Mohd Ayub A. F. et al. (Eds.) (2011). Workshop Proceedings of the 19th International Conference on Computers in Education. ChiangMai, Thailand: Asia-Pacific Society for Computers in Education.

Stepwise Selection of English Multiple-choice Cloze Questions Based on Difficulty-based Features for Keeping Motivation

Tomoko KOJIRI[a*], Yuki WATANABE[b] & Toyohide WATANABE[b]
[a]*Faculty of Engineering Science, Kansai University*
[b]*Graduate School of Information Science, Nagoya University*
* kojiri@kansai-u.ac.jp

Abstract: We had constructed the automatic generation system of English multiple-choice cloze questions. By using the system, plenty number of questions can be generated automatically, but learners become difficult to find appropriate questions. Therefore, objective of this research is to develop the method for providing questions that fit for learners. In the self-learning, questions that increase motivation for learners are effective. This research determines features that affect to difficulties of questions (difficulty-based features) and proposes the method for selecting questions according to the difficulty-based features for the stepwise learning. In order to manage the relations among questions, a question network is introduced in which questions are structured based on differences of each difficulty-based feature. Questions are selected by following appropriate links according the learners' answers..

Introduction

Multiple-choice cloze questions are often used in English learning. Such type of question is effective for checking the knowledge of grammar and lexicon. In addition, by tackling these questions repeatedly, the knowledge of English grammar and lexicon is able to be acquired. Only limited number of knowledge is included in one question, so many questions are needed to be solved for the purpose of acquiring whole grammar and lexicon knowledge. We have constructed the automatic generation system of English multiple-choice cloze questions; MAGIC [1]. By using the system, multiple questions can be generated automatically. However, to fit questions to learners' understanding situation is not focused. If difficult questions are posed to learners repeatedly, they do not feel like studying with the system for a long time.

Since questions contain plenty knowledge of grammar and lexicon, it is sometimes difficult to determine acquired/ in-acquired knowledge. In addition, learners' motivation is affected by their feelings whether they think "difficult" or "easy" for the questions. Such feelings may arise from the superficial features of questions. If the features of questions indicate that the question is too difficult, learners do not feel like tackling the questions. If the question is too easy, learners think questions are meaningless for them.

Traditional Intelligent Tutoring System or computer-adaptive testing tends to provide learning con-tents/test items that are appropriate for learners' understanding knowledge [2-4]. These systems analyze learners' acquired/in-acquired knowledge from their learning activities, such as answers of exercises. However, questions that are selected based on such knowledge-based approach do not always keep learners' motivation. Therefore, the objective of this research is to develop the method that provides questions based on the features that affect to learners' motivation (difficulty-based features). Difficulty-based features consist of more than one feature, and learners' feelings for these features may be different for each learner. So, the basis for selecting questions should be dynamically changed according to the learners.

Currently, target questions are questions that are generated automatically by MAGIC. So, the difficulty-based features need to be acquired systematically from the generated questions. Target learners are non-native speakers who do not understand basic grammatical knowledge.

1. Difficulty-based Features of English Multiple-choice Cloze Question

Figure 1 is an example of English multiple-choice cloze questions. A question consists of *sentence*, *blank part*, and *choices*. Choices include one correct choice and three distracters. Learners select one from choices for filling in the blank part.

There are various definitions or findings about difficulty features of English questions. Kunichika et al. defined difficulty features of English passage reading questions for non-native speakers as difficulties of *understanding of original texts*, *understanding of question sentences*, and *understanding of answer sentences* [5]. In English multiple-choice cloze questions, both original text and questions sentence correspond to question sentence, and answer sentences correspond to distracters. Therefore, following difficulty-based features are defined.

1) Difficulty of sentence --- Readability is one of the features that prevent learners of understanding the meaning easily. Researches about readability of English sentences insisted that lengths of sentences or difficulties of words affect to the readability [6]. Based on this result, *lengths of sentence* and *difficulties of words* are defined as one of the difficulty-based features of a sentence.

