital environment). DBR seeks eventual adoption in real environments and therefore, must be situated in real-life settings where there is no attempt to hold variables constant.

Another important aspect in this method is that a DBR team's methodological approach changes with the development of theoretical knowledge as initially unpredicted observations arise among predicted ones, which leads to intervention designs that are better fit to their intended settings [55]. Collins *et al.* [12] related such a methodology to the term "*Design Sciences*" coined by Simon [55], as opposed to "*Analytic Sciences*", which is associated with typical experimental versus-control group studies.

DBR researchers try to optimize the design as much as possible to observe how different variables and elements may work out. Under such a methodology, the learning design-enactment-reflection-refinement (or, invention-revision) cycles are iteratively conducted; thus, as conjectures are generated and perhaps refuted, new conjectures are developed in the next cycle and again, they are subjected to further testing. Achieving such a design science outcome requires sufficient understanding of the underlying variables at all relevant layers of a complex social system.

Therefore, one of the initial steps of DBR is identifying and in case of a new theory development, setting such an environment as natural laboratory to study intervention. The primary goal of our initial prototype described in this paper is to bootstrap such a natural laboratory with the target user-group so that they become stakeholders in our future iterations of the system. This helps us conduct an initial needs analysis of the system with the correct TD and gather insight about the user-base and the system requirements.

4. Our Proposed Design

The proposed system can be divided into four modules:

- Lexical Pattern Extraction Module.
- Learner Model Update Module.
- Social Learning Module.
- Presentation Module.

4.1. Lexical Pattern Extraction Module

This is an independent and offline module designed to facilitate the language expert to prepare relevant lexical items, their dependencies and concordances in the format that can be used by the main system. The expert will have an interface where s/he can enter grammatical patterns for the system to extract those patterns from the corpora. S/he can also ask for clusters of lexical items with similar features for the system to generate from a certain segment of the corpus. Then s/he would determine the validity of the lexical item and edit the presentation form for the learner depending on the grammatical patterns based on the lexical approach [18]. At this time, system will generate the transliteration of the particular item using the learner's native alphabet for the expert to verify.

The expert will then categorize and place the lexical item into a relevant learning group. The system will auto-generate the smaller lexical chunks that this item is dependent on, so that the expert can verify the items. For instance, for an idiomatic phrase, the expert will unlink the dependencies of the smaller chunks. Then, the system will display the concordance for that specific lexical pattern and the expert will choose the relevant examples of that pattern to be presented to the student during the exposure to that learning pattern.

Figure 1 shows the interactions and relationships among various entities in our Qur'anic Arabic learning system. This system will have a default idiomatic and interlinear translation, transliteration and recitation sound file for the particular example. The expert will be able to change and edit some of this information. The examples will be shown to the user during learning sessions to bridge the breath-depth paradox of learning lists of lexical items. After the verification of the expert, these data will be preserved in the pedagogically relevant lexical item database with all the dependencies and examples from the Qur'an. Figure 2 shows a prototype of the lexical pattern extraction module built for the system.

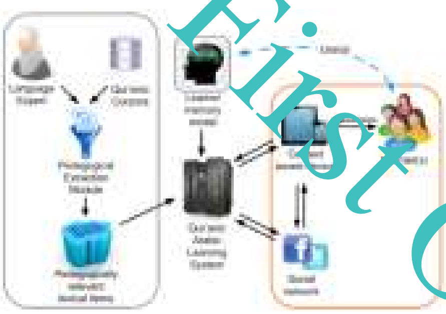

Figure 1. Model of Qur'anic arabic learning system.

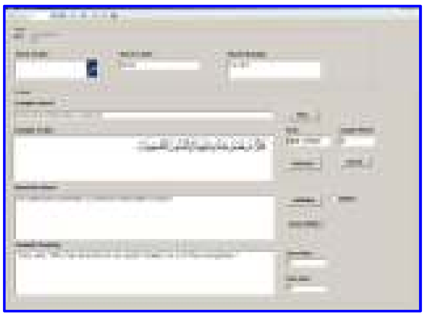

Figure 2. Lexical pattern extraction module.

