Attention-based Deep Learning Model for Text Readability Evaluation

1st Yuxuan Sun
School of Computer and Computing Science
Zhejiang University City College
Hangzhou, China
sunyuxuan866@163.com

2nd Keying Chen
School of Computer and Computing Science
Zhejiang University City College
Hangzhou, China
boencky111111@163.com

3rd Lin Sun*
School of Computer and Computing Science
Zhejiang University City College
Hangzhou, China
sunl@zucc.edu.cn

4th Chenlu Hu
School of Computer and Computing Science
Zhejiang University City College
Hangzhou, China
646223066@qq.com

***Abstract*—Text readability is useful to measure the comprehensibility of a piece of writing and plays an important role in the field of education. The classic evaluation methods are linear functions with two or three parameters. However, it needs laborious human tests to find the appropriate parameters. This paper presents a recurrent neural network to quickly build a text readability evaluation model. The model consists of bi-directional gated recurrent unit (bi-GRU) and attention layer. It can predict the readability level of the sentences or paragraphs. The weight of the attention layer can finely locate the distribution of difficulty in the comprehension of the sentences. We train the model on rough leveled reading materials without tremendous reading comprehension tests. Our experiments are performed on various text materials. Compared with the popular readability formulas, the results show good performance on readability measurement and visualization.**

I. Introduction

Text readability assessment is important for many institutional and individual users in a wide range of fields. In the education field, it can score the difficulty of textbooks [1] and match students to books or articles that they could read at an acceptable rate. In government documents, financial reports, and technical manuals, the comprehensibility of written plain texts needs to be measured and improved in a clear, concise, and exactly readable style. The quantitative measure is a popular way to assess text readability. Most of the readability evaluation methods are formulas which consist of the parameters, such as average sentence length, average word length, or the percentage of complex words. The formula style has some limitations:

- Traditional readability formulas use statistical correlation and regression analysis to study a language. The factors affecting readability are limited to word length, sentence length, and difficult words. The formulas can only give a global assessment on paragraph.
- Hand-crafted parameters should be fit by multiple regression. It requires tremendous tests of human reading comprehension [2].
- A readability formula is always limited to one language. For instance, Flesch Reading Eases and ARI formulas are designed for English and inaccurate for other languages.
- Once the parameters are calculated, they are fixed and used for a long time. It is not flexible.

In this paper, we explore the task of text readability classification using attention-based bi-GRU model. The contributions of this paper can be summarized as follows:

- We propose an attention-based bi-GRU method that has the ability to quickly establish a text readability assessment model for one language and can be adjusted flexibly according to different scenarios. The quantitative metric is calculated by a neural network instead of hand-crafted formulas. The model can be trained on rough leveled reading materials, such as the graded textbooks. Our model can not only assess the documents but also a single sentence.
- We perform a set of readability tests on dialogues, reviews, news, and English test papers. The results show that the readability of our model on text materials and correlation between hand-crafted formulas. We also visualize the readability of words or phrases by attention weights and measure the sentences which are ordered in wrong manner.

II. Related Work

The hand-crafted formula is well-known for measuring text readability [3]. Features are selected from sentence length, difficult words, syllables, etc. Flesch Reading Ease formula [4] was firstly developed by Rudolph Flesch, a supporter of the plain English movement. Later, Kincaid [2] revised the Flesch Reading Ease and developed Flesch-Kincaid Grade Level for the U.S. Navy in 1975. In their experiments, 569 subjects were

* Lin Sun is the corresponding author.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

tested for their comprehension of eighteen passages. Automated Readability Index (ARI) [5] considered the parameters of average word length and average sentence length. Gunning Fog Index [6] consisted of the percentage of hard words and average sentence length. Läsbarhetsindex (LIX) was proposed by Carl-Hugo Björnsson [7]. It was developed for the Swedish language but existing research indicated that the formula also performs well on most of the non-English languages.

In recent years, some researchers tried to build readability assessment models using machine learning techniques. Si and Callan [8] presented unigram language models to estimate the text associated with each readability level. The training data were 91 web pages downloaded from the Internet and divided into three readability levels: Kindergarten-Grade2, Grade3-Grade5, and Grade6-Grade8. Larsson [9] presented a Support Vector Machines (SVM) classifier for Swedish readability levels. The corpus consisted of three different levels, i.e. morning paper texts as difficult-level, high school student texts as medium-level, easy newspaper texts as easy-level. The extracted features were sentence-length, syntactic depth, etc. Qumsiyeh and Ng [10] presented a readability assessment tool using the multi-class SVM model. They constructed training dataset from standardized English Language tests of 14 grade levels from kindergarten to college. Sung et al [11] used SVM with 24 linguistic features to predict the Chinese text readability. Liu et al [12] employed CNN and LSTM to evaluate Chinese teaching material with levels of elementary, middle or advanced. Iram et al [13] evaluated the text readability using the fuzzy logic.

