



Predicting Machine Translation Comprehension with a Neural Network

Milam Aiken, Jamison Posey, Bart Garner, Brian Reithel
University of Mississippi, School of Business Administration, University, MS 38677
maiken@bus.olemiss.edu
jposey@bus.olemiss.edu
bgarner@bus.olemiss.edu
breithel@bus.olemiss.edu

ABSTRACT

Comprehension of natural language translations is dependent upon several factors including textual variables (grammatical, spelling, and word choice errors, sentence complexity, etc.) and human variables (language fluency, topic knowledge, motivation, dyslexia, etc.). An individual reader's understanding of machine-generated translations can vary widely because of the lower accuracy usually associated with this technology. Prior studies have had mixed results in predicting which variables have the greatest influence on translation comprehension. In the current study, we employ an artificial neural network to analyze survey responses and reading test scores, resulting in a significantly correlated forecast of reading comprehension. Thus, we are able to offer better predictions to identify which readers might have a better grasp of content from garbled translations.

Indexing terms/Keywords

Comprehension, Neural network, Machine translation

Academic Discipline And Sub-Disciplines

Computer Science (machine translation, neural networks)

TYPE (METHOD/APPROACH)

Survey

Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol. 15, No. 2

www.ijctonline.com, editorijctonline@gmail.com



INTRODUCTION

Machine translation often provides fast and cheap transliteration that is particularly useful for situations in which human interpreters are not available or the foreign source text might not be important enough to pay for professional translation services[8]. This automated translation is often performed with Web programs such as *Google Translate* that are becoming increasingly accurate and provide support for more and more languages, but performance still suffers with many language pairs[7,16].

Even if machine translation accuracy is low with incorrect word choices, poor grammar, and other errors, some meaning can often be found and these translations can still be useful to the reader[11]. In addition, some people are better able to understand poor translations than others, with wide-ranging comprehension rates for equivalent texts [1,12]. However, very little is known about why some readers have better comprehension than others. With more research, we might be able to predict which target audiences are more likely to benefit from machine translation services. For example, those with more motivation to understand might be able to grasp meanings more readily than other readers.

Prior research has shown that many variables might affect reading comprehension, including topic knowledge, critical thinking skills and language fluency (e.g., [10, 13, 14]). One study [3] using 80 undergraduate students found a significant but weak and negative correlation between results from their performance on a Cloze comprehension test [17] and understanding of some, but not all, machine-translated text ($R = -0.225$, $p = 0.045$). In another study [15], 96 undergraduate college students read poor machine translations and were asked to write in English what they thought was meant, using correct grammar and word choices. Results showed a significant, positive correlation between grammatical ability and comprehension of some of the text ($R = .207$, $p = .043$) and a significant, but negative, correlation between their self-assessed foreign language fluency and comprehension of other parts of the text ($R = -0.210$, $p = 0.040$).

In this paper, we conduct a new study to assess which variables are most important for human understanding. An artificial neural network and multilinear analysis are applied in an attempt to yield a more predictive model. The study then concludes with a summary, limitations, and directions for further research.

SURVEY

Subjects and Task Description

The purpose of this study is to predict how well a reader might comprehend a passage of text that could be translated by machine. We recruited 121 undergraduate business students from a university in the southern United States to assess possible factors that might impact comprehension, including language fluency, motivation, and subject knowledge. Motivation was measured by a 7-item (1=disagree, 7=agree), self-assessed Likert scale with two statements: "I was interested in this task." (INT) and "I was motivated to understand the sentences." (MOT). Subject knowledge was evaluated based upon the statement "I know the subject matter of these sentences." (SUBJ), and language fluency was assessed based upon the statements "I know English very well." (ENG), "I have a large English vocabulary." (VOC), "I rarely make grammatical mistakes." (GRAM), and "I know a non-English language well." (LANG). In addition, the students took the short Cloze test shown below as an additional means of assessing their language fluency (Source: <http://www.testyourenglish.net/english-online/cloze-reading/cloze1.html>). Correct answers are shown in bold.

