PCA-based analysis on factors of English translation ability

Ting Hong*

Zijin College, Nanjing University of Science and Technology, Nanjing, 210000, China

Received 1 March 2014, www.cmnt.lv

Abstract

This work studied the relationship between English proficiency and translation ability as well as that between English and translation teaching among English majors. PCA is utilized to quantify and analyze the relationship between translation ability and English skills such as listening, reading, error correction and writing. Then we obtain the quantitative relation between translation ability and its factors, providing decision basis for the improvement of English translation skills and teaching.

Keywords: PCA analysis, translation ability, English teaching, quantitative analysis

1 Introduction

With the deepening of economic globalization, there is a growing demand for talents of foreign language translation. Therefore, the improvement of students’ English translation ability has become an important task for college English major and common English teaching.

For a long time, translation ability has been regarded as a language skill due to unclear concepts about translation and English teaching. Currently, the mainstream of translation teaching still focuses on language skills. While strengthening the teaching of language knowledge, it neglects the variability of texts for translation, differences in translators’ cognition and some non-linguistic factors (social, historical and cultural). However, translation behavior is a dynamically-generated process of translation and also a reflection of translators’ comprehensive abilities. As many scholars interpreting the nature of translation capabilities, they made it clear that comprehensive ability is not equal to language application ability.

2 Principal component analysis (PCA)

2.1 BASIC IDEAS OF PCA

Principal component analysis utilizes the method of mathematical dimension reduction to substitute comprehensive variables for original ones. These unrelated, comprehensive variables contain the information of the original ones as much as possible. The statistical analysis method converting several variables into few unrelated, comprehensive ones is called as principal component analysis (PCA).

2.2 MATHEMATICAL MODELS OF PCA ANALYSIS

For a data sample, \( p \) variables \((x_1, x_2, \ldots, x_p)\) are observed to form the data matrix of \( n \) samples as follows.

\[
X = \begin{pmatrix}
    x_{11} & x_{12} & \cdots & x_{1p} \\
    x_{21} & x_{22} & \cdots & x_{2p} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{n1} & x_{n2} & \cdots & x_{np}
\end{pmatrix} = (x_1, x_2, \ldots, x_p),
\]

where

\[
x_j = \begin{pmatrix}
    x_{j1} \\
    x_{j2} \\
    \vdots \\
    x_{jn}
\end{pmatrix}, \quad j = 1, 2, \ldots, p.
\]

PCA can integrate \( p \) observed variables into \( p \) new comprehensive ones.

\[
F_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p
\]

\[
F_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p
\]

\[
\vdots
\]

\[
F_p = a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p
\]

Namely

\[
F_j = \alpha_{j1}x_1 + \alpha_{j2}x_2 + \cdots + \alpha_{jp}x_p, \quad j = 1, 2, \ldots, p
\]

The model should meet the following criteria.

1) \( F_i, F_j \) \((i \neq j, i, j = 1, 2, \ldots, p)\) are unrelated to each other.

2) \( F_1 \) has a larger variance than \( F_2 \), and \( F_2 \) has a larger variance than \( F_3 \), etc.

3) \( a_{11}^2 + a_{21}^2 + \cdots + a_{pp}^2 = 1, \quad k = 1, 2, \ldots, p.\)

There are \( p \) principal components. \( F_1 \) is called as the first principal component; \( F_2 \) as the second principal

*Corresponding author’s e-mail: zj_ht123@163.com
component, and so on; \( a_{ij} \) as the principal component factor.

The model can be represented by a matrix. \( A \) is the principal component coefficient matrix.

\[
F = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix},
\]

\[
F = AX , \quad \text{where}
\]

\[
A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}.
\]

3 Data sources

As psychological characteristics, English translation and language skills cannot be measured only by one’s external manifestations or characteristics.

As the study objective of this work, TEM-8 score of a college includes a total score and individual score. TEM-8, with high reliability and validity, is the most authoritative, standardized test for measuring students’ English proficiency.

For data processing and analysis, listening and dictation are merged into listening comprehension, measuring students’ listening ability; Reading One and Two are merged into reading comprehension to measure English reading ability; translation from English-Chinese and Chinese-English translation are also combined together.

In terms of differences in English translation ability of candidates, translation scores are divided into two groups— one with scores not less than 48 points (qualifying criterion) and the other less than 48 points.