4.2. Learner Model Update Module

This module will update the learner's memory model as shown in Figure 3 as the user interacts with the system. Depending on the strength of each lexical item in the learner's memory model, this module will schedule the Learn, Review and Test (LRT) operations in the most efficient way to guarantee that the user achieves the desired degree of learning within the desired time period. The scheduling and functionality of the operations will adapt and change based on the user's performance. The scheduler will eliminate all the cognitive workload of planning the LRT operations which are normally the responsibility of the user so that the learner can completely focus on the tasks.

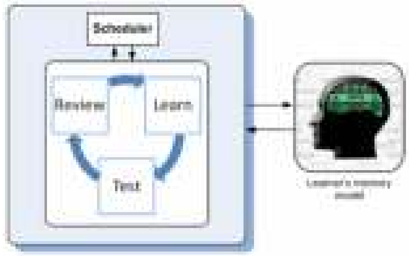

Figure 3. Learner model update module.

4.3. Social Learning Module

The social learning module would integrate the system with social networking platforms like Facebook and Twitter: To create a learners' community and to motivate the learner. The role of the social integration will be to provide feedback on a user's performance and comparison with the performances of other learners to stimulate social competition. The user's activity and progress will be shared with the friends of the users. By allowing participants to post their progress to their Facebook newsfeed, this module will enable social support from friends, in the form of structured support in Facebook (e.g., "*liking*" someone's progress or commenting on it) as well as unstructured interactions between individuals through other communication channels. With this module, the system would make learning more gratifying by [58]:

- Making it game-like, a mixture of challenge and fun.
- Enhancing the feeling of success by providing constant feedback on performance.
- Communicating performance in peer community, and comparing performances among peers.

4.4. Presentation Module

The presentation module would utilize the context-aware data gathered from the learner's device and customize the learning experience sensitive to learner history as well as the immediate cues of time, location, and motion. For example, it might be safer to listen via headphones when navigating busy public spaces, but politer to read from the screen in social situations where some conversation is anticipated [15].

5. Initial Prototype-based Results

The goal of this initial prototype [1] was to conduct an initial needs analysis of software that focuses on teaching Arabic of Qur'an and gather insights about the user-base and the system requirements. Initial software implementation was designed to accompany the understandquran.com website that teaches the language of Qur'an by mainly focusing on vocabulary and morphology of words, and some syntax.

5.1. Implementation

We implemented the prototype Lexical pattern extraction module using the C# programming language as shown in Figure 2. This module was developed for pre-processing and generating data for the test and the learning module. The language expert would enter the canonical form of the word and the program auto-generated the transliteration in English. For the particular example, the expert could choose a verse from the Qur'an, where the word occurred, and then select both the interlinear and idiomatic translation of the verse. The priority was to choose an example where the Arabic word is in its canonical form but in case of its absence in the Qur'an, other forms were used. The expert could choose any number of adjacent tokens to create a collocation and enter its interlinear translation. S/he would also select an audio file for this example. All the data were stored in a backend database. Because of the differences in Arabic font support in different web browsers, this system generated all the Arabic tokens as images which were then combined with the audio file into an Adobe Flex downloadable module.

A prototype test driven module was created using Adobe Flex as shown in Figure 4. The test module was divided into multiple quiz modules each containing 12 vocabulary questions. Before beginning the test, the user could review all the words for that module by accessing the list of words and their meanings for the chosen quiz module.

Figure 4. Prototype test interface.

In the test module, the target Arabic word was presented in its canonical form and one correct answer with three distracters was presented to the user. If the user answered correctly, the next question was presented. In the case of a wrong answer, the sequence of the answers was randomly changed so that the user cannot sequentially try out all the answers without paying proper attention. Since, the prototype was designed to promote test driven learning, the user could choose to access an example of the current word being tested as a hint. The example presented:

- The target word in the context of a sentence in Arabic with the usage of the word being highlighted in red.
- Interlinear translation of the lexical collocations and words around the target word.
- An idiomatic translation of the whole sentence.
- Audio recitation of the whole sentence.

This prompted the user to guess the meaning of the word from the context of the sentence and neighbouring lexical collocations, promoting incidental vocabulary learning [42]. After finishing the quiz, the users were presented with their performance data, for the completed quiz module. The performance displayed each row containing the Arabic word, its meaning and the number of trials it took the user to respond with a correct answer. Social media integration with Facebook, Twitter and StumbleUpon was implemented to build a user community. An email subscription system and a Facebook page were set up to collect user information and to notify the user of new content when it became available. An Youtube video describing the goal and interaction of the software was published as a how-to guide for the website.

