mScrabble: A Multilingual Scrabble and Lexicon Collection Generator

Ritesh Kumar[1, 2], Bornini Lahiri[3], Siddharth Singh[1]
[1]Centre for Transdisciplinary Studies, Dr. Bhimrao Ambedkar University, India
[2]Department of Linguistics, KMI, Dr. Bhimrao Ambedkar University, India
[3]Department of Humanities and Social Sciences, Indian Institute of Technology - Kharagpur, India
ritesh78_llh@jnu.ac.in, bornini@hss.iitkgp.ac.in, sidd435@gmail.com

Abstract

In the situation of language endangerment, especially because of various kinds of pressure from surrounding majority languages and a low language prestige among the community members, language games of various kinds could prove to be an effective tool enhancing the prestige, providing an additional domain of language use to the community members and also for the researchers working with the communities for language documentation and possibly revitalisation. Keeping these in mind, we have developed a word game - mScrabble, a substantially changed and adapted version of the popular game of Scrabble for a large number of languages - as a mobile app. In this paper, we present this game for two endangered Indian languages - Koda and Mahali - and discuss its features and rules, its technical specifications and its initial reception in the community. We also present a generic generator of this game which, given a word list and a few translations (of the items on the interface), could generate the game for any language. In addition to its primary function as a game, it could also be utilised for collecting lexical items from the community.

1 Introduction

While working with endangered languages, the community members (and the researchers as well) face two-fold problem -

- If the language is critically endangered and the number of speakers is sparse, speakers may not have sufficient opportunity to use the language in different domains. A similar situation could also be witnessed in those situations when the speakers have a negative attitude towards the language, and the language, in general, has a low prestige. In such situations, speakers prefer to use more prestigious language in as many domains as possible.
- Partly because of the above-mentioned reasons and also because of other reasons, speakers may not be able to retain and recall the lexical items in the language. Generally grammatical structures are not lost because of restricted usage of language; however, retention and retrieval of lexical items tend to be affected because of extremely low use of the language in regular conversation.

Language attitude and language prestige are a result of multiple socio-political factors including the discourse related to the language in the society and the perceived 'power' of the language, which, among other factors, is dependent on the availability of resources and technologies for the language. Negative attitude or low prestige for the language could, at least, partially be alleviated by making certain kinds of resources or technologies available for the language and ensuring a digital presence of the language. One such resource is the availability of games related to language – availability of language-related games or gaming apps might actually be perceived as a matter of pride and prestige by the speakers.

In addition to this, it has been observed (Bettinson and Bird, 2017; Bird et al., 2014; Junker and Torkornoo, 2012) that the language-based games also provide an opportunity to the speakers to interact among themselves as well as with the game in their own language – it increases the overall time given to the language as well as the domain of the language usage. Moreover, it provides a context within which the recall of different kinds of linguistic constructs could be higher than that in an isolated instance. Thus it could not only give cognitive nudge to the speakers but also helps in accelerating the data collection process for the researchers working with the community.

In this paper we present the development of

Proceedings of the 4th Workshop on the Use of Computational Methods in the Study of Endangered Languages: Vol. 2 Resource Papers and Extended Abstracts, pages 10–15,
Online, March 2–3, 2021.

a mobile app / browser-based language game – mScrabble (short for multilingual scrabble). The app is an adaptation of the popular word-based language board game - Scrabble. However because of the wide differences in the scripts used for most of the languages in India in comparison to English and also the need to make it easier to play among the community members, it required changing and adapting the interface as well as rules of the game for the present purposes. These changes and adaptations are discussed in Section 3.2. In Section 3 we discuss the game interface and its different properties in the specific context of Koda and Mahali. We also briefly discuss the initial reception of the game among the community members in this section. Section 4 discusses the process of generating the game for any language using a user-friendly web-based interface developed for this purpose. Finally in Section 6 we discuss some of the new features and upgrades that we are working on to make this game better and easier for the community members.