III. Methodology

Given a sentence $X = (w_1, ..., w_T)$, where T is the length of the sentence and w_i denotes a word token that is represented as an embedding vector. The learning problem is to map the input sentence X to a target $y \in RL$, where $RL = \{1, 2, ..., l\}$ is a set of readability level labels. We introduce a bi-GRU with attention mechanism shown in Figure 1 for text readability classification.

Gated Recurrent Unit (GRU) [14] is a recurrent neural network architecture (see Figure 2) which is simpler compared to LSTM [15]. A GRU unit is composed of a candidate state $\tilde{h_t}$ which depends on current input word x and the hidden layer state vector at the previous moment h_{t-1}, a reset gate $\mathbf{r} \in \mathcal{R}^h$ which is responsible for the decision of how important h_{t-1} is to the $\tilde{h_t}$, an update gate $\mathbf{z} \in \mathcal{R}^h$ which determines whether to ignore the current input x and a hidden state $\mathbf{h} \in \mathcal{R}^h$ which is determined by previous hidden state, candidate state and the update gate $\mathbf{z}$. The forward pass of a GRU unit can be described by the following equations:

$$\begin{aligned}
z_t &= \sigma(W_z x_t + U_z h_{t-1} + b_z) \\
r_t &= \sigma(W_r x_t + U_r h_{t-1} + b_r) \\
\tilde{h_t} &= \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \\
h_t &= z_t \odot \tilde{h_t} + (1 - z_t) \odot h_{t-1}
\end{aligned} \tag{1}$$

, where $W \in \mathcal{R}^{h \times d}$ and $U \in \mathcal{R}^{h \times d}$ are weight matrices, $b \in \mathcal{R}^h$ is bias vector , $x_t \in \mathcal{R}^d$ is an input word embedding, h_t is a hidden-state vector, $\sigma(\cdot)$ is a sigmoid function, and operator $\odot$ denotes an element-wise product.

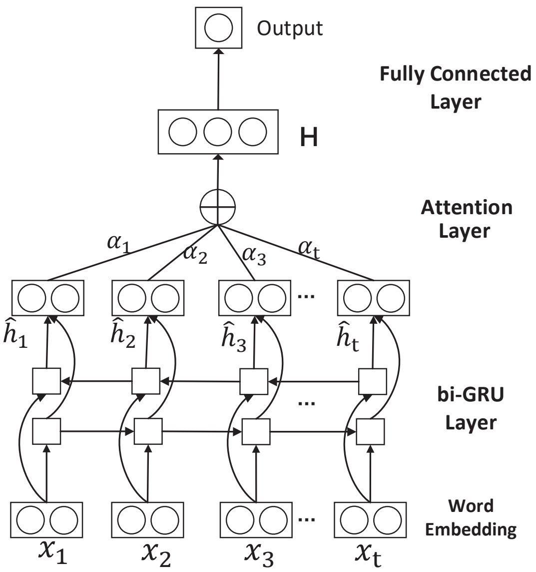

Fig. 1. Illustration of Attention-based bi-GRU Model

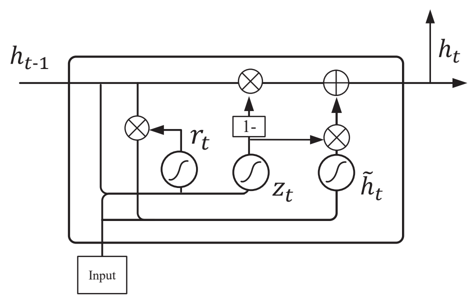

Fig. 2. Gated Recurrent Unit

Bidirectional recurrent network [16] computes the forward hidden sequence $\overrightarrow{h_t}$ from $t = 1$ to T and the backward hidden sequence $\overleftarrow{h_t}$ from $t = T$ to 1. Then the output $\hat{h}_t$ concatenates $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ together which can be used to provide more contextual information.

$$\hat{h}_t = <\overrightarrow{h_t}, \overleftarrow{h_t}> \tag{2}$$

The basic idea of attention mechanism is to match the readability level with the specific words or phrases in the sentence. We compute the vector H, a weighted sum of all the previous time steps, instead of using the last hidden state vector $\hat{h}_T$.

