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 $(x_1, x_2, \cdots, x_p)$ are observed to form the data matrix of n samples as follows.

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

where

$$x_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}, \qquad j = 1, 2, \cdots, p.$$

PCA can integrate p observed variables into p new comprehensive ones.

$$\begin{cases} F_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \\ F_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \\ \cdots \\ F_p = a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p \end{cases}.$$

Namely

$$F_j = \alpha_{j1}x_1 + \alpha_{j2}x_2 + \cdots + \alpha_{jp}x_p,$$
$$j = 1, 2, \cdots, p$$

The model should meet the following criteria.

1) F_i, F_j $(i \neq j,\ i, j = 1, 2, \cdots, p)$ 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) $a_{k1}{}^2 + a_{k2}{}^2 + \cdots + a_{kp}{}^2 = 1, \qquad k = 1, 2, \cdots, p.$

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.

$$F = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{pmatrix}, \qquad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}.$$

$F = AX$, where

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

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

Translation Score	N	Mean	Std.Deviation	Std.Error Mean
>=48.00	40	6.35	.796	.140
<48.00	94	5.71	1.031	.112

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

Levene's Test for Equality of Variances		T-test for Equality of Means							
	F	Sig.	t	df	Sig.(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variance assumed Equal variance not assumed	.772	.381	3.303	122	.001	.64	.194	.257	1.024
			3.693	80.182	.000	.64	.173	.295	.985

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.

According to Table 3 and Table 4, the results of Levene variance homogeneity test are F = 2.534 and P = 0.115> 0.05, so variances of two groups of Chinese-English translation scores are equal: t = 3.158, and P = 0.002 <0.05. Thus, the difference between the two groups in Chinese-English translation has statistical significance.

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

Translation Score	N	Mean	Std.Deviation	Std.Error Mean
>=48.00	40	6.74	1.067	.183
<48.00	94	5.89	1.457	.161

TABLE 4 Independent samples T-test of two groups of Chinese-English translation scores

Levene's Test for Equality of Variances		T-test for Equality of Means							
	F	Sig.	t	df	Sig.(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variance assumed	2.534	0.115	3.158	122	.002	.86	.275	.320	1.391
Equal variance not assumed			3.613	85.121	.002	.86	.237	.385	1.330

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.

TABLE 5 Pearson correlation analysis results

		Listening Comprehension	Reading Comprehension	Error Correction	Writing	Translation Score	Chinese to English	English to Chinese
Chinese to English	Pearson Correlation	.273(**)	.214(*)	.332(**)	.124	.333(**)	1	.454(**)
	Sig.(2-tailed)	.002	.017	.000	.171	.000	.000	.000
English to Chinese	Pearson Correlation	.403(**)	.348(**)	.271(**)	.041	.431(**)	.454(**)	1
	Sig.(2-tailed)	.000	.000	.002	.650	.000	.000	.000

** Correlation is significant at level 0.01 (2-tailed)
* Correlation is significant at level 0.05 (2-tailed)

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.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.604
Bartlett's Test of Sphericity	Approx. Chi-Square	56.843
	df	28
	Sig.	.001

FIGURE 1 KMO and Bartlett's Test of Sphericity

In Figure 1, KMO statistic is 0.604 > 0.5; Bartlett spherical statistic 56.843; significant probability 0.001 <0.05. Then the null hypothesis in Bartlett Test of Sphericity 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 80.571% (> 80%) of cumulative rate. Therefore, the four factors can reflect the influencing factors of English translation ability.

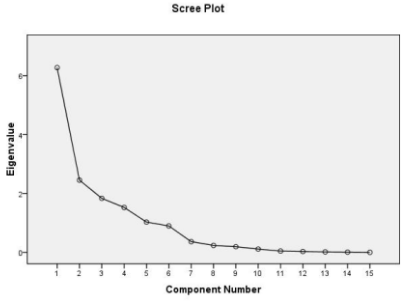

FIGURE 2 Scree plot

TABLE 6 Eigenvalues of correlation matrix R

	Initial Eigenvalue			Eigenvalue after extracting four factors and orthogonal rotation		
	Eigenvalue	**Proportion**	**Cumulation**	**Eigenvalue**	**Proportion**	**Cumulation**
1	6.274	41.829	41.829	5.278	35.189	35.189
2	2.453	16.352	58.181	2.477	16.511	51.700
3	1.834	12.227	70.408	2.462	16.413	68.112
4	1.525	10.164	80.571	1.869	12.459	80.571
5	1.028	6.853	87.424			
6	0.897	5.977	93.402			
7	0.362	4.415	97.816			
8	0.235	2.184	100			

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

	Component			
	1	2	3	4
Chinese to English x1	.926	-.084	.058	-.183
English to Chinese x2	.914	-.197	.187	.204
Correction x3	.906	.142	.227	.059
Writing x4	.901	-.129	.220	.270
Reading 1x5	.881	.359	.171	.002
Reading 2x6	.215	.888	.184	.060
Listening x7	-.437	.758	-.060	.375
Dictation x8	.358	.562	-.277	-.210

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.

$$F_1 = 0.170x_1 + 0.047x_2 + ... - 0.101x_8,$$

$$F_2 = -0.039x_1 - 0.052x_2 + ... + 0.298x_8,$$

$$F_3 = -0.018x_1 - 0.470x_2 + ... + 0.063x_8,$$

$$F_4 = 0.120x_1 + 0.087x_2 + ... + 0.216x_8,$$

$$F = 0.35189F_1 + 0.16511F_2 + 0.16413F_3 + 0.12459F_4.$$

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

	Component			
	1	2	3	4
Chinese to English x1	.170	-.039	-.018	.120
English to Chinese x2	.047	-.052	-.470	.087
Correction x3	.180	.150	-.021	-.019
Writing x4	.212	-.031	-.114	-.121
Reading 1x5	.090	-.031	.094	-.327
Reading 2x6	-.017	-.023	.020	.252
Listening x7	.145	-.118	-.295	.356
Dictation x8	-.101	.298	.063	.216

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

[1] Lao, She. Camel Xiangzi. trans. Shi Xiaoqing. Beijing: Foreign Languages Press. 1981 Leech G N Style in Fiction *New York Longman* 2010 11-16

[2] Li C N, Thompson S A 2011 Subject and topic: a new typology of language *In Charles N Li (ed) Subject and Topic London/New York Academic Press* 457-90

[3] Newmark P 1988 A Textbook of Translation Hertfordshire: Prentice-Hall Newmark Peter Approaches to Translation *Oxford Pergamon* 200 95-101

[4] Nida E 2011 A Language, Culture and Translating *Shanghai: Shanghai Foreign Language Education Press* 312-5

[5] Nord C 2007 Translating as a Purposeful Activity: Functionalist Approaches Explained *Manchester St Jerome* 67-74

[6] Quirk R 2011 A Grammar of Contemporary English *Longman* 34-8

[7] Robinson D 2002 Western Translation Theory: from Herodotus to Nietzsche *Cornwall: St. Jerome Publishing* 112-20

[8] Snell-Hornby M 2011 Translation Studies An Integrated Approach *Revised ed Amsterdam John Benjamins* 321-30

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.