Can we see (1) ___ the earth is a globe? Yes, we can, when we watch a ship that sails out to sea. If we watch closely, we see that the ship begins (2) ___. The bottom of the ship disappears first, and then the ship seems to sink lower and lower, (3) ___ we can see only the top of the ship, and then we see nothing at all. What is hiding the ship from us? It is the earth. Stick a pin most of the way into an orange, and (4) ___ turn the orange away from you. You will see the pin disappear, (5) ___ a ship does on the earth.

1. A. if B. where **C. that** D. whether E. when
2. A. being disappeared B. to be disappeared C. to have disappeared **D. to disappear** E. having disappeared
3. **A. until** B. since C. after D. by the time E. unless
4. A. reluctantly B. accidentally **C. slowly** D. passionately E. carefully
5. A. the same B. alike **C. just as** D. by the way E. similar to

Next, 62 of the students (Group 1, 51.2%) were given a reading comprehension test with no errors in the text while the remainder of the subjects (Group 2) were given the reading comprehension test shown in Appendix 1 with the same text translated by *Google Translate* to Bengali and back again to English, resulting in several grammatical and word-choice errors, evident in the Appendix. Bengali was chosen because of the relatively poor translation accuracy obtained with the software [1]. The original source text was not considered to be difficult (Flesch Reading Ease = 44.3 on a scale of 0 = difficult to 100 = easy, Flesch-Kincaid Grade Level = 11.3, or high-school level).

Summary Results

Table 1 shows a summary of the variables. The students considered themselves to be fluent in English, as they self-reported knowing their native language well, making few grammatical errors, and having a good vocabulary. In addition, the 79.3% overall score on the Cloze test was well above the 58% threshold for an "independent reading" level [5]. However, few students reported knowing a foreign language well. The students were also motivated and interested. There



were no significant differences between the two groups of students on these measures. There was a significant difference between the easy-text and difficult-text groups in terms of scores on the reading comprehension tests, however, because the latter contained several errors ($F = 14.40, p < 0.001$).

A Pearson correlation analysis (Table 2) shows that only their self-assessed grammatical knowledge was significantly correlated with text comprehension. However, the association was very weak. Interest and motivation were almost significantly correlated with text comprehension, but the relationships were weak as well. Thus, the results could not find strong determinants of reading comprehension; this result is consistent with the prior studies discussed earlier.

Table 1. Overall summary of variables

Group		Cloze	Text	INT	MOT	ENG	VOC	GRAM	LANG	SUBJ
1	Mean	78.4%	63.2%	4.63	5.23	5.71	5.24	4.31	2.24	4.95
	StdDev	15.1%	27.9%	1.56	1.43	1.26	1.25	1.37	1.63	1.53
2	Mean	80.3%	45.8%	4.24	4.85	5.88	5.32	4.53	2.44	4.32
	StdDev	13.1%	22.3%	1.32	1.46	1.12	1.15	1.30	1.74	1.61
All	Mean	79.3%	54.7%	4.44	5.04	5.79	5.28	4.41	2.34	4.65
	StdDev	14.1%	26.7%	1.45	1.45	1.19	1.20	1.34	1.68	1.59

(All Likert-scale variables significantly different from neutral value of 4.0 at $\alpha = 0.05$)

Table 2. Correlation analysis (with p-values)

	Text	Cloze	INT	MOT	ENG	VOC	GRAM
Text	1	0.061 (.503)	0.159 (.081)	0.152 (.096)	0.055 (.552)	0.120 (.191)	0.197** (.030)
Cloze		1	0.063 (.493)	0.148 (.106)	0.101 (.271)	0.169 (.065)	0.191** (.036)
Int			1	0.619** (.001)	0.096 (.294)	0.063 (.495)	0.036 (.321)
Mot				1	0.024 (.791)	0.012 (.892)	0.013 (.885)
Eng					1	0.696** (.001)	-0.047 (.605)
Voc						1	0.509** (.001)
Gram							1

NEURAL NETWORKS

Artificial neural networks are mathematical models of the biological processes that occur in brains when learning new material and have been proven to be more accurate than competing forecasting tools such as multi-linear regression, autocorrelation, and logistic regression [9]. While the models do not provide predictive significance values, they do not require any statistical assumptions. Therefore, when using a neural network, no tests need to be conducted for the standard assumptions required by other parametric statistical techniques.

