Multilingual Age of Exposure

Robert-Mihai Botarleanu[1], Mihai Dascalu[1,2](✉), Micah Watanabe[3], Danielle S. McNamara[3], and Scott Andrew Crossley[4]

[1] University Politehnica of Bucharest, 313 Splaiul Independentei, 060042 Bucharest, Romania
robert.botarleanu@stud.acs.upb.ro, mihai.dascalu@cs.pub.ro
[2] Academy of Romanian Scientists, Str. Ilfov, Nr. 3, 050044 Bucharest, Romania
[3] Department of Psychology, Arizona State University, PO Box 871104, Tempe, AZ 85287, USA
{micah.watanabe,dsmcnama}@asu.edu
[4] Department of Applied Linguistics/ESL, Georgia State University, Atlanta, GA 30303, USA
scrossley@gsu.edu

Abstract. The ability to objectively quantify the complexity of a text can be a useful indicator of how likely learners of a given level will comprehend it. Before creating more complex models of assessing text difficulty, the basic building block of a text consists of words and, inherently, its overall difficulty is greatly influenced by the complexity of underlying words. One approach is to measure a word's Age of Acquisition (AoA), an estimate of the average age at which a speaker of a language understands the semantics of a specific word. Age of Exposure (AoE) statistically models the process of word learning, and in turn an estimate of a given word's AoA. In this paper, we expand on the model proposed by AoE by training regression models that learn and generalize AoA word lists across multiple languages including English, German, French, and Spanish. Our approach allows for the estimation of AoA scores for words that are not found in the original lists, up to the majority of the target language's vocabulary. Our method can be uniformly applied across multiple languages though the usage of parallel corpora and helps bridge the gap in the size of AoA word lists available for non-English languages. This effort is particularly important for efforts toward extending AI to languages with fewer resources and benchmarked corpora.

Keywords: Natural language processing · Age of acquisition · Age of exposure · Multilingual

1 Introduction

The quantification of textual complexity is a crucial step toward better understanding the relations between text comprehension, the reader, and the nature of the text. Words are the fundamental building blocks of texts, and thus analysis of word complexity in a text can provide insight into the difficulties that readers might have in understanding certain documents. However, many of the tools used to estimate word complexity are created specifically for the English language. While simple measures such as number of characters in syllables can be easily identified regardless of the language, other measures

I. Roll et al. (Eds.): AIED 2021, LNAI 12748, pp. 77–87, 2021.
https://doi.org/10.1007/978-3-030-78292-4_7

of word complexity can only be measured by examining the relations between words and how words are used within the context of the language. Creating new tools to measure word complexity in multiple languages can aid in the crafting of better online instruction materials and techniques as well as interventions for a broader range of students. This is an important objective, particularly for under-resourced countries and languages.

Numerous approaches to quantifying word complexity have been proposed. These range from simple surface-level measurements, such as the number of syllables or characters, to measurements such as a word's frequency in a corpus or the number of synonyms for a given word. Previous studies have demonstrated detrimental impacts of complex words on reading comprehension. People tend to spend more time focusing on ambiguous or infrequent terms [1], which directly impacts reading speed. Certain words are more easily learned by L2 speakers [2] and various measures of word complexity are employed in evaluating of the complexity of phrases and texts [3].

"Age of Acquisition" (AoA) is an indicator of a word's complexity from the perspective of language learning. AoA is an estimate of the average age an average language learner acquired a given word. Word lists of AoA scores are typically collected using adults' estimates of when they learned the word [4]. The production of AoA lists is costly, time-consuming, and reflects adults' memories of word learning, and not the actual process of word learning. Like AoA, Age of Exposure (AoE) [5] is also an estimate of the average age that an average language learner acquires a given word. However, AoE scores are derived from a machine learning model that is trained on increasingly large corpora of texts, which simulates the process of learning a language to provide an automated measure of word complexity.

Age of Exposure is an extension of the Word Maturity model created by Landauer et al. [6]. In the Word Maturity model, Latent Semantic Analysis [7] was used to generate word vectors on increasingly larger, cumulative, corpora of texts. By performing Procrustes rotation between the vector spaces given by the LSA word vectors, one is then able to measure the cosine distance between the representation of a word at a given step in the trajectory and the final, "adult", representation. In AoE, Latent Dirichlet Allocation (LDA) [8] is used instead of LSA [6]; LDA affords better estimates of polysemy, with lower computational costs. In addition, AoE also introduces additional statistical features extracted from the learning trajectories.