2) Difficulty of distracters --- There are various relations between distracters and a correct choice. In some questions, all distracter types are the same. The number of the distracter types affects to the difficulty of questions. If all distracter types are the same, it is easier to find the correct choice. On the other hand, questions become more difficult if similar types of distracters exist in it. Therefore, *the number of distracter types* in choices is defined as a difficulty-based feature. As the distracter types, 12 types defined in MAGIC are applied.

2. Question Selection Method Based on Question Network

Learners are motivated to learn repeatedly if difficulties of questions become gradually increasing. If questions seem easy, learners feel that they cannot acquire new knowledge from it. Appropriate questions for learners should contain a little difficult difficulty-based feature than those in solved questions.

In order to represent the stepwise relations among questions, a question network is introduced that organizes all questions based on difficulty levels for each difficulty-based feature. In the question network, questions in the same levels for all difficulty-based features form one node, and nodes whose levels are next to each other are connected by links. By following this question network, learners are able to tackle questions from easier one to more difficult one according to their understanding levels. Figure 2 illustrates the conceptual framework of the question network. Nodes without incoming links correspond to the easiest questions. Nodes without outgoing links have the most difficult questions.

The levels of each difficulty-based feature are defined as follows.

- Length of sentence---The number of words is regarded as one of the viewpoints of defining the length of sentence. Based on the analysis of 1500 questions in the database of our laboratory, it is revealed that sentences consist of 4 to 32 words. Thus, we

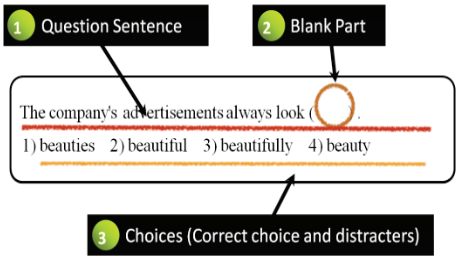

Figure 1: Example of English multiple-choice cloze question

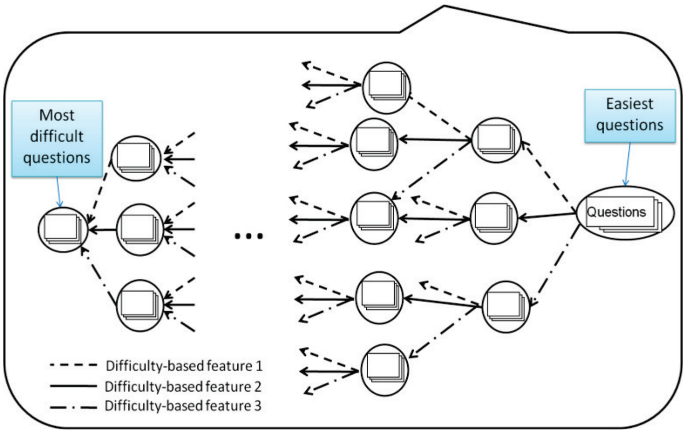

Figure 2: Conceptual framework of question network

categorize the length of sentence into four levels according the number of words. Table 1 shows the levels of the length of sentence.

- Difficulty of words---In this research, the difficulty of words followed SVL12000[7], which is the list of word difficulties defined by ALC. In SVL12000, 12000 words that are useful for Japanese are classified into five levels of difficulty. The level of a question is defined as the highest level in all words including the sentence and choices.
- The number of distracter type---Distracter types correspond to generation rules to generate distracters in MAGIC. Since choices of the same distracter types may be more difficult than that of the different one, the difficulty based on the number of distracter type is set as Table 2

Table 1: Levels of length of sentence

Level	1	2	3	4
# of words	less than 11	12 to 18	19 to 25	more than 26

Table 2: Levels based on the number of distracter type

Level	1	2	3
# of distracter types	3	2	1

Using the question network, learners' next questions are selected based on the answers of former questions. Figure 3 shows the process of selecting questions from question network. Currently, we assume that the set of questions is given at one learning.