2 The Game and its Adaptation

In general, scrabble is played with 2-4 players, wherein each player is given a rack of 7 letters and they are required to form a new word by using the letters of the alphabet available on their rack and the letters already on the board, which are placed by the other players for forming their words. The game is considered to be an excellent vocabulary building and recall mechanism for both the native as well as second / foreign language speakers and could prove to be a useful tool for recalling lexical items as well as providing greater opportunity for using the language (in case of endangered languages). Moreover, it is widely used as a novel vocabulary learning tool for children where children are exposed to new lexical items while playing the game.

For mScrabble, we have made certain modifications to the rules of the game in order to make it easier to play, especially for the community members of the endangered languages. These modifications were partly based on the feedback received from the community members. We give a list of modified interface and rules here -

- The rack size in our game is 8 as against 7 in the original game. It is further complemented by a persistent rack of matras (discussed in a little more detail in Section 3.
- There are no bonus points involved in calculating the scores (unlike Scrabble where some cells on the board give additional points to the player). This is done to simplify the scoring system and make it more accessible for the community members. [1]
- The score for each letter is not fixed or arbitrarily decided. These scores are assigned dynamically based on the normalised frequency of each letter in the given dictionary. The frequencies of each letter are normalised in between 1 - 10 and they are expected to be updated as the dictionary gets updated.
- There is an option to include a rack of persistent non-scoring characters in the game. In the present case, this is used to accommodate certain features of Indic writing systems (discussed in Section 3.2); however, it could be used in other ways as well (such as using it for providing multiple levels of difficulty in the game).
- There is no 'blank' letter (tile) in this game, which is arbitrarily assigned a letter by the players in a specific game. While technologically it is possible to do that, again, for the sake of simplicity this was done away with.
- In the original scrabble, 100 letters are part of the bag. We have increased this to 200 letters per game. Combined with an infinite number of matras, the number actually becomes even larger.
- In the original Scrabble, the rules of deciding on the correctness of a word is rather strict and there is, in fact, an official Scrabble dictionary (for English) for this purpose. However, given the sparseness of the lexicon in most of the endangered languages, this is not enforced; in fact, players are allowed to add new words to their local dictionary which might be later added to the main dictionary post-validation.

In addition to these, other minor alterations such as doing away with the formation of 'parallel'

[1] A secondary reason for doing away with the bonus point was also to get around the issues of trademark and copyright of the original makers of Scrabble - this point was raised by one of the reviewers and we have tried to address it in the overall design of the game at various places.

words, giving word suggestions etc are included in the game for the sake of simplification and also adapting for languages using non-Latin alphabets.

3 The mScrabble Game and Lexicon Collector for Koda and Mahali

3.1 The Five Ideals and mScrabble

The game was originally developed using JavaScript and HTML and could be played on one's local browser. It was then automatically converted to Android App using the Apache Cordova library [2]. Since the app was meant to be used by the community members who have recently started using technology and also the linguists who do not necessarily work in the field of technology, we tried to be mindful of what we call the five ideals of technologies for endangered languages and the communities speaking those -

- **The Relevance Ideal:** This is the first and the most fundamental ideal that any technology for endangered languages must aspire towards. The technology must be relevant for the community for which it is being developed and must be developed in association with the community members, addressing their needs and aspiration. In the process, it may strive towards supporting the language revitalisation efforts, if any, and also contribute to the overall improvement of language prestige. Achieving this ideal will also be a fundamental step towards realising a greater decolonisation of language and speech technologies (see (Bird, 2020) for a detailed discussion on this aspect). Even though one of the co-authors of this paper been working with Koda and Mahali community for last 3 - 4 years and the community members were involved in certain aspects of the development of mScrabble, we did not quite achieve this ideal mainly because the idea of the language game of this kind itself is a borrowed, colonial idea and may not be applicable at all in certain communities (such as those where the very idea of 'word' and 'word formation' are not similar to the Indo-European languages). Moreover, the extent of exact effects and relevance of mScrabble or any piece of technology for any piece of technology can only be gauged in its long-term usage and adoption and any claim regarding the achievement of this ideal can be made only over a relatively longer period of time.