While AoA and AoE scores are related to measures of reading comprehension and writing skill, the majority of published lists of AoA scores are for English words, and previous iterations of the AoE model have only been trained on English text corpora [6]. Thus, the aim of this study is to expand on the AoE models by providing a method of directly estimating the AoE scores from the learning trajectories, generated using unsupervised language models of words in English, German, French and Spanish AoA word lists. We investigate the similarities between these word lists and show that our method can generalize accurate AoA estimations for different languages, allowing for the creation of approximate AoA word lists on the entirety of a language's (known) vocabulary. The differences between the distributions of AoA scores in different languages are expected to impact the performance of modeled learning trajectories; however, our method shows that simulated word learning trajectories generated by applying unsupervised language models on multi-lingual corpora can capture similarities as well as differences between the word learning processes in those languages. We thus aim to answer the following research questions: a) Are AoA word lists in different languages sufficiently similar to afford using the same statistical modeling technique? and b) Can we estimate, within reasonable error, the AoA scores for words in a language automatically and how do these models relate in terms of the features used?

2 Method

2.1 Corpora

To perform the iterative model training necessary to estimate learning trajectories, we required a corpus that was both sufficiently large and also similar between languages. To this end, selected the "ParaCrawl" [9] dataset which provides documents that are aligned between various languages (i.e., they are equivalent through translation), extracted from a large number of webpages. Of these, we used three aligned corpora, English-German (en-de), English-French (en-fr), and English–Spanish (en-es).

In order for the trained models to estimate learning trajectories for various languages, the texts in the corpora must present sufficient variety in terms of complexity. One means of evaluating text complexity, independent of the AoA, is to use an automatic readability formula such as the Flesch Reading Ease [10], which uses simple surface-statistics of the structure of an English text to estimate its difficulty. By plotting the distributions of the Flesch Reading Ease scores across the three corpora we selected, we observed a uniform distribution of readability on the English documents in the dataset (see Fig. 1). Some of the documents exceed the 0–100 range that Flesch defined in the original paper; however, this possibly resulted from the documents being automatically crawled from webpages resulting in syntax errors (i.e., sentences not terminated properly or whitespaces between words missing). Nevertheless, the three corpora present relatively uniform distributions with the majority of texts being located in the 50–75 range. Given that the Flesch Reading Ease formula was constructed for English, applying it directly to directly to the other three languages is not uniformly reliable. We elected, instead, to assume that the aligned texts had readability levels similar to their English counterparts.

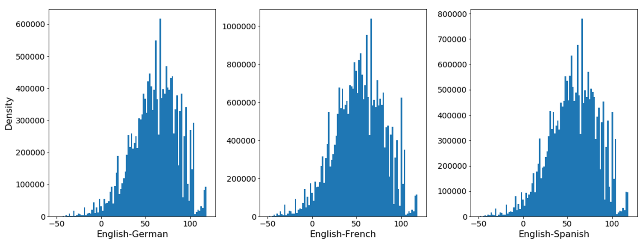

Fig. 1 Flesch Reading Ease distributions for the English dataset

In the AoE paradigm, language models are trained on increasingly larger subsections of a corpus. This is intended to simulate the way in which humans are exposed to more texts (or discourse) as they learn to speak, read, and write. In our experiments, we elected to split each of the three corpora into 20 different stages. Each stage included all of the texts in the previous ones, with the final model being trained on the entirety of the corpus of a language. In Fig. 2, the progression of the size of the three corpora as language acquisition is simulated has been plotted. All three are large, with the English-German corpus having 813,223 documents in the first stage and 16,264,448 documents in the final stage; English–Spanish 1,099,364 in the first and 21,987,267 documents in the final stage; and English-French 1,568,709 in the first and 31,374,161 documents in the final stage. Here, a “document”, means a pair of aligned texts in two languages. We also considered two different orders for the documents: an arbitrary ordering and one based on Flesch Reading Ease, with the most readable texts being *seen* first, with the least readable ones being left for the latter stages.