$$H = \sum_{t=1}^{T} \alpha_t \hat{h}_t. \tag{3}$$

where α_t is attention weights which computed in a feed-forward neural network,

$$\alpha_t = \frac{exp(w^T tanh(\hat{h}_t))}{\sum_{i=1}^{T} exp(w^T tanh(\hat{h}_i)}, \tag{4}$$

where $w \in \mathcal{R}^{d\times 1}$ a parameter vector. H is fed to a full connected layer and then a softmax layer that outputs the probability distribution $p = (p_1, ..., p_l)$ of the readability levels. Figure 1 shows the illustration of our model.

For a sentence, the readability level (RL_s) is defined as the label whose probability is the maximum in p, For a passage or document, the readability level is defined as the average of RL_s of all sentences,

$$RL_d = \frac{1}{N}\sum_{i=1}^{N} RL_{s_i}, \tag{5}$$

where N is the number of the sentences.

IV. Experiments

A. Readability Formulas

In experiment, we also demonstrate the results of classical readability formulas for comparison. The brief information of four popular readability formulas are listed as follow:

- Flesch Reading Ease formula: Flesch Reading Ease formula was introduced in 1948 by Rudolph Flesch and one of the most popular formula for measuring the text readability. The score of readability can be computed by the following equation:

$$\begin{aligned} &Flesh \\ &= 206.835 - 0.846WL - 1.015SL, \end{aligned} \tag{6}$$

  where WL is the average number of syllables per 100 words, and SL is the average number of words per sentence.
- Flesch-Kincaid Grade Level:
  Flesch-Kincaid Grade Level was developed in 1975 by J. Peter Kincaid [2]. This formula was revised from Flesch Reading Ease formula. The score is equivalent to the U.S. grade level and easy to choose books for children. The grade of readability can be computed by following equation:

$$\begin{aligned} &Grade\ Level \\ &= 0.39SL + 11.8WL - 15.59, \end{aligned} \tag{7}$$

  where WL is the average number of syllables per 100 words, and SL is the average number of words per sentence. The grade computed by the formula represents the US grade level of education.
- Gunning Fog Index: Gunning Fog Index formula was proposed by Robert Gunning in 1952, also aims to approximate the grade level of education a person needs to understand the text. The grade of readability can be obtained by the following formula:

$$Fog = (SL + HD) * 0.4, \tag{8}$$

  where SL is the average sentence length in a sample of 100 words and HD is the number of words with syllables greater than or equal to 3 in the sample of 100 words.
- Automated Readability Index (ARI): The Automated Readability Index was devised to provide an easy and automated method to evaluate the readability of textual material [5]. This formula also outputs the grade level of a text which is equivalent to the formal education level. The Automated Readability Index is given by the following equation:

$$ARI = 4.17WL + 0.5SL - 21.43 \tag{9}$$

  where WL is the average number of letters in each word and SL is average sentence length.

B. Datasets

We perform the experiments on Lexile books, reviews, dialogues, news, and English test papers. The brief introduction of the datasets are listed as follows:

- **Lexile Book dataset**: The Lexile Framework is an educational tool to match readers with books, articles and other leveled reading resources [1]. In the United States, the Common Core State Standards recommend the Lexile Framework as a quantitative measure for selecting books for students [17]. The Lexile scale is from BR300 to 2000L [18], matching readers from beginning to college and career level. We map the Lexile Scale (LS) into five readability levels (RL), as shown in Table I. To reduce the possibility of confusion of neighboring Lexile scale, we separate the adjacent RLs with an interval of 200L. We select books of the different levels with corresponding Lexile range.

TABLE I
Readability levels and Lexile ranges in Lexile book dataset.

RL	LS	Age
1-very easy	150-350L	5-6 years old
2-easy	550-750L	7-9 years old
3-medium	950-1150L	10-14 years old
4-hard	1350-1550L	15-19 years old
5-very hard	1750-1950L	20+ years old

- **Review, dialog and news**: We use IMDB Movie reviews[2] which is a dataset contains movie reviews along with their associated binary sentiment polarity labels, Cornell Movie Dialog which is a corpus with data-rich collection of dialogues extracted from the raw movie scripts[3], and AG's corpus of news articles which is a collection of more than 1 million news articles which gathered from more than 2000 news sources [4] to evaluate our model.
- **TOEFL English test papers**: TOEFL is a standard English language proficiency test which is highly recognized in the world. More than 10000 universities and other

[1]https://lexile.com/
[2]https://www.imdb.com/interfaces/
[3]http://www.cs.cornell.edu/∼cristian/Cornell_Movie-Dialogs_Corpus.html
[4]https://www.di.unipi.it/∼gulli/AG_corpus_of_news_articles.html

institutions in over 150 countries accept TOEFL scores. In this paper, we collect 20 TOEFL test papers and test the reading comprehension part.

