Pict-net Abstraction for EFL Writings and Classification Using the Normalized Compression Distance

Hoshi Journal of General Education

Volume 32, Pages 33-43, 2014

URL: http://id.nii.ac.jp/1240/00000372/
Pict-net Abstraction for EFL Writings and Classification Using the Normalized Compression Distance

Sayuri Yoshizawa-Watanabe
(Hoshi University, School of Pharmacy and Pharmaceutical Sciences)

1. Introduction

Foreign language teachers have been doing research on different ways to educate students to improve their composition skills. Since the tasks of rectifying students’ essays and assessing the degree of improvement in writing are time-consuming for most language teachers, P. Elbow introduced a ‘teacherless’ collaborative learning approach in 1973 [1]. The effectiveness of the approach has been examined and verified [2]-[7], and I have also conducted a number of researches on the effectiveness of the method and found that the peer review activities help students to improve logical thinking skills as well as self-correction skills [8]-[17]. Furthermore, I proposed a novel tool to assist the evaluation of essays by the Normalized Compression Distance (NCD) [18], as well as a new method called “pict-net abstraction”, a helping aid for students with inefficient comprehension level of the language in composition education [19]-[21]. This new collaborative approach was found successful in making students realize the discrepancies between the writer’s intention and the readers’ understanding by comparing their pictogram networks, helping to detect the ambiguity of the argument in their writings. This approach was developed in reference with the one using concept maps, graphical tools for organizing and representing knowledge [22][23].

The remainder of this paper is organized as follows. The next section presents a brief description of visualizing peer review processes in English as a Foreign Language (EFL) composition, and the NCD measure is briefly
described in the third section. The purpose of the research is stated in the fourth section, and the method of this experiment is in the fifth section. The sixth section shows the results and discussion, and the conclusions of this study and implications are in the final section.

2. Visualizing Peer Review Processes in EFL Writings

Language instructors have adopted peer review activities hoping that the activities stimulate students to have a broader perspective on the content and form of their essays. However, the writers’ motivation and self-confidence may increase or even decrease depending on the types of feedback. Students who do not enjoy learning the language tend to have much difficulty in making proper comments. Therefore, visualizing peer review processes in EFL writing classes would be very useful. A new method using pictogram network abstraction (hereinafter referred to as “pict-net abstraction”) can be used quite easily because it can be done by simply selecting pictograms from the provided pictogram lists and linking them. The good point is that the pict-net abstraction has excellent visibility and the results using this method will not be easily affected by surface errors such as minor spelling errors, typographical errors, grammatical or syntax errors.

Pictograms are pictorial signs and instructions that represent words, objects, or concepts, and speakers of different native languages can understand the meanings equally and effectively. Since they can be used without considering age, sex, and nationality, drawing a pict-net can become a helping guide to convey the understanding of others’ intentions.

A pict-net is a directed graph, a set of nodes connected by edges, where the edges have a direction associated with them [19]. Pict-nets can visualize knowledge and concepts with pictograms that are interconnected with one-direction and/or two-direction arrows. Figure 1 is an example of “pictogram network” showing that two people meet/met and become/became friends (and cooperate/cooperated with each other).
The distance of the pict-nets between a writer and a reviewer can be found using matrix representation and determined by the matrix distance measure:

\[
\sqrt{\frac{\text{sum}(\text{net}_{\text{writer}} - \text{net}_{\text{reviewer}})^2}{\text{average}_{\text{weight}}}}
\]

As shown in Table 1, “one” is inputted to the table if the pictograms are linked, and “zero” is inputted to the table if the pictograms are not linked. The following table shows a part of the matrix representation of a pict-net.