STEP1: Based on the answers for questions in the last learning, learners' levels for each difficulty-based feature are determined. Levels for each difficulty-based feature *i* for time *t* are calculated as Equation 1. The average differences of solved questions from current level are added to the current level. In the first learning, *Level (i, t-1)* is zero.

$$Level\ (i,t) = Level\ (i,t-1) + Av.distance\ of\ solved\ questions\ from\ current\ node \qquad \dots (1)$$

STEP2: The number of solvable nodes becomes large if the learner solved questions in farther node, while it becomes small if the learner only could solve the questions in the nearer nodes. The range of solvable nodes at time *t* is calculated by Equation 2.

$$Range(t) = Range(t-1) + \left(\begin{array}{l} Ave.distance\ to\ solved\ questions - \\ Ave.distance\ to\ incorrectly\ answered\ questions \end{array}\right) \qquad \dots (2)$$

STEP3: Questions are selected from several solvable nodes. More questions should be selected from nodes that are nearer to the learner's current node. The probabilities for selecting questions for each node *i* is calculated by Equation 3. The ratio of questions from node *i* in all questions is determined by following the normal distribution based on the distance from the current node.

$$Probability(i) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{Distance\ of\ node\ i\ from\ current\ node)^2}{2}\right) \qquad \dots (3)$$

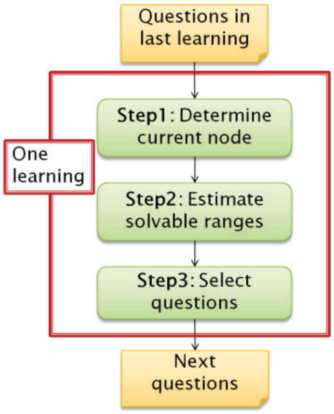

Figure 3: Process of selecting questions

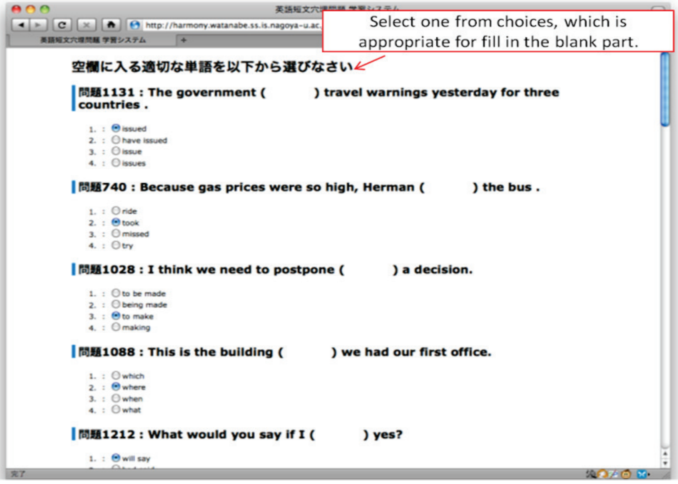

Figure 4: Interface of Prototype System

3. Evaluation

Experiments were conducted using the prototype system. The prototype system is implemented as a web-based system. When the learning starts, the selected questions are shown in the web page as shown in Figure 4. Currently, 10 questions are selected in one learning phase. Learners answer the questions by selecting the radio buttons of the correct choice. After learners select all answers and push the send button, their answers are evaluated, and the result and explanation are displayed.

12 members in our laboratory became examinees of the experiment. First, examinees were asked to solve a pretest which consists of 20 questions and examinees' initial levels were calculated based on the result of the pretest. The questions were carefully prepared by authors to include all levels of difficulty-based features as the equal ratio. In the learning phase, they were asked to answer 10 questions for 10 times. All 10 questions are different.

As the counter methods, we have prepared following two methods:

- **Random link selection method (RLSM)** which selects links randomly in selecting nodes in the question network,
- **Random question posing method (RQPM)** which selects questions randomly from the database.

In RLSM, the movement of the node occurs when the learner can solve 70 percent of the questions in the node. 4 examinees were assigned for each method. Averave understanding levels of examinees who assigned for each method were almost the same.