5.2. Creating User Community

We used email lists, forums, and Facebook pages to inform the potential users about our page. Other sites helped us spread the information for example, the original understandquran.com site, oxford university language centre (http://www.lang.ox.ac.uk/links/arabic.html) etc. The users of the site also helped us by posting about the site in their blogs, Facebook pages and Twitter feeds. During the data collection period for the prototype, using Google analytics, there were 82,655 unique visitors who visited the site and total number of pageviews during this period was 174,586. Because the whole page loaded as an adobe flex application, the user interaction with different quiz modules and examples were not counted as different pageviews by Google analytics. There were in average 3000-4000 users visiting our site per month and the user visits doubled when the social network integration was completed. The month in which over 10,000 users visited the site was during Ramadan (the Muslim holy month of fasting).

5.3. User Demography

We gathered some basic demographic data from both the users who visited the site and the users who subscribed to the Facebook page. We have found a stark difference between the top ten countries where the visitors came from and the top ten countries of user-subscription. This refers back to our assumptions about the target demography: The site visitors include a lot of incidental visitors but among the people who subscribed to the site, the majority is from Muslim countries where the native language is not Arabic and the tradition of teaching children how to recite the

Qur'an is very strong. Almost 80% of the subscribers are in the age range of 18-34. However, among that group, the age range of 18-24 constitutes about 50% of all the subscribers. This indicates that young people are early adopters of this technology.

5.4. User Environment

The user environments were predominantly desktop browsers because the adobe flash player was not supported on mobile devices during the prototype testing. The major shift we are seeing in the user behavior is towards mobile devices. Although, the site did not support mobile browsing, the percentage of visitors trying to access the site with mobile devices was gradually increasing. We tested a prototype: Word-a-day component built using HTML5 applying responsive design for different mobile form-factors. This resulted in more than 60% users accessing this component using mobile devices.

5.5. User Perception

A total of 321 comments were collected from the users, including emails, blog comments, Facebook comments, and forum replies. The overall response from the user community was very positive. Following are some of the issues and limitation that came up in the user comments.

In almost 34% of the comments, the users wanted to store their progress information and share it online with their friends. The primary goal of our next prototype is to implement these two components.

Since, it was an initial prototype, more than 50% of the comments had some element suggesting the lack of content. Since, the comprehensive analysis of the corpus was not done for the prototype, we only tested some of the modules by collecting data available from the web. In our future implementation by using the lexical pattern extraction module, we would like to generate more comprehensive material for the system.

Around 70% of the comments asked for the mobile device implementation of the test module. This is in line with the global trend of user-shift from the desktop environment. Although small, around 6%, another significant comment was the request for offline implementation of the software. This was due to the fact that the users situated in developing world did not have affordable access to Internet facilities.

Finally, more than 20% of the comments included requests for different language implementation of the prototype which is a major limitation of our implementation. The highest language requests, in order, were for Urdu, Malay, French, and Turkish.

6. Conclusions and Future Works

The main focus of this study was to examine the potential for a learning system for learning the Arabic of Qur'an and to bootstrap a learners' community of the target user group so that they become stakeholders in our future iterations of the system. We have successfully created an online user community of almost 10,000 users who have subscribed to receive updates of the project by different means.

The implication, i.e., the conclusion that can be drawn from this work could be summarized as:

- We discovered that there is an overwhelming need and potential for software of this nature.
- There is a growing need for mobile accessibility for the purpose of e-learning.
- It is understood that user locations are often distributed and very diverse.
- There is a need for better user modelling to customize the user experience and give users an opportunity to track their progress.
- A deeper integration with the social media is needed so that the user has control over sharing the user data with their friends on the social network to create an environment of social learning.

While this study was exploratory in nature, blended multiple disciplines, and had a limited scope, we have only just begun to explore how technology, especially from the angle of Information and Communications Technology (ICT) can facilitate the learning goal of this target demography.

7. Acknowledgments

Supported partially by RMC research grant EDW B13-084-0969, Project Title: "Teaching and Learning Qur'anic Arabic Utilizing Adaptive and Intelligent Systems for Collaborative Learning". Please direct all correspondences to the Principal Investigator, Al-Sakib Khan Pathan.