4 Results analysis

4.1 MEAN CONTRAST OF CHINESE-ENGLISH TRANSLATION SCORES

Mean data contrast of the two groups and T-test results of independent samples are shown in Table 1, Table 2, respectively.

According to Table 1 and Table 2, the results of Levene variance homogeneity test are \( F = 0.772 \) and \( P = 0.381 > 0.05 \), so the two groups of students have equal variance in Chinese-English translation score; \( t = 3.303 \), and \( P = 0.001 < 0.05 \). Thus, the difference between two groups in Chinese-English translation has statistical significance.

### TABLE 1 Mean contrast for two groups of Chinese-English translation scores

<table>
<thead>
<tr>
<th>Translation Score</th>
<th>N</th>
<th>Mean</th>
<th>Std.Deviation</th>
<th>Std.Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;=48.00</td>
<td>40</td>
<td>6.35</td>
<td>.796</td>
<td>.140</td>
</tr>
<tr>
<td>&lt;48.00</td>
<td>94</td>
<td>5.71</td>
<td>1.031</td>
<td>.112</td>
</tr>
</tbody>
</table>

### TABLE 2 Independent samples T-test for two groups of Chinese-English translation scores

<table>
<thead>
<tr>
<th>Levene’s Test for Equality of Variances</th>
<th>T-test for Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Sig.</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
</tr>
<tr>
<td>Equal variance assumed Equal variance not assumed</td>
<td>.772</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2 MEAN CONTRAST OF CHINESE-ENGLISH TRANSLATION

Table 3 and Table 4 show the mean contrast, independent sample T-test of two groups of data, respectively.

### TABLE 3 Mean contrast of two groups of Chinese-English translation scores

<table>
<thead>
<tr>
<th>Translation Score</th>
<th>N</th>
<th>Mean</th>
<th>Std.Deviation</th>
<th>Std.Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;=48.00</td>
<td>40</td>
<td>6.74</td>
<td>1.067</td>
<td>.183</td>
</tr>
<tr>
<td>&lt;48.00</td>
<td>94</td>
<td>5.89</td>
<td>1.457</td>
<td>.161</td>
</tr>
</tbody>
</table>
The statistical contrast in Chinese-English and English-Chinese translation shows that students with stronger English ability are doing better in English translation.

4.3 PEARSON CORRELATION ANALYSIS

For data of all the samples, Table 5 shows pearson correlation coefficients among Chinese-English and English-Chinese translation, listening comprehension, reading comprehension, error correction, writing and translation score.

According to Table 5, Chinese-English translation has the biggest correlation coefficient (0.454) with English-Chinese translation. It indicates that Chinese-English and English-Chinese ability are correlative, so teaching for the two kinds of translations should be complementary; Chinese-English translation has strong correlation (0.333) with translation score, so it is closely related to listening, writing and reading; error correction and Chinese-English score have the strongest correlation (0.332) among all individual scores. With language knowledge including grammar, rhetoric and language structure, error correction can greatly influence translation ability by identifying faulty wording or formulation and proposing corrective approach.

4.4 PCA ANALYSIS

4.4.1 Statistical test

After setting the significance level as 0.05, KMO and Bartlett's Test of Sphericity are conducted on the selected data. Results are shown in Figure 1.

![Score Plot](image)

**KMO and Bartlett's Test**

<table>
<thead>
<tr>
<th>Measure of Sampling Adequacy</th>
<th>KMO</th>
<th>Bartlett's Test of Sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approx. Chi Square</td>
<td>56.845</td>
<td>df = 26</td>
</tr>
<tr>
<td>Sig.</td>
<td>.001</td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 1 KMO and Bartlett’s Test of Sphericity**

In Figure 1, KMO statistic is 0.604 > 0.5; Bartlett’s Test of Sphericity is significant at level 0.05, the null hypothesis is rejected, and PCA analysis is more suitable for the index data of translation ability.

4.4.2 Determination of common factors

In this work, eigenvalues is greater than one, and the cumulative contribution rate greater than 80%. Figure 2 shows the PCA extraction results.

In Figure 2, the characteristic curve shows a turning point at factor 5, so common factors can be extracted from the first five factors. In Table 6, among the five factors with eigenvalue greater than one, the first four factors contribute to 80.57% (>) 80%) of cumulative rate. Therefore, the four factors can reflect the influencing factors of English translation ability.
### 4.4.3 Establishment of factor loading matrix

Varimax rotation method is used to calculate factor loading matrix. Results are shown in Table 7.