The correct questions in each learning were evaluated. Table 3 is the average number of correct questions and its variance for each learning. The average numbers are almost the same for all three methods. However, the variance of our method is the smallest of the three. This indicates that the number of correctly answered questions is almost the same for every learning. This result shows that our method could provide questions whose levels are similar to the learners, even if the understanding levels of learners change during the 10 learning.

Table 3: Result of learning phase

	Average # of correct questions	Variance of # of correct questions
Proposed method	5.725	1.585
RLSM	5.850	2.057
RQPM	5.825	2.665

The questionnaire result for acquiring the consciousness of examinees for the proposed questions is shown in Table 4. In each questionnaire item, 5 is the best and 1 is the worst.

Items 1 and 2 got high values. Based on the result of item 1, examinees felt questions become difficult as the learning proceeded. Based on the item 2, they also felt that words were getting more difficult. Table 5 shows the number of links that examinees who use the prototype system with proposed method followed during the learning. All examinees follow links of *difficulty of words* more than 2 times. The worst result of item 4 may be caused by the small number of following links based on *the number of distracter type*. Based on the result, if links are followed, learner can feel the difficulties of questions. Therefore, questions are arranged appropriately by its difficulties in the question network.

Table 4: Questionnaire result

	Contents	Average value
1	Did the questions become difficult?	4.00
2	Did the words in questions become difficult?	4.00
3	Did the question sentences become difficult?	3.50
4	Did the distracters become difficult?	2.75

Table 5: # of links that examinees followed

	Difficulty of words	Length of sentence	The number of distracter type
Examinee 1	3	1	1
Examinee 2	2	2	0
Examinee 3	2	3	1
Examinee 4	3	1	0

4. Conclusion

In this paper, the method for posing questions based on the subjective difficulty-based features was proposed. Based on the experimental result, defined features are intuitive and match to learners' consciousness. In addition, using the question network which arranges questions according to the levels of difficulty-based features, questions that fit for learners' levels were able to be selected in spite of change of learner's situation during the learning. In our future, we need farther experiments with students of various understanding levels for confirming the effect of our method.

Currently, three difficulty-based features have been prepared. However, there are still several other features in questions, such as grammatical structure. Thus, for our future work, to investigate other features of questions is necessary if they become difficulty-based features or not.

References

[1] Goto, T., Kojiri, T., Watanabe, T., Iwata, T., and Yamada, T.: "Automatic Generation System of Multiple-choice Cloze Questions and its Evaluation", An International Journal of Knowledge Management and E-Learning, Vol.2, No.3, pp.210-224 (2010).

[2] Suganuma, A., Mine, T., and Shoudai, T.: "Automatic Generating Appropriate Exercises Based on Dynamic Evaluating both Students' and Questions' Levels", Proc. of ED-MEDIA, CD-ROM, (2002).

[3] Chen, C.M., Lee, H.M. and Chen, Y.H.: "Personalized E-learning System Using Item Response Theory", Computers & Education, Vol. 44, No. 3, pp.237–255 (2005).

[4] Songmuang, P., Ando, M., Nagamori, M., Ueno, M., Okamoto, T.: "E-testing Construction Support System with some Prediction Tools", Proc. of ICALT, pp.437-438 (2007).

[5] Kunichika, H., Urushima, M., Hirashima, T., and Takeuchi, A.: "A Computational Method of Complexity of Questions on Contents of English Sentences and its Evaluation", Proc. of ICCE 2002, pp.97-101 (2002).

[6] Kate, R. J., Luo, X., Patwardhan, S., Franz, M., Florian, R., Mooney, R. J., Roukos, S., and Welty, C.: "Learning to Predict Readability Using Diverse Linguistic Features", Proc. of ICCL 2010, pp.546-554 (2010).

[7] "Standard Vocabulary List 12000: SPACE ALC", http://www.alc.co.jp/eng/vocab/svl/index.html [in Japanese].