Neural networks are trained by exposing the network to individual examples of the data to be used for predictions or classifications. The process is repeated until the neural network recognizes underlying patterns between inputs (independent variables) and outputs (dependent variables).

In this study, only one significant, weak correlation could be found to forecast reading comprehension, but a combination of several factors might add predictive power. First, as a basis for comparison against traditional statistical techniques, we developed a multilinear regression model using SPSS. A random sample of 100 observations was used for development of the model, giving the results shown in Table 3 ($F = 3.640, p < 0.001$):



Table 3: Multilinear coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	24.048	19.672		1.222	.225
Group	-4.711	1.741	-.254	-2.706	.008
Cloze	3.096	3.645	.080	.849	.398
Int	4.782	16.740	.035	.286	.776
Voc	-.254	3.122	-.011	-.081	.935
Eng	-2.679	2.836	-.121	-.945	.347
Mot	1.851	2.449	.092	.756	.452
Subj	5.552	1.831	.310	3.032	.003
Lang	-.343	.164	-.192	-2.089	.040
Gram	3.659	2.415	.179	1.515	.133

The remaining 21 observations were used for evaluation of the multilinear regression model's efficacy. The mean absolute percentage error (MAPE) between the actual and forecasted values was 40.48%, and the correlation between these values was weak and insignificant ($R = 0.275$, $p = 0.228$).

Next, a neural network was developed using NeuroForecaster, a commercial product from NIBS, Inc. [6]. The network used the genetic algorithm. Training was conducted until the 100-observation in-sample mean absolute percentage error MAPE was reduced down to approximately 11.72%, as shown in Figure 1. The forecast of the 21-item hold-out sample is shown in Figure 2, with a MAPE of 12.2%. Further, the correlation between the actual and forecasted values was relatively strong and significant ($R = 0.489$, $p = 0.025$).

Figure1. Neural network training
(actual values shown in red, forecasted values shown in blue)