**TABLE 7** Factor loading matrix

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese to English</td>
<td>.926</td>
<td>-.084</td>
<td>.058</td>
<td>-.183</td>
</tr>
<tr>
<td>English to Chinese</td>
<td>.914</td>
<td>-.197</td>
<td>.187</td>
<td>.204</td>
</tr>
<tr>
<td>Correction</td>
<td>.906</td>
<td>.142</td>
<td>.227</td>
<td>.059</td>
</tr>
<tr>
<td>Writing x4</td>
<td>.901</td>
<td>-.129</td>
<td>.220</td>
<td>.270</td>
</tr>
<tr>
<td>Reading 1x5</td>
<td>.881</td>
<td>.359</td>
<td>.171</td>
<td>.002</td>
</tr>
<tr>
<td>Reading 2x6</td>
<td>.215</td>
<td>.888</td>
<td>.184</td>
<td>.060</td>
</tr>
<tr>
<td>Listening x7</td>
<td>-.437</td>
<td>.758</td>
<td>-.060</td>
<td>.375</td>
</tr>
<tr>
<td>Dictation x8</td>
<td>.358</td>
<td>.562</td>
<td>-.277</td>
<td>-.210</td>
</tr>
</tbody>
</table>

### 4.4.4 Scores of calculated factors

Score linear calculation model can be constructed for each factor based on the factor score matrix in Table 8, and the specific steps are as follows.

\[
F_1 = 0.170x_1 + 0.047x_2 + \ldots - 0.101x_8 ,
\]

\[
F_2 = -0.039x_1 - 0.052x_2 + \ldots + 0.298x_8,
\]

\[
F_3 = -0.018x_1 - 0.470x_2 + \ldots + 0.063x_8,
\]

\[
F_4 = 0.120x_1 + 0.087x_2 + \ldots + 0.216x_8,
\]

\[
F = 0.35189F_1 + 0.16511F_2 + 0.16413F_3 + 0.12459F_4 .
\]

Then mathematical evaluation model of English translation ability can be concluded based on variance contribution rate of common factors in Table 6.

**TABLE 8** Factor score matrix

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese to English</td>
<td>.170</td>
<td>-.039</td>
<td>-.018</td>
<td>.120</td>
</tr>
<tr>
<td>English to Chinese</td>
<td>.047</td>
<td>-.052</td>
<td>-.470</td>
<td>.087</td>
</tr>
<tr>
<td>Correction</td>
<td>.180</td>
<td>.150</td>
<td>-.021</td>
<td>-.019</td>
</tr>
<tr>
<td>Writing</td>
<td>.212</td>
<td>-.031</td>
<td>-.114</td>
<td>-.121</td>
</tr>
<tr>
<td>Reading 1x5</td>
<td>.090</td>
<td>-.031</td>
<td>.094</td>
<td>-.327</td>
</tr>
<tr>
<td>Reading 2x6</td>
<td>-.017</td>
<td>-.023</td>
<td>.020</td>
<td>.252</td>
</tr>
<tr>
<td>Listening x7</td>
<td>.145</td>
<td>-.118</td>
<td>-.295</td>
<td>.356</td>
</tr>
<tr>
<td>Dictation x8</td>
<td>-.101</td>
<td>.298</td>
<td>.063</td>
<td>.216</td>
</tr>
</tbody>
</table>

### 5 Conclusions

In this work PCA analysis reveals the relationship between English proficiency and translation ability, as well as the quantitative relationship between English and translation teaching among English major students. Then the quantitative relationship between translation ability and English abilities – listening and reading comprehension, Chinese-English and English-Chinese translation – is obtained through extracting factors and calculating factor scores. In the end, quantitative calculation and evaluation for English translation are achieved, thus providing decision basis for teaching English translation.
References


Author

Ting Hong, born in Oktober 1980, Nanjing, Jiangsu Province, China.

Current position, grades: Master, a lecturer of Zijin College, Nanjing University of Science and Technology, China.

University studies: American literature, English teaching and Research.

Scientific interest: English teaching and research, computer aided Instruction.

Publications: 15 papers.