# Selecting English Multiple-Choice Cloze Questions Based on Difficulty-Based Features

Tomoko Kojiri, Yuki Watanabe, and Toyohide Watanabe

**Abstract.** English multiple-choice cloze questions require learners of various grammatical and lexical knowledge. Since the knowledge of learners is different, it is difficult to provide appropriate questions suitable for learners' understanding levels. This research determines features that affect to difficulties of questions and proposes the method for selecting questions according to the 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 feature. Questions are selected by following appropriate links according the learners' answers. By following this question network, learners are able to tackle questions from easier one to difficult one according to their understanding levels.

## 1 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 and thus it is used in TOEIC or TOEFL. 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, many questions need to be solved for the purpose of acquiring whole grammar and lexicon knowledge. However, it is difficult to select questions that are appropriate for individual learners' understanding levels. If all knowledge in the question has been already acquired, a learner cannot acquire new knowledge from it. If all knowledge is new to a learner, it may be difficult for him to understand plural knowledge from one question. Appropriate questions for learners should contain some acquired knowledge and a few in-acquired one.

Tomoko Kojiri
Faculty of Engineering Science, Kansai University 3-3-35 Yamate-cho,
Suita, Osaka, 564-8680, Japan
e-mail: kojiri@kansai-u.ac.jp

Yuki Watanabe · Toyohide Watanabe
Graduate School of Information Science,
Nagoya University Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan

J. Watada et al. (Eds.): Intelligent Decision Technologies, SIST 10, pp. 577–587.
springerlink.com 

Intelligent Tutoring System (ITS) which provides learning contents that fit for learners' understanding levels has been developed for different learning domain. Suganuma et al. developed the system that estimates dynamically difficulties of exercises and learners' understanding levels according to the learners' answers, and provides exercises based on the estimated understanding levels [1]. This system assigned difficulties of exercises based on the learners' answers. However, the reasons for the incorrect answer may be different among learners. The incorrect knowledge should be evaluated in detail. Since English multiple-choice cloze questions need various knowledge to solve and it is difficult to find out which knowledge is used to derive the correct answer, to specify correctly acquired/inacquired knowledge from learners' answers may be difficult. We have constructed the knowledge network that arranges multiple-choice cloze questions according to the quantity of their grammatical knowledge [2]. This research assumed that the difficulties of questions become high according to the number of their grammatical knowledge. By solving questions along the knowledge network, English grammar could be learned from a simple question to a complicate one. However, this knowledge network considers only grammatical knowledge and is not able to cope with all types of multiple-choice cloze questions. Other features of questions are needed to characterize questions.

This paper proposes features of multiple-choice cloze questions that affect to difficulties of questions (*difficulty-based feature*). Then, it proposes a method for providing questions based on the features for the stepwise learning. In order to represent the relations among questions, a question network is introduced that structures all questions based on difficulty levels for each difficulty-based feature. In the question network, questions that situate 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.

In our system, in order to estimate correctly the learners' understanding levels in the ongoing learning process, learners are required to solve plural questions at one learning process. In addition, questions are selected not from one node but from several nodes that are similar to the learners' understanding levels (*solvable nodes*). Questions in solvable nodes have larger possibilities to be solved by learners and they are determined after each learning process.

## 2 Approach

### *2.1 Difficulty-Based Features of English Multiple-Choice Cloze Question*

English multiple-choice cloze questions require learners to understand the meaning of English sentences and grammatical structure, and to select word/s for a blank part that forms the correct sentence from choices. Figure 1 is an example of English multiple-choice cloze questions. A question consists of *sentence*, *blank part*, and *choices*. Choices consist of one correct choice and three distracters. Learners select one from choices for filling in the blank part.