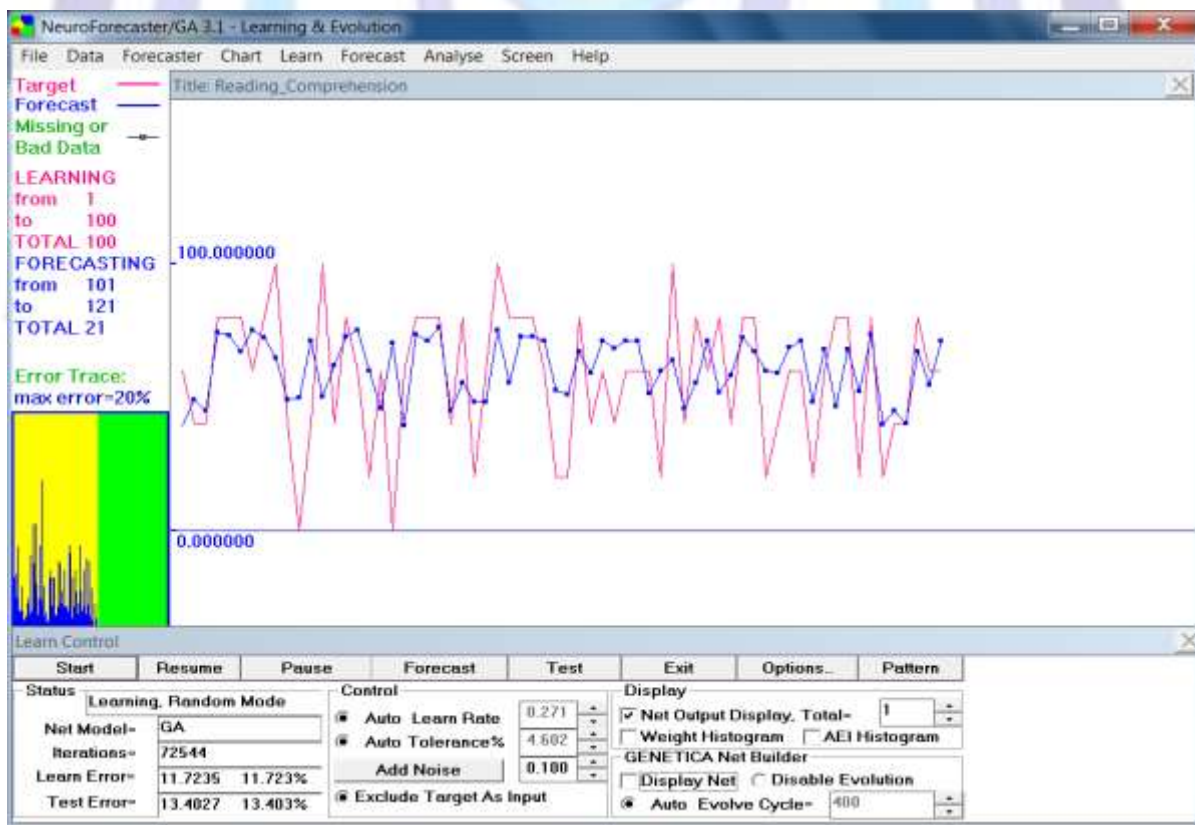
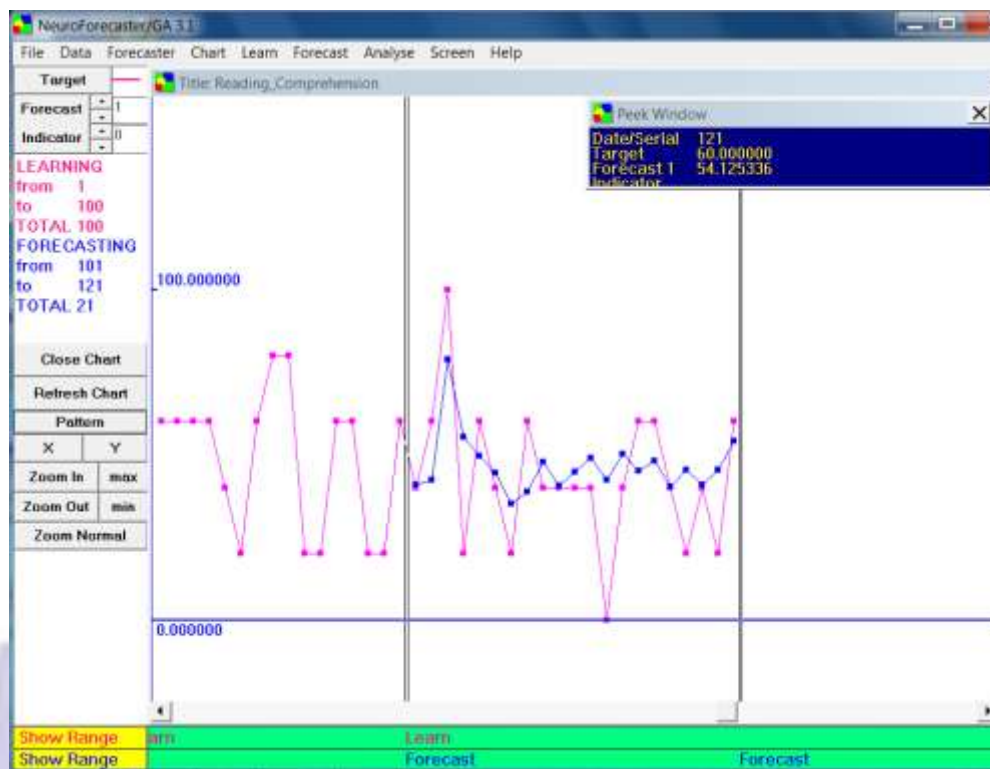