### Table 1 Matrix representation of pictogram networks

<table>
<thead>
<tr>
<th>Net_WA1</th>
<th>Pict_1</th>
<th>Pict_2</th>
<th>Pict_3</th>
<th>Pict_4</th>
<th>Pict_5</th>
<th>Pict_6</th>
<th>Pict_7</th>
<th>Pict_8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pict_1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pict_2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Pict_3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pict_4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pict_5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pict_6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pict_7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Pict_8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pict_9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pict_10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pict_11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pict_12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pict_13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pict_14</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 1 Example of “pictogram network”
3. Classifying Compositions by Using the Normalized Compression Distance

The normalized compression distance (NCD) is an outcome of the mathematical theoretical development based on the incomputable notion of algorithmic complexity Kolmogorov developed in the late 1960s [24]. It is applicable for clustering objects of any kind, such as music, texts, or gene sequences [25]-[27]. If x and y are two objects, and c(x) is the compressed size of x using compressor c, c(y) is the compressed size of y using compressor c, and c(x∙y) is the compressed size of concatenated objects x and y, then the NCD is:

\[
\text{NCD}(x, y) = \frac{c(x \cdot y) - \min\{c(x), c(y)\}}{\max\{c(x), c(y)\}}
\]

The NCD is expected to be between 0 and 1. The closer the NCD(x,y) is to zero, the more similar x and y are.

Some researchers have applied the NCD method to heterogeneous sets of web data, and their results indicated a high degree of similitude between the clustering structures obtained by the NCD [28]. In one research with 35 participants, the revised essays with the NCD values of .23 or lower contained few changes in content and structure and received no holistic score changes [18]. In this research, the NCD measure is also applied to the EFL compositions to see whether or not the value .23 is a valid threshold for classifying compositions in preestimating and evaluating the degree of difference in the contents of writings before and after the activity.

4. Purpose of the Research

The purpose of this research is to analyze the effectiveness of the pict-net abstraction as well as the adequacy of adopting the NCD measure for classifying essays. The experiments were conducted to see whether the visual summaries of the essays could be of help in detecting the ambiguity
and polysemy of the argument, to verify whether the main purport of the statements would be reflected in the pict-nets the reviewers created, and to find out whether compositions could be classified using the NCD measure.

5. Method of the Experiments

On the day of the experiment, a material with sample pictograms and explanations of how to make pict-nets was distributed. To minimize the difference in their English comprehension level, each participant was provided with an electronic dictionary that contained unabridged dictionaries of English-Japanese, Japanese-English, and English Collocations. In addition, a handout was distributed that described a basic style of essay writing and some useful information about the conjunctions.

5.1 Participants

The experiment was conducted by recruiting participants at a graduate school in Japan. The essay writers were five males and one female enrolled in the first year of a master’s course in engineering and two male adult working students. The reviewers were the same as above, and they were all Japanese.

5.2 Procedure

To verify whether or not what the writers wanted to tell was properly conveyed to the readers, the following procedures were used for the experiment.

1. An instructor determined an essay theme and two instructors prepared a list of pictograms.
2. On the day of the experiment a material regarding pict-net and the essay theme were provided.
3. Participants created a pict-net on the essay theme, and then wrote an essay.
4. Participants exchanged their essays and created the pict-net based on the essay.
5. Essay writers compared their own pict-nets with the reviewers’ pict-nets and revised their essays.
6. One new reviewer read each revised essay and created a pict-net based on the essay.
7. Participants stated opinions about this experiment.
8. The instructor disclosed the purpose of the experiment.

5.3 Essay Theme
The description of the essay theme in the handout was:

“There exist female-only carriages on some railways and subways. The introduction of such carriages was intended to reduce the number of sexual harassment cases and make women feel safe to use trains. In principle, only women can get on female-only carriages, but a number of railway companies allow young children, the aged, the injured, physically handicapped persons and their caregivers to use them. Considering the advantages and disadvantages of introducing female-only carriages, and the pros and cons of introducing male-only carriages, decide whether you are for or against the existence of female-only carriages. After deciding your position, create a pict-net to describe the structure of your composition by the use of provided pictograms interconnected with one-direction arrows (→) and/or two-direction arrows (←→). Then, write an essay of one hundred English words. The time limit for these tasks is a total of fifty minutes.”

The participants were allowed to use the provided 40 pictograms in Figure 2 to create pict-nets. They were permitted to repeatedly use one pictogram if necessary, but were told to mimic and draw the pictogram him/herself.

Figure 2  Pictogram lists distributed to participants
6. Results and Discussion

6.1 Distance between Pictogram Networks before and after the Peer Review Activities

The distance of the pict-nets between the essay writers and the reviewers before and after the peer review activity shortened except one case as shown in Table 2. This means that peer review using pict-net abstraction contributed to the better comprehension of the essay contents by the reviewers.