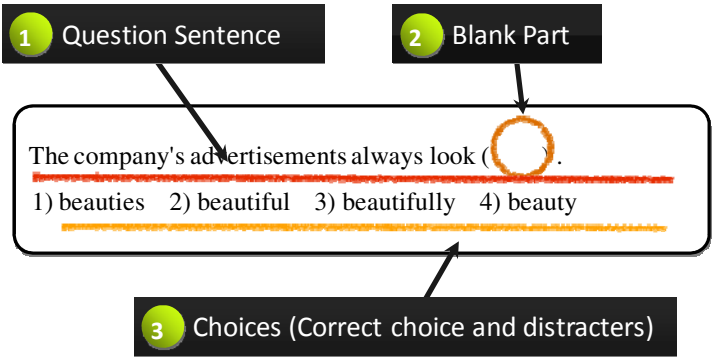


**Fig. 1** Example of multiple-choice cloze question

There are various definitions or findings about difficulty features of English questions. Kunichika et al. defined difficulty features of English reading questions as difficulties of *understanding of original texts*, *understanding of question sentences*, and *understanding of answer sentences* [3]. This paper follows this definition and defines three difficulty-based features of English multiple-choice questions illustrated. *Understanding of original texts* for English reading question corresponds to *understanding of sentence* for multiple-choice questions. *Understanding of question sentences* is regarded as understanding of intention of a question, so *understanding of blank part* corresponds to this feature. *Understanding of answer sentence* is regarded as similar to the *understanding of differences among choices*.

– 1) Difficulty of sentence

Especially for questions that ask words, it is important to grasp the meaning of a sentence correctly. 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 [4]. 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 blank part

A blank part indicates required knowledge to answer the question. In some questions, answers in the blank part can be estimated, most of which ask grammatical knowledge such as word class. Therefore, the difficulty of the blank part depends on which word class is asked.

3) Difficulty of choices

There are various relations between distracters and a correct choice. One distracter may belong to the same word class as the correct choice and another one may have the same prefix as the correct choice. The difference between choices needs to be grasped in selecting the correct answer. We adopt types of distracters defined in [5]. The types of distracters represented differences between the correct answer and the distracters. 12 types of distracters exist, which were derived by analyzing

existing questions statistically. 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.

## 2.2 *Question Network*

In learning with multiple-choice cloze questions, it is desirable that learners acquire knowledge step by step according to their understanding levels. If the difficulty of a question is determined without considering the knowledge included in the question, to support learners to acquire knowledge may become difficult. In our research, levels of difficulty-based features are assigned to each question. Such difficulty-based features are more related to the knowledge used to solve the question, but it is still acquired from the features of questions. By determining the level of learners for each difficulty-based features and selecting questions, learners are able to solve questions that are appropriate for their levels.

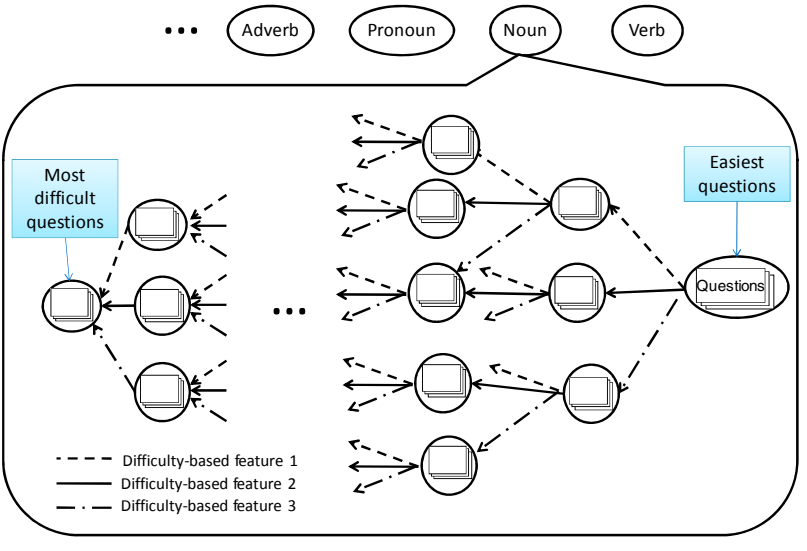


**Fig. 2** Question Network

In order to provide questions along levels of difficulty-based features, questions need to be organized according to the levels. In this paper, a question network is introduced that structures questions along the levels of each feature. Nodes in the question network contain the questions of the same levels for all difficulty-based features. Nodes whose levels are next to each other are linked by directed links. Since the difficulty order cannot be defined uniquely for the word class of the blank part, links based on word class are not attached. Instead, question networks

are constructed for all kinds of word classes. Figure 2 illustrates the question network. Nodes without incoming links correspond to the easiest questions. Nodes without outgoing links have the most difficult questions.

