A Game-based Data Collecting Framework for the Recommendation of Kids' Second Language Learning

Weizhi Ma, Min Zhang*, Chenyu Zhang, Yixin Chen,
Qian Xie, Weiyue Sun, Yiqun Liu, Shaoping Ma
State Key Laboratory on Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science & Technology, Tsinghua University, Beijing, China
z-m@tsinghua.edu.cn

ABSTRACT

Under the background of globalization, more and more people are learning a second language. Word learning is essential in language learning. It is valuable and helpful if a recommendation system for the language word learning is proposed for learners. In previous work, the research of language cognition is often based on the dataset of kids' first language development, such as: WordBank, to summarize and study the word learning process. However, there lacks proper large-scale second language vocabulary development dataset, and it is very difficult to collect a big dataset with traditional collecting methods. This limits the study of second language learning and the construction of words recommendation systems. In this paper, we designed a data collecting framework for kids based on the idea of games with a purpose, to collect kids' vocabulary development status and his/her attributes. We have implemented the second language vocabulary development collecting system for English. Moreover, we proposed an idea of word learning recommendation system for kids.

KEYWORDS

Word Recommendation; Kids; Second Language Learning; Game with a Purpose;

1 INTRODUCTION

Language learning is a fundamental part of human life (especially in childhood), during which kids gradually learn how to express their minds, abstract or concrete concepts. Obviously, language learning has a strong relationship with the development of kids' cognitive ability. And there are studies about recommending the proper words of native language (L1) to help the learners [2].

While the second language (L2) studying can be more difficult compared to the first, for lacking idealistic learning environment (E.g.: learner's family and friends may not use the language to communicate with him/her), especially when L2 is not the native language of the country. Thus, we want to provide a word recommendation algorithm for L2 learning based on related psychological and neural cognitive theories. Our studies are concentrated on L2 which is not a native language, because it is easier to learning a language in the native environment (E.g.: learning English in America as second language).

In previous work, the order of word learning for L1 is able to predict the trajectory of one's language learning, which is useful in recommending the most proper words for learners to learn [7]. The system is driven by large-scale data and a modeling method. For example, Wordbank[1] is an open database of kids' vocabulary established by researchers, featuring data from contributors around the world. It archives data from the MacArthur-Bates Communicative Development Inventory (CDI), a family of parent-report questionnaires. In building recommendation system for word learning, we can use the collected dataset to decide whether a word is proper for a 15-month child. Apparently, the dataset is the base of prediction.

However, to the best of our knowledge, there lacks proper L2 language learning open dataset. The dataset should be collected in countries that do not take L2 as native language, which is essential for this study. In fact, collecting a big dataset is difficult and costly. To solve this problem, we take several measures to provide data support for the study of second language learning recommendation system. Our main contributions are listed as follows:

- As far as we know, this is the first work aiming to specialize researches about L2 learning process of preschool-children in countries that do not take L2 as native language. We proposed an online data collecting framework, which aims to collect big data for L2 word learning recommendation.
- The collecting framework is game-based. Based on psychological and linguistic theories, bonus features (E.g.: cognitive abilities and learning methods) of users are collected.
- The whole framework could be applied to other data collecting problems after a few adaptations.

2 RELATED WORK

2.1 Word Recommendation for Native Language Learning

CDI is a family of parent-report questionnaires which can measure kids' native language (English or Spanish) development level [6]. Scoring charts are based on the statistical data during the last several

*Corresponding author.
*This work is supported by Natural Science Foundation of China (Grant No. 61532011, 61672311) and National Key Basic Research Program (2015CB358700).

DOI: 10.475/123_4

[1] http://wordbank.stanford.edu

years. It depicts an "average" child in different cognitive levels and is efficient. Many L1 studies are conducted based it, such as: [3], [9], and [13].

Several scientists propose hypotheses about the semantic and phonology network structure in the human brain and how human memorize and learn a new word. Hills finds the law of preferential acquisition of semantic network using CDI as tools [7]. Large-scale dataset of CDI is applied to support the proposed ideas. Other researchers also proposed different network and word learning assumptions [2]. We prefer that the proposed preferential acquisition law could be helpful in recommending suitable words to learners.