- **The Transparency Ideal:** The technology must be absolutely transparent with no hidden functioning, logging and sharing of private data. We have not included any background data collection, logging or anything which the user is not aware of in mScrabble. The sharing of new words in mScrabble also involve a multi-step and explicit click on the button to be executed. If the researchers wish to include some logging details, they must all be clearly and prominently available on the interface and switched off by default such that the users will have the option of explicitly enabling and disabling those at all times. Moreover, users must have the option of looking at and sharing such logs explicitly by clicking on the button to share those. Along with this, the source code of the technology must be made available publicly for wider scrutiny as well as development in order to achieve the transparency ideal. Given the sensitive nature of data and the trust of the community members on the researchers, it is absolutely essential that this ideal be achieved and we have managed to achieve this.

- **The Simplicity Ideal:** The user interface must be friendly and intuitive enough that the users may comprehend and start using the technology without (or with minimal) instructions and training (of not more than an hour). While the interface of mScrabble was considered to be simple and intuitive, it still took some training (more so for playing the game than using the interface itself) before speakers started using it.

- **The Data Ideal:** The technology must work without an active internet connection and it should not entail any additional internet data

[2] https://cordova.apache.org/ Cordova is a wrapper library that converts HTML/JavaScript app into a native container which can access the device functions of several platforms including Android, iOS and Windows. Even though it allows to simultaneously publish the same app for multiple devices available, as of now, we have only used it to create and publish the Android apps of the game primarily because the community members we are working with have access to only Android phones. However, given that the code has been made freely available, we believe it would be rather straightforward to publish apps for other platforms as well.

cost for any user. At this point, we have managed to almost achieve this ideal and except for the download size of little over 6 MB, there is no additional internet data requirement for using the app.

- **The Infrastructure Ideal:** The technology must entail minimum requirements in terms of hardware (viz. storage, memory, screen size, etc.) and software (viz. OS version, additional dependencies, etc.) requirements. mScrabble, currently takes less than 10 MB of storage and works with relatively older versions of Android and smaller screens. However, it is still difficult to play the game on smaller screens and it might lead some difficulties among the players. One way of handling this is to reduce the size of the board but that would entail additional difficulty in playing the game and also reducing its overall functionality. In this case, it was a trade-off between the simplicity ideal and the infrastructure ideal and we have tried to achieve a balance in between the two.

3.2 The Gaming Interface

mScrabble currently works for four languages - English, Hindi (in Devanagari script), Mahali and Koda (both in the Bangla script). The user could select the language of their choice at the splash screen, followed by a selection of the number of players who will play in the game. The main gaming interface (see the screenshot of the gaming interface in Figure 2 included in Appendix A) consists of the following parts and buttons -

- **Part A:** For each player there is a 'rack' of letters available to them, which could be filled by touching on the button adjacent to the rack. There is also a 'Scorecard' for each player that shows the complete score accumulated over the course of the game by the player.

- **Part B:** A persistent rack of characters is included in the game. In comparison to English scrabble, which is meant to be played with an alphabetic script, the scripts used by language in India are syllabic such that each letter represents a syllable. In order to adapt for this in a principled way, all the vowel letters attached to the consonant letters (generally called 'matras' in the Indic languages) are always available on the rack for all the players throughout the game and they also do not carry any point.

- **Part C:** A 15x15 grid that works as the canvas for forming the words. The players could select one letter or a combination of a letter and a matra and put that on the grid to start forming the word. This process is to be repeated one-by-one for each letter / syllable until a valid word is formed.

- **Part D:** Once the word is formed then the player needs to click on the 'Check' button. At the backend, this interface makes use of a dictionary of lexical items in the concerned language for checking the validity of a word formed by the player. If the word is not in the dictionary, the player has an option to add the word to the dictionary. Thus it allows for collecting vocabulary that is not already in the dictionary.