References

[1] Abdullah, M. S. and Pathan, A.-S.K., "Learning Qur'anic Arabic through Interactive Web-Based Software: A Pragmatic Approach in Language for Specific Purpose," *The 4th ICT4M*, March 25-27, 2013, pp. 1-6.

[2] Al-Yaari, S.A.S., Al Hammadi, F.S., and Alyami, S. A., "Written Grammatical Errors of Arabic as Second Language (ASL) Learners: An Evaluative Study," *Int. Jrnl of English Language Education*, vol. 1, no. 2, p. pp–143, 2013.

[3] An, S., "Schema Theory in Reading," *Theory and Practice in Language Studies*, vol. 3, no. 1, pp. 130–134, 2013.

[4] Árabe, C., "Teaching Collocation in the Arabic as a Foreign Language Class," *Enseñanza y aprendizaje de la lengua árabe*, p. 163, 2010.

[5] Baddeley, A. D., *Human memory: Theory and practice*. Psychology Pr, 1997.

[6] Bahrick, H. P., and Phelps, E., “Retention of Spanish vocabulary over 8 years,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 13, no. 2, pp. 344–349, 1987.

[7] Bergstrom-Lehtovirta, J., Oulasvirta, A., and Brewster, S., “The effects of walking speed on target acquisition on a touchscreen interface,” in *Proc. of the 13th MobileHCI*, 2011, pp. 143–146.

[8] Boticki, I., Wong, L. H., and Looi, C.-K., “Designing Technology for Content-Independent Collaborative Mobile Learning,” *IEEE Trans. on Learning Tech.*, v. 6, no. 1, pp. 14–24, 2013.

[9] Buck, G., *A guide for the teaching of second language listening*. SD: Dominie Press, 1995.

[10] Byrnes, H., *The Role of Listening Comprehension: A Theoretical Base*. Foreign Language Annals, vol 17, 1984.

[11] Cobb, T., “Is there any measurable learning from hands-on concordancing?,” *System*, vol. 25, no. 3, pp. 301–315, 1997.

[12] Collins, A., Joseph, D., and Bielaczyc, K., “Design research: Theoretical and methodological issues,” *The Journal of the learning sciences*, vol. 13, no. 1, pp. 15–42, 2004.

[13] Dukes, K., E, A., and Habash, N., “Supervised Collaboration for Syntactic Annotation of Quranic Arabic,” *Language Resources and Evaluation*, 47 (1), pp 33-62, 2013.

[14] Dukes, K., and Habash, N., “Morphological annotation of quranic Arabic,” *Language Resources and Evaluation Conference*, 2010.

[15] Edge, D., Fitchett, S., Whitney, M., and Landay, J., “MemReflex: adaptive flashcards for mobile microlearning,” in *Proc. of the 14th MobileHCI*, 2012, pp. 431–440.

[16] Ellis, N. C., “The psychology of foreign language vocabulary acquisition: Implications for CALL,” *Computer Assisted Language Learning*, vol. 8, no. 2–3, pp. 103–128, 1995.

[17] Emery, P. G., “Collocation in modern standard Arabic,” *Journal of Arabic Linguistics*, vol. 23, no. 1, pp. 56–65, 1991.

[18] Flowerdew, L., *Corpus-based research and pedagogy in EAP: From lexis to genre*. Language Teaching, 2012.

[19] Forte, A., and Bruckman, A., “Why do people write for Wikipedia? Incentives to contribute to open–content publishing,” *Proc. of GROUP*, vol. 5, pp. 6–9, 2005.

[20] Ganji, M., “The effects on reading comprehension of lexical collocation instruction, subject matter knowledge, and cultural schema,” *Theory and Practice in Language Studies*, vol. 2, no. 1, pp. 105–112, 2012.

[21] Godwin-Jones, R., “Emerging technologies from memory palaces to spacing algorithms: Approaches to second-language vocabulary learning,” *Language Learning and Technology*, vol. 14, no. 2, pp. 4–11, 2010.

[22] Goverover, Y., Hillary, F. G., Chiaravalloti, N., Arango-Lasprilla, J. C., and DeLuca, J., “A functional application of the spacing effect to improve learning and memory in persons with multiple sclerosis,” *Journal of Clinical and Experimental Neuropsychology*, vol. 31, no. 5, pp. 513–522, 2009.