2.2 Previous Studies on Second Language Learning

Many datasets, such as Wordbank, are collected for L1 learning process of preschool children. Many studies on language cognition are based on these statistics. However, less large-scale datasets are available for L2 study by far. Existing papers often propose some hypothesis about details of L2 studying while lacking enough convincing statistical results to support them [10].

There are already some L2 studies about English as a Second Language [1] and [4]. While many of them are based on datasets collected in countries that take English as native language (E.g.: learning English in America). We have more interests in helping speakers from other countries to learn L2. Thus, the ideal dataset should be collected in countries which do not take the English as a native language.

L1 studies often take preschool children as targets for survey, because it is easier to control those variables (which may affect learning results) than adults. Thus, we take kids' L2 word learning status as data collecting target as well. We need to design an effective and efficient data collection method first.

2.3 Game with a Purpose for Data Collecting

To collect a large-scale dataset, it is expensive and time consuming to hire people completing special tasks or questionnaires. For reasons above, some researchers using web online game-based system to collect high-quality data provided by users in the Internet.

These systems perform well on attracting online clients into playing games. Some questions are hidden in the gaming procedure, answered by users when they are unconscious, which is a win-win work. Users will have fun and researchers get the desired data. A typical instance is reCAPTCHA system designed by Luis, requiring users to input verification code from pictures. Billions of matching tags for pictures are available via this system every year, and achieving a high accuracy as 99.1% [16]. ESP Game system[14] and Phetch system[15] designed by him and his colleague are other great applications. ESP Game system lets two online users find as many objects as they can in the same image, which is aiming to collect picture object labeling results. Users are inspired to find more things and collaborate with others while having great fun.

We follow the idea of collecting data in the game, and use it to design the L2 word learning data collecting framework.

3 PROBLEM ANALYSIS

Our current research is aimed at developing a game-based data collecting system which collects second language (English here) vocabulary learning data from kids in early childhood (aged 3 - 6), which is a fundamental work of words recommendation system for second language learning.

3.1 Ethical Considerations

We adopted four ethical principles in our study:

- **Safety Principle**: Any potential harm to kids' physical and mental health is not allowed, including (1) negative feedback on kidsfi performance; (2) long-time engagement.
- **Validity Principle**: The collected data should be valid and able to serve for research purpose. Particularly, parents's engagement is helpful to ensure that kids understand and actually follow the instructions.
- **Comprehension Principle**: Comprehensive consideration of relevant variables is required given the complex nature of language learning.
- **Privacy Principle**: kids' privacy should be protected.

3.2 How to Design the Game-based Framework

Basically, there are two widespread methods in collecting language learning data as we known: **A.** Record the language learning performance of early childhood kids by audios or videos. Those records are then analyzed manually. **B.** Require kids or their parents to complete some tasks in a face-to-face setting with the assistance of specially-trained professionals.

However, the methods have the following limits:

- Time and labor consuming, it is hard to obtaining a large scale dataset.
- Audio or video records of kids' language learning performance may infringe their privacy.

To overcome the disadvantages, we propose a game-based data collecting framework here. The system provides a English vocabulary test which require the engagement of both parents and kids. The feedback about kids' second language vocabulary ability will be provided as an incentive for participants and the whole test can be completed easily with a mobile phone connected to Internet. In such way, the system satisfies parents and kids' need for entertainment and language ability assessment, and is able to collect large-scale data at a low cost. Notice that to keep the results trustable, the operations on the system are finished by kids' parents.

3.3 Strategy of Word Selecting

To measure kids' English vocabulary validly and collect sufficient information, it is advisable to test kids with a large number of words. However, to given the short concentration span of kids participants, it is desirable to develop a compact test. Thus, how to design the words in questionnaire is very important. We have to design a word selecting strategy.

To solve such conflict, we could use the words which are adapted from the Wordbank database (including 558 words from the CDI vocabulary). Moreover, we make the words in one's vocabulary test adaptive. Each word is classified into a recognition Level, and participants will face different words decided by their performance. In each level, several words are common that every participant has to answer, and others are randomly chosen. In this way, we are able to measure participants' English vocabulary across a wide range while control the overall length of the test. Also, by designing common and random words, we could obtain data which cover a wide range of words while still relatively concentrate.