- **Part E:** The user has the option to 'retry' mid-way (after they started forming a word but realised that it cannot be done), pass their chance to the other user and also change the letters on their rack (which entails that they also pass their chance to the next person in queue).

- **Part F:** Part F shows the points of each letter in the given language at that point in the game - this could be used by the users to cross-check the points earned by each player.

- **Part G:** Finally users have the option to revise, edit and share the new words that are added by them in Part D with the researchers and other community members using the share button. This is the only information that gets stored and that could be shared by the users.

4 The mScrabble Generator

In addition to the game itself, we are also providing an mScrabble generator interface that allows the community members and researchers to generate the game for any language by giving a list of words, list of persistent characters (if needed) and translations of the gaming interface elements. The generator is a browser-based application, which runs as a Flask App in the backend. The code

for the generator as well as the already generated games are available on the project's Github page [3], which could be downloaded and installed locally to generate new games. The generator will also be hosted on our servers for using it out-of-the-box [4].

The mScrabble generator is a simple HTML form (a partial screenshot is given in Figure 3 included in Appendix A) that asks for the relevant information and files from the user and based on the input information and data files, a new game is generated. It is possible to generate the game for a single language or include multiple languages in the same interface.

5 mScrabble in the Community: Exploring the Relevance Ideal

In order to understand the acceptance and the relevance of the game among the community members, it has been distributed among the members of the two communities in India - Koda and Mahali. Both of these languages are critically endangered Austro-Asiatic languages spoken by only a few hundred speakers in the Eastern state of West Bengal (Lahiri, 2020). While the the size of community is much larger, multiple fieldwork among the community members by one of the co-authors of this paper over the last few years have revealed that most of the members have now shifted to the other major neighbouring languages such as Bangla and Hindi and only a few hundred speakers could now speak the language in some form. The rate of language attrition is further accentuated by the lack of domains of usage of language and other speakers with whom the language could be used.

The speakers of both of these languages mainly stay in rural areas and it is only within the last few years that they have started using smartphones. Since the community members have not been exposed to such apps and games earlier, it was a bit of an effort to make them understand the rules and regulations of the game. Moreover, both the languages, just like most of the other endangered languages, are mostly oral without an established writing system. But slowly over a period of time a large number of community members have become 'literate' in the other official languages, especially Bangla. As a result, they are now fluent in the Bangla script and also use it for writing their own language, wherever needed. This was the primary reason why we used the Bangla script for the game as well.

As expected, the game has been taken with great enthusiasm by the community. A preliminary set of short unstructured interviews and feedback revealed that the game has piqued the interest of the community members but they find it quite difficult and tricky to play initially. Based on their feedback, we included the buttons to pass their chance and reshuffle the rack (Part E of the game discussed in Section 3.2) in the interface. We are also now working on including a word suggestion module to increase the ease of playing. We are currently observing the usage pattern and the long-term engagement of the community members with the app and we hope to get more concrete results on this over the next few months. We also plan to conduct more detailed ethnographic interviews and getting the feedback from different sections of the community over a relatively longer period of time. We hope to get better understanding of the relevance, impact and reception of the gaming apps like these.

6 The Way Ahead

Currently we are working on generating apps for and piloting with some more languages in India. We are also currently working on a word recommendation system (based on the community feedback) that will be integrated with the game. This recommendation system will suggest words (including their automatically-generated inflected forms) to the players in case they are stuck and not able to form a new word. While this will add an additional level of ease for game players, more importantly, it will enable the app to be used as a vocabulary learning tool for both children and adult speakers of endangered languages. Since, unlike in case of thriving languages, the adult speakers of endangered languages may also not have a large lexical stock (because of several reasons including the inability to use language regularly and in multiple domains), this recommendation system is expected to perform the function that an adult does in the classic scrabble - expose the players to new lexical items which they may not be aware of or may have forgotten.