Table 2 Difference of the pict-nets between the writer and the reviewers

<table>
<thead>
<tr>
<th>Essay</th>
<th>Size of the original draft (byte)</th>
<th>Size of the revised draft (byte)</th>
<th>c(x)</th>
<th>c(y)</th>
<th>c(x−y)</th>
<th>NCD</th>
<th>Scores of the original draft</th>
<th>Scores of the revised draft</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>444</td>
<td>516</td>
<td>372</td>
<td>398</td>
<td>417</td>
<td>.11</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>661</td>
<td>695</td>
<td>469</td>
<td>484</td>
<td>510</td>
<td>.08</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>521</td>
<td>570</td>
<td>427</td>
<td>414</td>
<td>557</td>
<td>.33</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>703</td>
<td>603</td>
<td>538</td>
<td>466</td>
<td>756</td>
<td>.54</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>724</td>
<td>761</td>
<td>487</td>
<td>520</td>
<td>616</td>
<td>.25</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>442</td>
<td>435</td>
<td>382</td>
<td>372</td>
<td>457</td>
<td>.22</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>575</td>
<td>608</td>
<td>455</td>
<td>451</td>
<td>614</td>
<td>.36</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>548</td>
<td>561</td>
<td>367</td>
<td>380</td>
<td>430</td>
<td>.17</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

6.2 Size, NCD Results, and Holistic Scores of the Essays before and after the Peer Review Activities Using Pict-net Abstraction

The sizes, the NCD results, and the holistic scores of the original and the revised essays are shown in Table 3. The NCD values ranged from .08 to .54, which reflected the degree of similarities of the essays before and after the peer review activities with pict-net abstraction.

Table 3 Size, NCD results, and holistic scores of the essays

<table>
<thead>
<tr>
<th>Essay</th>
<th>Size of the original draft (byte)</th>
<th>Size of the revised draft (byte)</th>
<th>c(x)</th>
<th>c(y)</th>
<th>c(x−y)</th>
<th>NCD</th>
<th>Scores of the original draft</th>
<th>Scores of the revised draft</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>444</td>
<td>516</td>
<td>372</td>
<td>398</td>
<td>417</td>
<td>.11</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>661</td>
<td>695</td>
<td>469</td>
<td>484</td>
<td>510</td>
<td>.08</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>521</td>
<td>570</td>
<td>427</td>
<td>414</td>
<td>557</td>
<td>.33</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>703</td>
<td>603</td>
<td>538</td>
<td>466</td>
<td>756</td>
<td>.54</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>724</td>
<td>761</td>
<td>487</td>
<td>520</td>
<td>616</td>
<td>.25</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>442</td>
<td>435</td>
<td>382</td>
<td>372</td>
<td>457</td>
<td>.22</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>575</td>
<td>608</td>
<td>455</td>
<td>451</td>
<td>614</td>
<td>.36</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>548</td>
<td>561</td>
<td>367</td>
<td>380</td>
<td>430</td>
<td>.17</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
The original and the revised essays were assessed by a native English instructor, based on the adequacy and organization of the content, the cohesion and logicality of the text, and the overall quality. Four out of eight revised essays received higher scores than the original draft, but the other four remained the same. The ones with no holistic score changes had the NCD value below .22 in this case.

7. Conclusions and Implications

The experiment revealed that the “pict-net abstraction”, a collaborative learning approach using peer review activities and pictograms, was effective in that the visualized essay summary helped the writers to easily notice the ambiguity of the argument in their own essays, which led to better revisions. Although some participants felt uncomfortable to make comments on peer’s essays during the traditional peer review activities due to the lack of confidence in their language level, the participants of the peer review activities using pict-net abstraction found the method intriguing and quite useful. This research also supported the results of the threshold value in the original paper on the assessment of the impact of student peer review in writing instruction by using the normalized compression distance. However, more exercises using pictograms should be created to augment the effectiveness of the method and check whether or not every participant would be able to create the same pict-net from the provided pictograms. Further research should be carried out with larger number of participants to establish a more effective pict-net abstraction method as well as the NCD measure.

References


17(2), 149-154.