- **China's National Postgraduate Entrance Examination (NPEE)**: China's National Postgraduate Entrance Examination is a national unified entrance exam for universities and scientific research institutions to recruit postgraduates. The purpose is to scientifically, impartially and effectively test the ability of using English language.
- **Eighteen passages**: Kincaid [2] selected eighteen passages from the U.S. navy training manual for developing Flesch-Kincaid Grade Level. These passages are presented as cloze and sent to testers for testing and use their scores as a measure of the difficulty of these passages.

C. Settings

We use GloVe [19] 300-dimensional word vectors as the input of our model. The size of mini-batches is 64 sentences. The size of the hidden layer is 300. We implement the attention-based bi-GRU network using PyTorch and run Adam [20] optimizer for 60 training epochs with the learning rate of 5e-4. The dropout rate is 0.75.

We use the Lexile Book dataset for model training. The books are downloaded in PDF format and converted to plain text files. We collect the complete sentences from the plain text file. All sentences in the book are marked as the label that matches the book's RL. We count the distribution of the sentence length in a book and filter the sentences if their length is out of [Q1, Q3] range, where Q1 is the lower quartile and Q3 is the upper quartile. Table II shows the number of books (#Book) , sentences (#Sentence), and average length of sentences ($\overline{L}$) in each level.

TABLE II
THE STATISTICS OF THE LEXILE BOOK DATASET.

RL	#Book	#Sentence	$\overline{L}$
1	150	1239	5.5
2	10	8011	9.6
3	10	5823	18.2
4	10	7262	24.2
5	10	4463	31.5

D. Classification Performance Evaluation

Precision, recall and F1 score are used for the evaluation metrics in the experiments. First, we split the sentences of the Lexile Book dataset into train/dev/test with the ratio of 8:1:1 to test the performance of sentence-level classification. Table III shows the comparison of CNN, RNN, GRU, bi-LSTM [16], bi-GRU, and bi-GRU with attention. Table IV shows the precision (P), recall (R) and F1 score for each reading level.

Second, we divide the books of each level into five folds and perform cross-validation. Table V shows the confusion matrix of book-level classification. Each row of the matrix corresponds to an actual class, and each column of the matrix corresponds to a predicted class.

E. Evaluation on Other Text Materials

We apply the model on reviews, dialogues and news articles. Figure 3 shows the readability testing results on different text materials. Dialogues are always simple and easy to understand, so the number of level 2 is the most. Movie reviews are more difficult than dialogues, and the part of level 3 is the largest. The readability of news articles matches 15+ years old, and the percentages of level 4 and 5 account for approximately 98%.

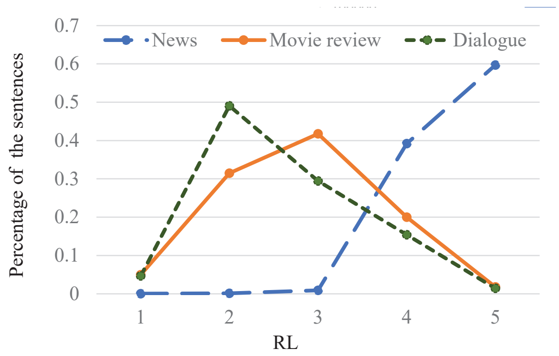

Fig. 3. Results on different text materials of bi-GRU+Attention.

We also calculate the correlation between our model and several classical formulas such as Flesch-Kincaid Grade Level, Flesch Reading Ease formula, Automated Readability Index (ARI) and Gunning Fog Index on TOEFL test papers, Chinese Postgraduate English test and 18 passages. Table VI shows the statistics of correlation between our model and several classical formulas.

F. Visualization of Attention Weights

We get the attention weights from the model and visualize them in Table VIII. The deeper color corresponds to a larger attention weight. We can find that the attention mechanism in our model tends to give larger weight to the words with more alphabets and syllabus which can be more difficult to understand. This phenomenon is consistent with readability formulas. For example, in the third sentence of AG news, the attention mechanism gives a relatively large weight to the word "dramatically", "deplete", "California", "heat-related" and "jeopardizing".