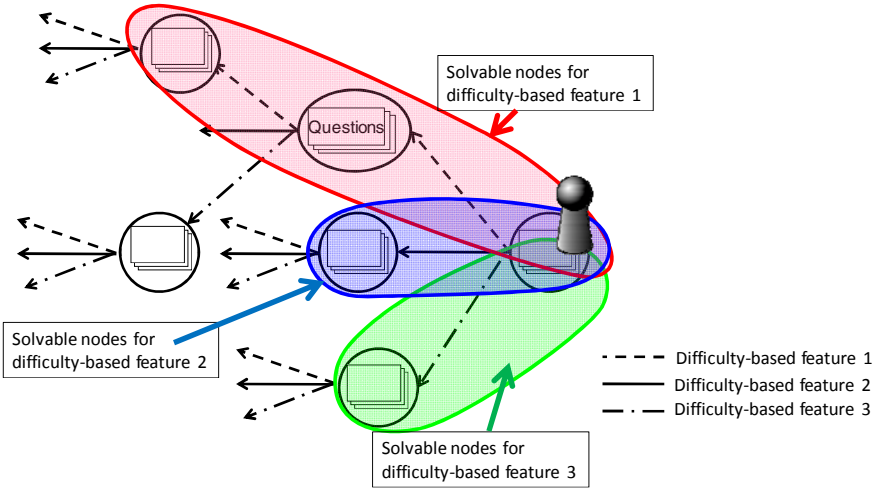


**Fig. 3** Selecting solvable nodes

Learners' current understanding levels are grasped by their answers and current their nodes are determined. If learners' levels are increased to the next level, current nodes of learners in the question network are changed by following the link of the corresponding feature.

The senses of difficulty for difficulty-based features vary for each learner. Some learners may feel a feature is critical, but others may not. If a learner does not feel the feature is difficult, learners move to the higher levels quickly, since it is a waste of time to follow the link one by one in such case. In order to determine the correct level of learners, several questions from several nodes that are estimated to be solvable are provided. The solvable nodes for each difficulty-based feature are estimated based on the learners' answers. Figure 3 is an example of selecting questions from solvable nodes. In this figure, two nodes are solvable for the feature 1, while only one node for the features 2 and 3.

## 3 Mechanism for Selecting Questions

### *3.1 Construction of Question Network*

Questions that have the same levels for all features are gathered as one node of the question network. Two nodes whose levels of one difficulty-based feature differs only one are connected by a link. Figure 4 shows a part of a question network.

This research defines length of sentence, difficulty of words, and the number of distracter types as the difficulty-based features. Followings are the method for acquiring these features from questions.

- 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 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[6], which is the list of word difficulties defined by ALC. In SVL12000, 12000 words that are useful for Japanese are selected and classified into 5 levels of difficulty. In addition to these five levels, the most difficult level 6 is attached to the words that are not in the list. 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 are attached to the questions in the database of our laboratory. People who achieved more than 700 points in TOEIC are asked to attach the distracter types based on the definition in [5]. 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