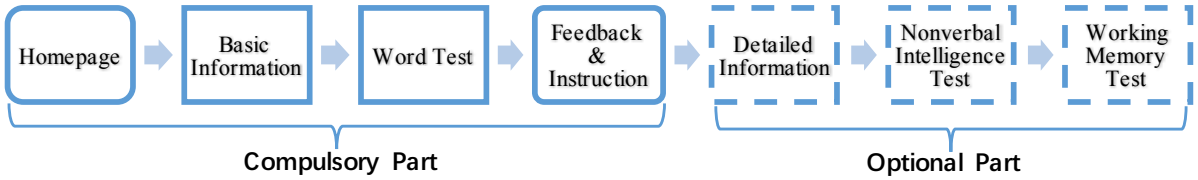

Figure 1: The Flow chart of Game-based Data Collecting System

Noticeably, the level of a word is determined by the age when it is mastered by native speakers currently, not a standard customized to second English learners. Additionally, we particularly cover both concrete and abstract words in the test for diversity purpose.

3.4 Other Relevant Variables

Many factors have effects on a second language learning, and we want to know more about the user. Thus, some valuable supplementary information is listed in below:

- Demographic information: child's gender, age, and family background, ways to learn L2...etc.
- Information about L2 learning: exposure to L2 (E.g.: L2 class in the kindergarten, extracurricular L2 courses and activities, media, etc.), starting age, average spent time and intensity of each kind of English exposure.
- General cognitive abilities: working memory and nonverbal intelligence [12].

We can profile a user better with these information, which could be helpful in achieve fine-grained word learning recommendation. Moreover, except demographic information, which is presented at the beginning, other information mentioned above could serve as an optional part presented after the vocabulary test.

4 FRAMEWORK OF GAME-BASED DATA COLLECTING SYSTEM

4.1 Framework

Figure 1 is the flow chart of our game-based data collecting system. Considering a disadvantage of online tests is that users may quit at any time, we avoid collecting too much information at the beginning. For this purpose, we provide both the compulsory and the optional part. Thus, users may be willing to finish the test.

The compulsory part is the key of our system. At first, we will collect some basic information of kids: age, sex and ways to learn second language. Then begins the word test, those words in the test are carefully chosen and graded. Moreover, we have devised adaptive word difficulty, in order to collect more valuable data. After answering a certain number (20 in our design) of words recognition rounds, the system will give feedback to users based on the accuracy. An instruction will encourage users to participate in the subsequent optional part.

The optional part consists of a questionnaire and two tests (the dashed squares in Figure 1). They are corresponding to Section 3.4. The questionnaire is the supplement of the previous questionnaire on basic information, we will further ask parents about the time their kids spent on each way to learn L2. The two tests are Raven's Progressive Matrices Test measuring nonverbal intelligence and working memory test measuring language learning ability.

4.2 Word Selecting Algorithm

As Section 3.2 mentioned, word grading is significant to produce a list of English words suitable for kids. We choose 558 CDI English words from Wordbank, arrange them into five recognition levels according to the statistics of more than 10,000 kids aged from 16 to 30 months. For each word w, denote the proportion of kids aged at x knowing w as $f_w(x)$. Find x^* such that

$$x^* = argmin_x(f_w(x) >= 0.5) \tag{1}$$

and the level of w satisfies that

$$w \in \begin{cases} L_1 & 16 < x^* <= 18 \\ L_2 & 18 < x^* <= 21 \\ L_3 & 21 < x^* <= 24 \\ L_4 & 24 < x^* <= 27 \\ L_5 & 27 < x^* <= 30 \end{cases}$$

The grading rule is based on the CDI test results from WordBank. We set the interface as 0.5, because we think half of kids know a word means that word is proper for kids in that age. The number of words in L_1 to L_5 are 41, 51, 194, 164, and 108 respectively.