Visualization of text readability can be used in the following ways, but not limited to:

- For reading books, the visualization of text difficulty can help readers locate reading difficulties in the sentences, and understand the text better. Also, it provides a more intuitive way for readers to learn difficult words and syntax.
- For language teachers, they can pay more attention on higher weighted words in a sentence when they design questions. Teachers can focus on explaining the difficult words or grammar structures to students.
- The visualization of text difficulty is also applicable to text simplification. The words or sentences which have

TABLE III

COMPARISON WITH THE OTHER MODELS IN SENTENCE-LEVEL CLASSIFICATION.

	dev			test		
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
CNN	79.7	82.0	80.3	80.0	81.9	80.3
RNN	81.0	81.0	80.8	79.9	79.5	79.3
GRU	91.3	91.8	91.5	89.4	90.2	89.7
bi-LSTM	89.6	90.8	90.0	87.4	89.4	88.2
bi-GRU	91.6	92.2	91.9	90.6	91.2	90.9
bi-GRU + Attention	**92.1**	**92.8**	**92.4**	**90.7**	**91.6**	**91.1**

TABLE IV

CATEGORY PERFORMANCE ON DEV AND TEST DATASETS IN SENTENCE-LEVEL CLASSIFICATION.

RL	dev			test		
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
1	96.7	90.2	93.3	98.0	89.2	93.4
2	96.0	94.2	95.1	94.4	94.2	94.3
3	90.0	86.5	87.2	86.9	84.2	85.6
4	90.2	92.2	91.2	88.9	89.7	89.3
5	93.1	97.2	95.1	89.9	96.2	92.9
Overall	92.1	92.8	92.4	90.7	91.6	91.1

TABLE V

CONFUSION MATRIX OF BOOK-LEVEL CLASSIFICATION.

RL	1	2	3	4	5
1	145	5	0	0	0
2	0	10	0	0	0
3	0	1	8	1	0
4	0	0	0	10	0
5	0	0	0	0	10

TABLE VI

STATISTICS OF CORRELATION BETWEEN OUR MODEL AND CLASSICAL FORMULAS.

	TOEFL	NPEE	18passages
Flesch-Kincaid	0.794	0.826	0.839
Flesch	-0.759	-0.732	-0.766
ARI	0.817	0.804	0.756
Fog	0.725	0.765	0.683

large weight can be replaced with synonyms that are easy to understand. We can also improve the reports in a clear and readable style for ordinary citizens.

G. Results of Sentences in Disorder

We randomly select 500 sentences from level 1 to 4 in books, dialogue, movie review, and news datasets respectively. We shuffle the sentences in random order to test the readability of disordered sentences. The scores of readability formulas are unchanged because the parameters of average length, average word length, syllables and difficult words are independent of order. Table VII shows the comparison of average level $\overline{RL}$ on ordered and disordered sentences. The results of $\Delta_{\overline{RL}}$ show $\overline{RL}$ increases on disordered sentences, where level 1 is the largest. Table IX shows the visualization examples between ordered and disordered sentences. We can find that the grammatical errors in the disordered sentences can be visualized by attention weights. The awkward phrases, such as ,“nervous me”, “teeth Corrective”, “of a More”, “and very story”, “to many think”, “it ends whole”, “about to you”, “had I this”, “her of see”, are shown in Table IX.

TABLE VII

COMPARISON OF AVERAGE LEVEL ON ORDERED AND DISORDERED SENTENCES.

	RL=1	RL=2	RL=3	RL=4
$\overline{RL}_{ordered}$	1	2	3	4
$\overline{RL}_{disordered}$	1.70	2.31	3.23	4.14
$\Delta_{\overline{RL}}$ (%)	70.0	15.5	7.7	3.5

V. CONCLUSION

This paper has presented an attention-based bi-GRU method for readability assessment. Instead of consuming a lot of man-power and material resources to build a hand-crafted formula, our method can quickly build a text readability measurement model from rough leveled reading materials. The proposed model can evaluate all kinds of texts. Visualization of attention weights can be used to locate where the difficult part is in a sentence. It is helpful to key highlight, article modification and text readability improvement in many language applications.

ACKNOWLEDGMENT

This work is supported by National Innovation and Entrepreneurship Training Program for College Students (No.201913021002)

REFERENCES

[1] N. G. Association *et al.*, “Common core state standards,” *Washington, DC*, 2010.

[2] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom, “Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel,” 1975.

[3] G. R. Klare *et al.*, “Measurement of readability,” 1963.