Figure2. Neural network forecasting
(actual values shown in red, forecasted values shown in blue)



CONCLUSION

Summary

Predicting reading comprehension is a complex problem with many interrelated, uncertain variables, which has made it difficult for previous studies to find strong predictive models using traditional statistical techniques. In this study, using data arising from self-reported assessments of language fluency, motivation, and topic knowledge, as well as results from a Cloze test, we have demonstrated that an artificial neural network can be used to make significantly better forecasts of who might be able to comprehend machine-translated passages of text.

Limitations

The study has several limitations. First, only one short reading comprehension test and one Cloze test were administered. This limited approach could yield an incomplete measure of the subjects' true understanding of the translated text. Future studies will need to examine these phenomena in a wider variety of topics and languages. Second, there was no variation of the reading difficulty level of the text or the number of errors in the passage, which could be additional sources of predictive power in more complicated experimental situations. Third, perhaps the college students' reading comprehension abilities were too similar to adequately explore the full range of effects arising from the independent variables. Finally, other unknown factors might be involved in the highly complicated mental process of comprehending machine translations.

Future Research

As mentioned above, a more thorough study with more comprehensive Cloze and reading comprehension tests might be necessary. It would also be beneficial to conduct additional studies that incorporate greater variation in subjects' abilities and levels of motivation.

REFERENCES

- [1] Aiken, M. and Balan, S. (2011). An analysis of Google Translate accuracy. *Translation Journal*. 16(2).
- [2] Aiken, M., Balan, S., Vanjani, M., and Garner, B. (2010). The effect of comment errors in multilingual electronic meetings. *Communications of the International Information Management Association*. 10(4), 49-60.
- [3] Aiken, M., Park, M., and Lindblom, T. (2013). Language fluency as a factor in machine translation comprehension. *International Journal of Computers and Technology*. 10(2), 1349-1355.



- [4] Aiken, M., Wang, J., Gu, L., and Paolillo, J.(2011). An exploratory study of how technology supports communication in multilingual groups. *International Journal of e-Collaboration*, 7(1), 17-29.
- [5] Chatal, R. (2001). Diagnostic and instructional uses of the Cloze procedure. *The Nera Journal*, 37(1), 3-6.
- [6] Dostál, P. (2013). The use of soft computing methods for forecasting in business, their applications in practice. In *Nostradamus: Modern Methods of Prediction, Modeling and Analysis of Nonlinear Systems* (pp. 49-60). Springer Berlin Heidelberg.
- [7] ElShiekh, A. (2012). Google translate service: Transfer of meaning, distortion or simply a new creation? An investigation into the translation process & problems at google. *English Language and Literature Studies*, 2(1), 56.
- [8] García, I. (2010). Is machine translation ready yet? *Target*, 22(1), 7-21.
- [9] Maren, A., Harston, C., and Pap, R. (2014). *Handbook of neural computing applications*. Academic Press.
- [10] McVay, J. and Kane, M. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology*, 141(2), 302-320.
- [11] Pepper, W., Aiken, M., and Garner, B. (2011). Usefulness and usability of a multilingual electronic meeting system. *Global Journal of Computer Science and Technology*, 11(10), 35-40.
- [12] Posey, J. and Aiken, M. (2015). Large-scale, distributed, multilingual, electronic meetings: A pilot study of usability and comprehension. *International Journal of Computers and Technology*, 14(3), 5578-5585.
- [13] Schmitt, N., Jiang, X., and Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26-43.
- [14] Sénéchal, M. (2006). Testing the home literacy model: Parent involvement in kindergarten is differentially related to grade 4 reading comprehension, fluency, spelling, and reading for pleasure. *Scientific Studies of Reading*, 10(1), 59-87.
- [15] Vanjani, M., Aiken, M., and Park, M. (2015). A study of factor sinfluencing machine translation comprehension. *Quarterly Review of Business Disciplines*, 2(2), 121-132.
- [16] Wade, R. (2011). Try Google Translate to overcome language barriers. *BMJ*, 343.
- [17] Williams, R., Ari, O., and Santamaria, C. (2011). Measuring college students' reading comprehension ability using Cloze tests. *Journal of Research in Reading*, 34(2), 215-231.