In each level L_i, we select 2 or 4 words as common word. The selection strategy considers not only abstract and concrete words but also nouns, verbs, adjectives, and prepositions. Take level 4 as an example, we select ride, broken, hamburger and cloud as common words. Besides, in order to dynamically adjust the difficulty of the test, we designed Algorithm 1 to choose a word to produce the test.

Algorithm 1 Word Selection Algorithm

if $C_x! = \emptyset$ **then**
  $w = \text{RANDOM}(C_x), C_x = C_x \backslash w$
  $result = \text{QUESTION}(w)$
  $Gain \mathrel{+}= (result == TRUE)\ ?\ A_w : -A_w$
  if $C_x == \emptyset$ **then**
    $x \mathrel{+}= (Gain > \alpha)\ ?\ 1 : 0$
    $x \mathrel{-}= (Gain < \alpha)\ ?\ 1 : 0$
    $Gain = 0, K = 0$
  end if
else
  $w = \text{RANDOM}(R_x), R_x = R_x \backslash w$
  $result = \text{QUESTION}(w), K \mathrel{+}= 1$
  $Gain \mathrel{+}= (result == TRUE)?A_w : -A_w$
  if $K >= 2$ **then**
    $x \mathrel{+}= (Gain > \beta)\ ?\ 1 : 0$
    $x \mathrel{-}= (Gain < -\beta)\ ?\ 1 : 0$
    $K = (Gain > \beta\ or\ Gain < -\beta)\ ?\ 0 : K$
    $Gain = (result == TRUE)\ ?\ A_w : -A_w$
    $Gain = (Gain > \beta\ or\ Gain < -\beta)\ ?\ 0 : Gain$
  end if
end if

In the algorithm, $QUESTION(w)$ is a function representing a query about word w to the user. All users start at level x = 1, A_w is the score of word w. In each level, the set of common words C_i has priority over other words. After common words in level i are used, we randomly choose words from R_i. In our actual system, the initial value is set to 0 for both Gain and K, the score of each word is set to 1 and the two parameters α and β are set to 2 and 1.

Each question needs parents to click on the icon to play the audio of English pronunciation and ask their kids to click on the picture best describes the word they hear. The three wrong choices are randomly chosen from our self-made picture library. To make sure that each question is unambiguous, we fix choices for each question and examine them by hand for at least three times.

4.3 Optional Tests

We use Raven's Progressive Matrices Test [12] and working memory test [11] as a method to test kids' nonverbal intelligence and short-term memory to help us know more about the kid. Raven's Progressive Matrices was formulated by British Psychologist Raven in 1938 and is still popular around the world today. It is usually a test used in measuring abstract reasoning and regarded as a nonverbal estimate offluid intelligence. We choose 8 representative items for kids in our system [8].

Working memory is a cognitive system with a limited capacity that is responsible for temporarily holding information available for processing. Existing research results show that strong working memory contributes to language learning [5]. A common approach to measure working memory is to find the digit span (the maximum length of numbers that a person can repeat correctly immediately). The average digit span for normal adults without error is five to nine. Our working memory test is based on the digit span test of Wechsler Adult Intelligence Scale, and we make a little adjustment to shorten the test time. The primary length of digit is 3. If a kid successfully answers one, then length += 1. If the kid wrongly answers 2, then the current length is his/her score.

4.4 Design of System Feedback

Considering our game-based goal and the safety principle, we focus not only on the design of variables in the questionnaire and algorithms in the tests, but also on the design of the feedbacks. We should make sure the feedbacks will not let the kids down, so that they will not be hurt and are more willing to finish the tests.

We will not tell each kid whether he is right or wrong immediately after a question is answered. This is because kids are easily frustrated by bad results. The following are feedbacks of three tests in our system respectively. E.g.: **Word test**: According to the corresponding month x of the hardest word that the user answers correctly and informs him/her that: Your honey has known English words that native English speakers aged at month x known. If all wrong, then the feedback is: It seems that your honey knows little about English, why not start from simple conversation words?

To summarize, we have developed the data collecting system and APIs adapted to mobile phones based on the design in Section 4.1 and 4.2, the whole system is under operation for data collecting now, you can visit the website (http://kidswords.thunlp.org:15010). We will use the collected data to support the related research of English language learning and cognition. We assume the users are Chinese kids here, so the interface is in Chinese.