[23] Haapalainen, E., Kim, S., Forlizzi, J. F., and Dey, A. K., “Psycho-physiological measures for assessing cognitive load,” in *Proc. of the 12th ACM ICUC*, 2010, pp. 301–310.

[24] Hassan, H., and Daud, N. M., “Corpus Analysis of Conjunctions: Arabic Learners’ Difficulties with Collocations,” [presentation slides], 2011.

[25] Hawwari, A., Bar, K., and Diab, M., “Building an Arabic Multiword Expressions Repository,” *ACL 2012*, p. 24, 2012.

[26] Heliel, M. H., “Collocations and Translation,” *Proceedings of the FIT Round Table “Professional Arabic Translation and New Technologies”, Tanger*, 1989.

[27] Hollan, P. C., “Cognitive versus stimulus-response theories of learning,” *Learning and behavior*, vol. 36, no. 3, pp. 227–241, 2008.

[28] Houle, C. O., *The inquiring mind*. University of Wisconsin Press Madison, 1961.

[29] Hunston, S., *Corpus Linguistics: Historical Development*. The Enc. of App, Ling,, 2013.

[30] Ikehara, C. S., and Crosby, M. E., “Assessing cognitive load with physiological sensors,” in *Proc. SS 2005. HICSS’05.*, 2005, p. 295a–295a.

[31] Jafarpour, A. A., Hashemian, M., and Alipour, S., “A Corpus-based Approach toward Teaching Collocation of Synonyms,” *Theory and Practice in Language Studies*, 3(1), pp. 51–60, 2013.

[32] Johnston, L., and Miles, L., “Assessing contributions to group assignments,” *Assess. and Eval. in H. Ed.*, 29(6), pp. 751–768, 2004.

[33] Le, T. H., *Learning lexical collocations with concordancing and scaffolding*. Retrieved December 20, 2011.

[34] Lewis, M., *The lexical approach*, vol. 1. Lang. Teaching Publications Hove,, UK, 1993.

[35] Lugo, L., “The Future of the Global Muslim Population,” Pew Research Center, USA, 2011.

[36] Maamouri, M., Bies, A., Buckwalter, T., and Mekki, W., “The penn arabic treebank: Building a large-scale annotated arabic corpus,” in *NEMLAR Conference on Arabic Language Resources and Tools*, 2004, pp. 102–109.

[37] Magnisalis, I., Demetriadis, S., and Karakostas, A., “Adaptive and Intelligent Systems for Collaborative Learning Support: A Review of the Field,” *IEEE TLT*, vol. 4, no. 1, pp. 5–20, 2011.

[38] McEnery, T., and Xiao, R., “What corpora can offer in language teaching and learning,”

Handbook of Research in Second Language Teaching and Learning, vol. 2, 2010.

[39] Milton, J., *Measuring second language vocabulary acquisition*. Cambridge Press, 2009.

[40] Moore, L., *Learning by Heart in Qur'anic and Public Schools in Northern Cameroon*. Social Analysis, 50(3), pp 109-126, 2006.

[41] Nakata, T., "English vocabulary learning with word lists, word cards and computers: implications from cognitive psychology research for optimal spaced learning," *RECALL-HULL THEN CAMBRIDGE-*, vol. 20, no. 1, p. 3, 2008.

[42] Nation, I., and Chung, T., "28 Teaching and Testing Vocabulary," *The handbook of language teaching*, p. 543, 2009.

[43] Öztina, S., *Effects of Input Flood and Negative Evidence on Learning of Make/Do Collocations: A Study with Seventh Grade Turkish EFL Students*. Master degree dissertation, Anadolu Univ. Institute of Educational Sciences, 2009.

[44] Pantelopoulos, A., and Bourbakis, N. G., "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Trans. on Sys., Man, and Cyber.*, 40(1), pp. 1-12, 2010.

[45] Pavlik, P. I., and Anderson, J. R., "An ACT-R model of memory applied to finding the optimal schedule of practice," *6th ICCM 2004*, 2008.

[46] Pavlik, P. I., Cen, H., and Koedinger, K. R., "Performance factors analysis–a new alternative to knowledge tracing," in *Proc. of the 14th AIED*, 2009, pp. 531–538.

[47] Pavlik Jr, P. I., Presson, N., Dozzi, G., Wu, S., MacWhinney, B., and Koedinger, K., "The FaCT (Fact and Concept Training) System: A new tool linking cognitive science with educators," in *Cognitive Science Society*, 2007.