[3] https://github.com/kmi-linguistics/mScrabble

[4] This is currently not available but will be soon made available on the project website - http://mscrabble.ctrans.in

Acknowledgments

We would like to express our sincerest gratitude to Mahali and Koda speakers who have worked with us towards the development of this app.

References

Mat Bettinson and Steven Bird. 2017. Developing a suite of mobile applications for collaborative language documentation. In *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 156–164, Honolulu. Association for Computational Linguistics.

Steven Bird. 2020. Decolonising speech and language technology. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3504–3519, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Steven Bird, Florian R. Hanke, Oliver Adams, and Haejoong Lee. 2014. Aikuma: A mobile app for collaborative language documentation. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 1–5, Baltimore, Maryland, USA. Association for Computational Linguistics.

M.O. Junker and D. Torkornoo. 2012. Online language games for endangered languages (jeux.tshakapesh.ca www.eastcree.org/lessons). In *EDULEARN12 Proceedings*, 4th International Conference on Education and New Learning Technologies, pages 6662–6673. IATED.

Bornini Lahiri. 2020. Effect of Bangla on Koda verbs. In Tariq Khan, editor, *Queries in Structure of Language*, pages 131 – 37. Central Institute for Indian Languages.

A The mScrabble Screenshots

Figure 1: Screenshot of the splash screen of mScrabble

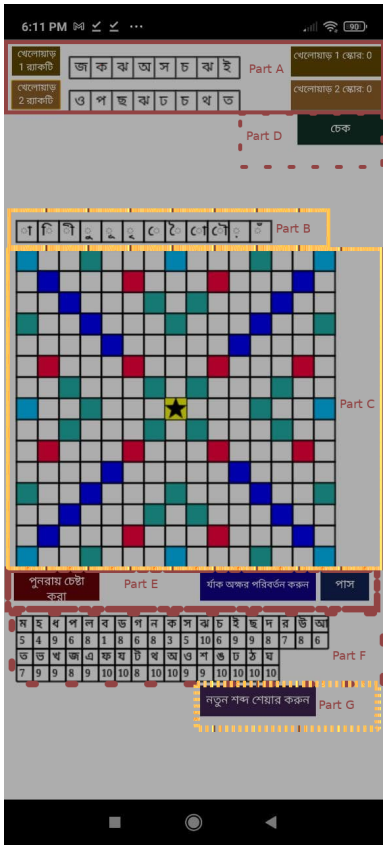

Figure 2: Screenshot of the main Gaming Interface of mScrabble

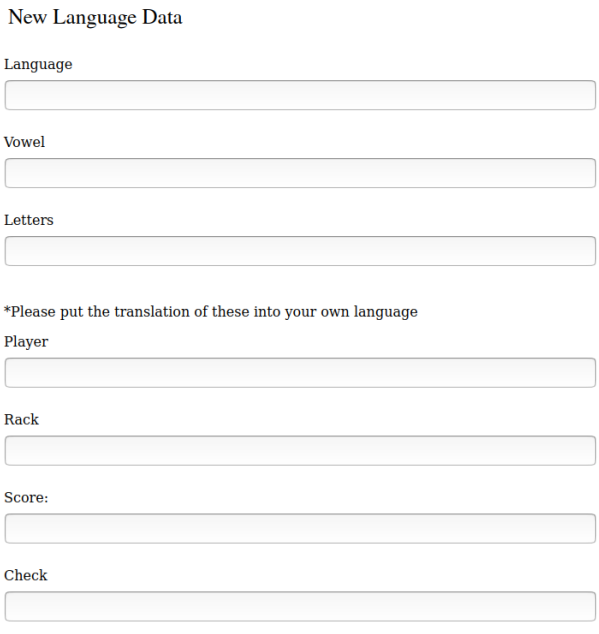

Figure 3: Screenshot of the mScrabble Generator