Authors' biographies



Dr. Milam Aikenis is a Professor and Chair of Management Information Systems in the School of Business Administration at the University of Mississippi. His research interests include machine translation and artificial neural networks.



Dr. Jamison Posey is a Clinical Assistant Professor of Management Information Systems in the School of Business Administration at the University of Mississippi. His research interests include machine translation, multilingual meeting systems, and impression management used in outsourced information systems development teams.



Dr. Bart Garner is a Clinical Assistant Professor of Management Information Systems in the School of Business Administration at the University of Mississippi. His research interests include machine learning and multilingual meetings in an organizational setting.



Dr. Brian Reithelis is a Professor of Management Information Systems in the School of Business Administration at the University of Mississippi. His research interests include digital forensics, software development processes (including team communication enhancement techniques), and the strategic use of information technology in an organizational setting.



Appendix

Reading Comprehension Test translated from Bengali

(English original source taken from <http://www.testprepreview.com/modules/reading1.htm>)

Marie is one of the most accomplished scientists. Together with her husband, Pierre, the radium, a material widely used for the treatment of cancer, uranium and other radioactive material was discovered and studied. Pierre and Marie's friendly cooperation has helped unlock the secrets of the Atom.

Marie, where his father was a professor of physics at Warsaw, Poland, in 1867, was born in. At an early age he displayed a brilliant mind and a jovial personality. His great enthusiasm for learning after high school asks her to continue with her studies. University of Warsaw when he learned that he was close to the women, however, became disgruntled. Due to a higher education, he left Poland and in 1891 he earned his master's degree and doctorate in physics at the Sorbonne, where, in order to enter a French university.

Marie and Pierre Curie, one of whom was his day, the greatest scientist was fortunate to have studied at the Sorbonne. Marie and Pierre married in 1895 and together with many productive years were spent in physics laboratory. A short time after the discovery of radium, Marie Pierre 1906, a horse-drawn wagon was shot stunned by this horrible misfortune and suffering endured deadly. Despondently their close relationship and share the joy of scientific research, they recalled. He was by himself to raise two young daughters that greatly increased her distress.

Curie's sense of solitude as a professor of physics at the Sorbonne, she requested her husband to succeed when starting to fade. She was the first woman to world-famous universities. In 1911, he was awarded the Nobel Prize for Chemistry for isolating radium. Finally, the long exposure to radium Marie Curie suffered from a serious illness, however, he became disillusioned about his work. Regardless of the outcome, he had dedicated himself to revealing the mysteries of science and the physical world.

1. The Curies' ____ collaboration helped to unlock the secrets of the atom.
A. friendly
B. competitive
C. courteous
D. industrious
E. chemistry
2. Marie had a bright mind and a ____ personality.
A. strong
B. lighthearted
C. humorous
D. strange
E. envious
3. When she learned that she could not attend the university in Warsaw, she felt ____.
A. hopeless
B. annoyed
C. depressed
D. worried
E. none of the above
4. Marie ____ by leaving Poland and traveling to France to enter the Sorbonne.
A. challenged authority
B. showed intelligence
C. behaved
D. was distressed
E. answer not available in article
5. ____ she remembered their joy together.
A. Dejectedly
B. Worried
C. Tearfully
D. Happily
E. Sorrowfully
6. Her ____ began to fade when she returned to the Sorbonne to succeed her husband.
A. misfortune
B. anger
C. wretchedness
D. disappointment
E. ambition



7. Even though she became fatally ill from working with radium, Marie Curie was never _____.

- A. troubled
- B. worried
- C. disappointed
- D. sorrowful
- E. disturbed