5 DISCUSSIONS ABOUT RECOMMENDATION

The designed data collecting system is for L2 learning recommendation, the performance of kids' word learning and some personal information are collected. There are several potential ways for recommendation, such as: A. Calculate the word learning status in different ages with different learning methods separately. When a user first use this system, record the words his/her already known, and recommend words according to his/her ages and learning methods. B. General a word trajectory queue, recommend users to learn words according to the queue. We will try different methods after data collecting and hope more researchers could construct helpful recommendation systems for L2 learning with us.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we present our work on designing a L2 word learning system based on game playing for kids, which is a fundamental work for language learning recommendation system construction. Moreover, we implement the framework for kids in China who are learning English.

The collected dataset will be published later. We hope that will be helpful in kids language learning and related studies and we will further construct L2 learning recommendation systems.

REFERENCES

[1] Selenay Aytac and Selenay Aytac. 2016. Use of action research to improve information literacy acquisition of international ESL students. *New Library World* 117, 7/8 (2016), 464–474.
[2] Nicole M Beckage and Eliana Colunga. 2015. Language networks as models of cognition: understanding cognition through language. *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks* (2015), 3–28.
[3] Mika Braginsky, Daniel Yurovsky, Virginia A Marchman, and Michael C Frank. 2015. Developmental Changes in the Relationship Between Grammar and the Lexicon.. In *CogSci*.
[4] Andrzej Cirocki, Sujeewa Tennekoon, Alicia Pena Calvo, and others. 2014. Research and reflective practice in the ESL classroom: Voices from Sri Lanka. *Australian Journal of Teacher Education (Online)* 39, 4 (2014), 24.
[5] Nick C Ellis. 1996. Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology: Section A* 49, 1 (1996), 234–250.
[6] Larry Fenson, Elizabeth Bates, Philip S Dale, Jeffrey P Hartung, Stephen J Pethick, Judy Snitzer Reilly, J Steven Reznick, and Donna J Thal. 1996. *The MacArthur communicative development inventories.* Singular publishing Gruop.
[7] Thomas T Hills, Mounir Maouene, Josita Maouene, Adam Sheya, and Linda Smith. 2009. Longitudinal analysis of early semantic networks: preferential attachment or preferential acquisition? *Psychological science* 20, 6 (2009), 729–739.
[8] RM Kaplan and DP Saccuzzo. 2009. Standardized tests in education, civil service, and the military. *Psychological testing: Principles, applications* 7 (2009), 325–327.
[9] Magdalena Luniewska, Ewa Haman, Sharon Armon-Lotem, Bartffomiej Etenkowski, Frenette Southwood, Darinka An‼elkovifj, Elma Blom, Tessel Boerma, Shula Chiat, and Pascale Engel De Abreu. 2016. Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words? *Behavior Research Methods* 48, 3 (2016), 1–24.
[10] Brian MacWhinney. 1997. Second language acquisition and the competition model. *Tutorials in bilingualism: Psycholinguistic perspectives* (1997), 113–142.
[11] Akira Miyake and Priti Shah. 1999. *Models of working memory: Mechanisms of active maintenance and executive control.* Cambridge University Press.
[12] John C Raven. 1936. Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive. *Unpublished masterfis thesis, University of London* (1936).
[13] Rose Schneider, Daniel Yurovsky, and Mike Frank. 2015. Large-scale investigations of variability in children's first words.. In *CogSci*.
[14] Luis Von Ahn and Laura Dabbish. 2008. Designing games with a purpose. *Communications of the Acm* 51, 8 (2008), 58–67.
[15] Luis Von Ahn, Shiry Ginosar, Mihir Kedia, and Manuel Blum. 2007. Improving image search with phetch. In *Acoustics, speech and signal processing, 2007. icassp 2007. ieee international conference on*, Vol. 4. IEEE, IV–1209.
[16] Luis Von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. 2008. recaptcha: Human-based character recognition via web security measures. *Science* 321, 5895 (2008), 1465–1468.