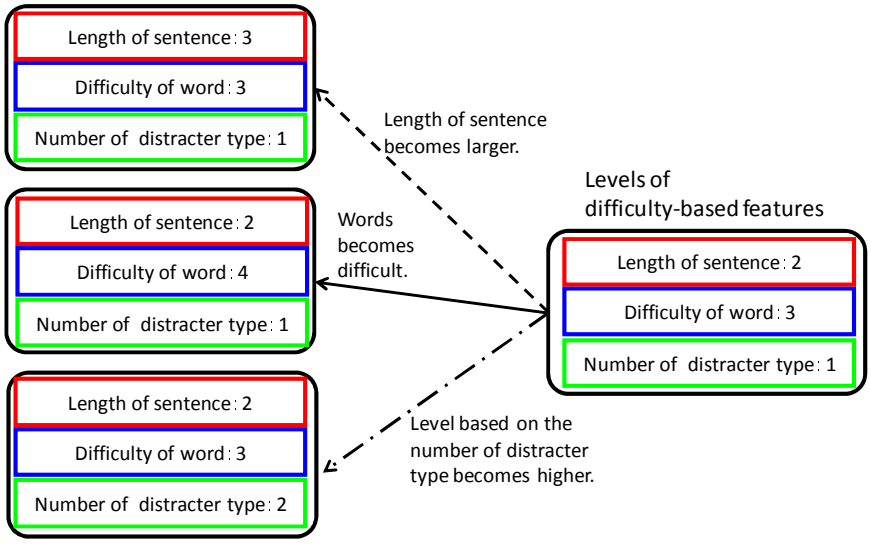


**Fig. 4** Example of attaching links

**Table 1** Levels of length of sentence

| Level | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| # of words | <11 | 12-18 | 19-25 | >26 |

**Table 2** Levels based on the number of distracter type

| Level | 1 | 2 | 3 |
|---|---|---|---|
| # of distracter types | 3 | 2 | 1 |

## *3.2 Selection of Questions*

In the begging of the learning process, the start node which fits for the learner's initial understanding level is estimated. The start node is calculated by the result of the pretest. Let $\theta_i$ be the level of difficulty-based feature *i* of a learner. $\theta_i$ is calculated by the following formula.

$$\theta_i = \frac{1}{n_i}\sum_{j=1}^{n_i} b_{i,j}P_{i,j}, \tag{1}$$

where $b_{i,j}$ represents the level *j* of difficulty-based feature *i*, $n_i$ is the number of levels in *i and* $P_{i,j}$ indicates the ratio of correctly answered questions whose levels are *j* of the feature *i*. By deriving average of levels and ratio of correctly answered questions, average level of the feature *i* can be derived.

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 the following formula.

$$S(l_i) = \beta\frac{1}{\sqrt{2\pi}}exp\left(-\frac{l_i^2}{2}\right), \tag{2}$$

$l_i$ is the number of the links from the learner's current node to the node *i*. $S(l_i)$ follows the normal distribution and $\beta$is a normalization factor. According to the probability for each node, questions are selected from the node.

In the learning phase, learners solve several questions. After the learning has been finished, learners' levels for each difficulty-based feature are re-calculated. Learners' levels for the difficulty-based feature *i* after *t*–th learning, such as$\theta_{i,t}$ is calculated by the following formula.

$$\theta_{i,t} = \theta_{i,t-1} + \frac{1}{\alpha|Q_t|}\sum_{q'\in Q'_t}\left(b_{i,q'} - \theta_{i,t-1}\right) \tag{3}$$

Average difference of solved questions' levels and the current level is added to the current level for each difficulty-based feature. Qt is the set of questions that are posed in the t–th learning and |Qt| represents the number of the questions. Q't is the set of the questions that learners answered correctly at the t–th learning. bi,q is the level of difficulty-based feature i of question q, and α is the ratio of correctly answered questions for judging the accomplishment of the node. If the value ofθi,t is bigger than that of the current node, the next node by following the link of the difficulty-based feature i until the level of the current node becomes bigger thanθi,t is selected as the current node.

The solvable nodes for the learner are also re-calculated after each learning. 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 distance of solvable nodes from the current node is calculated by the following formula.

$$r_{i,t} = r_{i,t-1} + \frac{1}{|Q_t|}\sum_{q \in Q_t} w_{i,q} l_{i,q} \quad , \tag{4}$$

This equation adds the certain number of links from the current node which is derived by subtracting the number of links to the incorrectly answered nodes from that of correctly answered ones. li,q represents the number of links from the current node to the node that question q belongs to for the difficulty-based feature i. wq corresponds to the correctness of the question q; wq is 1 if the answer to the question q is correct and -1 if the answer is incorrect. The initial value of ri,t are set as 1, which means only the next node is solvable. ri,t is set to 1 if the calculated ri,t becomes smaller than 1.