[48] Pope, N.K.L., "The impact of stress in self-and peer assessment," *Assess. and Eval. in H. Ed.*, 30(1), pp. 51–63, 2005.

[49] Pyc, M. A., and Rawson, K. A., "Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory?," *Journal of Memory and Language*, vol. 60, no. 4, pp. 437–447, 2009.

[50] Rafiq, Y. and Fullerton, H., "Peer assessment of group projects in civil engineering," *Assess. and Eval. in H. Ed.*, vol. 21, no. 1, pp. 69–81, 1996.

[51] Razak, Z., Ibrahim, N. J., Idna Idris, M., and others, "Quranic Verse Recitation Recognition Module for Support in J-QAF Learning: A Review," *IJCSNS*, 8(8), pp. 207-216, 2008.

[52] Saif, A. M., and Aziz, M. J., "An automatic collocation extraction from Arabic corpus," *J. of Computer Science*, vol. 7, no. 1, pp. 6–11, 2010.

[53] Schilit, B., Adams, N., and Want, R., "Context-aware computing applications," in *Mobile Computing Systems and Applications, 1994. WMCSA*, 1994, pp. 85–90.

[54] Short, J., Williams, E., and Christie, B., "The social psychology of telecommunications," 1976.

[55] Simon, H. A., *The sciences of the artificial*. MIT press, 1996.

[56] Son, L. K., and Simon, D. A., "Distributed Learning: Data, Metacognition, and Educational Implications," *Educational Psychology Review*, vol. 24, no. 3, pp. 379–399, 2012.

[57] Talmon, R., and Wintner, S., "Morphological Tagging of the Qur'an," in *Proc. of the Workshop on Finite-state Methods in Natural Language Processing, 10th Conference of the European Chapter of the Association for Computational Linguistics*, 2003, pp. 67–74.

[58] Thoms, B., "A Dynamic Social Feedback System to Support Learning and Social Interaction in Higher Education," *IEEE TLT*, 4(4), pp. 340-352, 2011.

[59] Tu, C.-H., and McIsaac, M., "The relationship of social presence and interaction in online classes," *The American journal of distance education*, vol. 16, no. 3, pp. 131–150, 2002.

[60] Wilson, G., Brewster, S. A., Halvey, M., Crossan, A., and Stewart, C., "The effects of walking, feedback and control method on pressure-based interaction," in *Proc. of the 13th MobileHCI*, 2011, pp. 147–156.

[61] Wong, W., *Input enhancement: From theory and research to the classroom*. Boston: McGraw-Hill, 2005.

[62] Wu, S., Franken, M., and Witten, I. H., "Supporting collocation learning with a digital library," *Computer Assisted Language Learning*, vol. 23, no. 1, pp. 87–110, 2010.

[63] Yang, J., and Park, J., "Software Review: SuperMemo UX-Extreme English Advanced and Proficient," *CALICO Journal*, vol. 29, no. 4, pp. 718–726, 2012.

[64] Zylberberg, A., Dehaene, S., Roelfsema, P. R., and Sigman, M., "The human Turing machine: a neural framework for mental programs," *Trends in cognitive sciences*, 15(7), pp. 293–300, 2011.

Matin Abdullah is pursuing Ph.D. degree in Computer Science at IIUM, Malaysia. He received M.Sc. degree in Computer Science in 2002 from University of Houston, USA. He received B.Sc. degree in Electrical Engineering from the same University. He is also currently an Assistant Professor in Computer Science and Engineering department and Associated Director of Centre for Research on Bangla Language Processing, BRAC University, Bangladesh.

Al-Sakib Pathan received Ph.D. in Computer Engineering in 2009 from Kyung Hee University, South Korea and B.Sc. in Computer Science and Information Technology from Islamic University of Technology (IUT), Bangladesh in 2003. He is currently an Assistant Professor at the CS department in IIUM, Malaysia. He is actively involved in numerous research events and journals. He is a Senior Member of IEEE, USA.

Imad Al Shaikhli is an Associate Professor at the Department of Computer Science, International Islamic University Malaysia, Malaysia. He is currently serving as the Head of Research (HoR) at the Kulliyyah (Faculty) of Information and Communication Technology. His research interest includes cryptography, steganography, genetic algorithms, neural networks and biometrics. He is a Senior Member of IEEE, USA.