[4] R. Flesch, “A new readability yardstick.” *Journal of applied psychology*, vol. 32, no. 3, p. 221, 1948.
[5] R. Senter and E. A. Smith, “Automated readability index,” CINCINNATI UNIV OH, Tech. Rep., 1967.
[6] R. Gunning, “Plain language at work newsletter,” 2004.
[7] J. Falkenjack, K. H. Mühlenbock, and A. Jönsson, “Features indicating readability in swedish text,” in *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013)*, 2013, pp. 27–40.
[8] L. Si and J. Callan, “A statistical model for scientific readability,” in *CIKM*, vol. 1, 2001, pp. 574–576.
[9] P. Larsson, “Classification into readability levels: implementation and evaluation,” 2006.
[10] R. Qumsiyeh and Y.-K. Ng, “Readaid: a robust and fully-automated readability assessment tool,” in *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence*. IEEE, 2011, pp. 539–546.
[11] Y.-T. Sung, T.-H. Chang, W.-C. Lin, K.-S. Hsieh, and K.-E. Chang, “Crie: An automated analyzer for chinese texts,” *Behavior research methods*, vol. 48, no. 4, pp. 1238–1251, 2016.
[12] H. Liu, S. Li, J. Zhao, Z. Bao, and X. Bai, “Chinese teaching material readability assessment with contextual information,” in *2017 International Conference on Asian Language Processing (IALP)*. IEEE, 2017, pp. 66–69.
[13] N. Iram, S. Zafar, and R. Zahra, “Web content readability evaluation using fuzzy logic,” in *2018 International Conference on Advancements in Computational Sciences (ICACS)*. IEEE, 2018, pp. 1–8.
[14] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using rnn encoder-decoder for statistical machine translation,” *arXiv preprint arXiv:1406.1078*, 2014.
[15] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[16] A. Graves, A.-r. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2013, pp. 6645–6649.
[17] E. H. Hiebert, “The common core state standards and text complexity,” *Teacher Librarian*, vol. 39, no. 5, pp. 13–19, 2012.
[18] A. J. Stenner, “Measuring reading comprehension with the lexile framework.” 1996.
[19] J. Pennington, R. Socher, and C. Manning, “Glove: Global vectors for word representation,” in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
[20] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

TABLE VIII

EXAMPLES OF VISUALIZING ATTENTION WEIGHTS.

[Movie_dialogue]

Because I want to get married .

Your father is a smart guy .

There might still be one standing .

Where did he get all that information ?

[IMDb]

It allows a person to enter both realms of the conflict simultaneously .

Although there are some violent scenes , it was the Nazi occupation of Hungary .

I soon got lost with all the governmental politics .

It makes the ideas of exploration exciting to us again .

[AG NEWS]

Iran's Arash Miresmaili withdraws from the Olympics amid confusion over his reasons .

Americans who served in the military will be well represented at the upcoming republican convention , more so than at last month's Democratic convention or in the U.S. population overall , according to the GOP .

Global warming could cause dramatically hotter summers and a depleted snow pack in California , leading to a sharp increase in heat-related deaths and jeopardizing the water supply , according to a study released monday .

In a bid to keep protesters from disrupting the Republican National Convention , the Bloomberg administration is offering " peaceful political activists " discounts .

[TOEFL paper]

The former is termed reintroduction and the latter reinforcement .

An example of apparent translocation success involves the threatened seychelles warbler .

The fundamental premise of modern architecture was that the appearance of the building should exhibit the nature of its materials and forms of physical support .

Both of these , variability in responses and functional redundancy , could be thought of as insurance against disturbances .

TABLE IX
VISUALIZATION EXAMPLES BETWEEN ORDERED AND DISORDERED SENTENCES.

simple very That .	⟹	That very simple .
Makes me nervous .	⟹	Makes nervous me .
Corrective teeth surgery .	⟹	teeth Corrective surgery .
They are not afraid !	⟹	not afraid ! are They
More of a gentleman ...	⟹	gentleman ... of a More
The story is very sensitive and touching .	⟹	sensitive and very story is The touching .
I think many would like to see him commander-in-chief .	⟹	see to many think commander-in-chief . I him like would
If you followed the whole thing , you know how it ends .	⟹	followed know If you how . it ends whole , thing the you
I need to know about your upbringing .	⟹	about to your need upbringing . know I
I had a personal interest in this movie .	⟹	personal had I this . movie in a interest
I recommend to all of you to see all of her movies .	⟹	recommend to her of see . all I you of all to movies