## 4 Prototype System

We have developed the web-based system based on the proposed method. When the learning starts, the selected questions are shown in the web page as shown in Figure 5. Currently, 10 questions are selected in one learning process. Learners answer the questions by selecting the radio buttons of the correct choice. If learners cannot understand the answer, they can check the checkbox which says "I could not solve the question". If this checkbox is checked, the answer of this question is not considered as learners' answers. After learners select all answers and push the send button, their answers are evaluated, and the result and explanation are displayed.

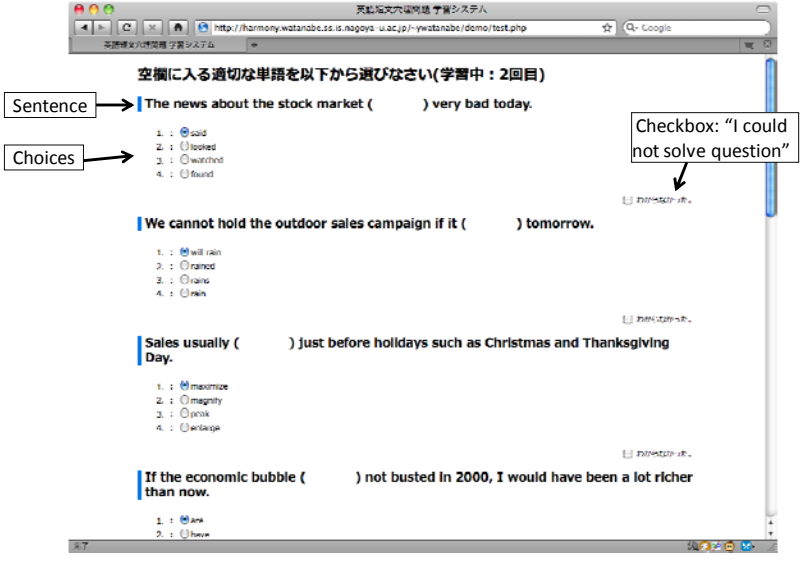


**Fig. 5** Interface of prototype system

## 5 Experimental Result

Experiments were conducted using the prototype system. This experiment focuses on the question network of verb words. 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. In the experiment, $\alpha$ in Equation 3 is set as 0.7, which means learners were regarded to accomplish the node if they could answer more than 70 percent of the posed questions.

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 solved correctly 70 percent of the questions in the node. The examinees were 12 members in our laboratory and 4 of them were assigned for each method; proposed method, RLSM, and RQPM.

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 the understanding levels of learners change during the 10th 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. For the item 3, examinee who answered 4 followed the link of *the length of sentence* more than 2 times, and examinees who answered 3

followed the link only once. 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 |

## 6 Conclusion

In this paper, the method for posing questions step by step according to the 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.

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. Moreover, our system only provides explanation for learner's answers to the question and does not support learners of acquiring the knowledge. If questions run out before learners can make correct answers for the certain amount of questions, they cannot proceed the learning process. Therefore, we have to provide the support tool that teaches necessary knowledge to learner.

## References

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

2. Goto, T., Kojiri, T., Watanabe, T., Yamada, T., Iwata, T.: English grammar learning system based on knowledge network of fill-in-the-blank exercises. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part III. LNCS (LNAI), vol. 5179, pp. 588–595. Springer, Heidelberg (2008)
3. Kunichika, H., Urushima, M., Hirashima, T., Takeuchi, A.: A Computational Method of Complexity of Questions on Contents of English Sentences and its Evaluation. In: Proc. of ICCE 2002, pp. 97–101 (2002)
4. Kate, R.J., Luo, X., Patwardhan, S., Franz, M., Florian, R., Mooney, R.J., Roukos, S., Welty, C.: Learning to Predict Readability Using Diverse Linguistic Features. In: Proc. of ICCL 2010, pp. 546–554 (2010)
5. Goto, T., Kojiri, T., Watanabe, T., Iwata, T., Yamada, T.: Automatic Generation System of Multiple-choice Cloze Questions and its Evaluation. An International Journal of Knowledge Management and E-Learning 2(3), 210–224 (2010)
6. Standard Vocabulary List 12000: SPACE ALC, `http://www.alc.co.jp/eng/vocab/svl/index.html` (in Japanese)