Our model simulates the manner in which humans are exposed to language, starting by reading simpler texts and increasing difficulty as their language mastery improves; nevertheless, this approach does not consider other channels for language learning (e.g., dialogue with other people, video and audio entertainment, writing). In the context of the Word Maturity and AoE models, word acquisition is modeled as the growth of the simulated vocabulary when the model is presented with increasingly more text. The simulated learning trajectories take a simplified view of human language learning because they do not take into account individual differences (e.g., personal interests, different educational systems) and are intended to model the average level of language exposure a language speaker might encounter solely by reading texts.

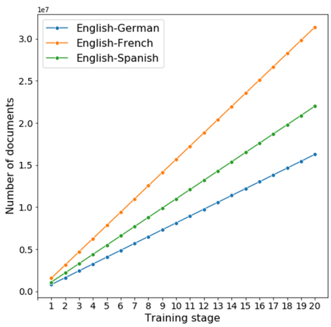

Fig. 2 Number of documents in each of the three corpora

AoE scores are correlated with AoA scores because they are assumed to reflect the language learning process. Thus, in order to estimate AoE word scores, we trained statistical regression models that required training and evaluation data – namely AoA word lists. We selected an AoA word list per language: English [4], French ([11], Spanish [12], and German ([13]. The three word lists varied in size (English: 30,121; French: 1,493; Spanish: 7,039: German: 3,200); however, our approach assumed that the model follows the same learning process for all languages (which is likely incorrect but necessary for the current analysis). To assess the viability of this assumption, we performed automatic word-to-word translations and measured the correlations between the English word list and the others. While not all the words could be automatically matched, the majority were, and we were able to confirm their correlation using Spearman Rank Correlations: English-German $r = 0.681$, English-French $r = 0.594$ and English–Spanish $r = 0.682$.

The distributions for the four AoA lists are provided in Fig. 3. The English word list scores are the closest to a normal distribution, while the Spanish scores appear almost bimodal. The ranges of the distributions also differ, with some English word scores exceeding 20, while the maximum Spanish scores are 11, and the German and French scores are approximately 15. In addition to their relative sizes, these differences in the distributions can impact attempts to train regression models to predict AoA scores.

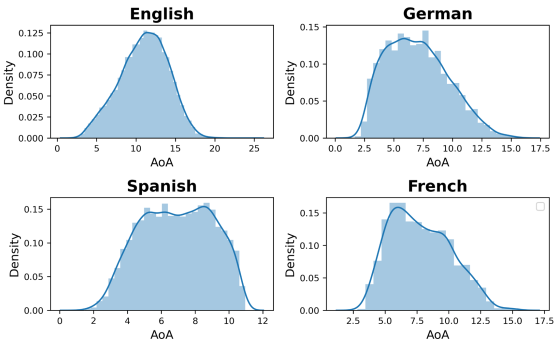

Fig. 3 Distribution plots for the four AoA word lists

2.2 Modeling Learning Trajectories

To model learning trajectories, we trained Word2Vec [14] language models utilizing the cumulatively increasing corpora, as outlined previously in Sect. 2.1. Of the two variants of Word2Vec, we chose to use the skip-gram architecture wherein the Word2vec model is used to predict context words for a given target term. Our choice of using Word2Vec instead of LDA as used in the first version of AoE was motivated by the inherent geometrical properties of the word vectors it produces. Word2Vec maps words into a multi-dimensional vector space wherein arithmetic operations between the vectors are used to represent semantic and syntactic relationships between words. As such, this method was a more a natural fit in the incremental training algorithm used to model learning trajectories. Specifically, the Word2Vec model could then be evaluated as it evolved (i.e., as it was exposed to more texts) by comparing intermediate vector spaces to the mature one.

Specifically, we utilized word embedding vectors of size 300, with a context window of 5 and trained each model for 50 epochs. Because the models were trained on incrementally increasing portions of each corpus, the final, “mature”, model was assumed to contain the most accurate word embeddings. With this in mind, the intermediate models offer snapshots into what Word2Vec was able to model at each “learning” step. Measuring the discrepancy between an intermediate word representation and its final, mature one can be done using cosine similarity. We trained our models in stages. Hence, there were 19 intermediate model similarities to the mature representation, which formed the learning trajectories. Prior to measuring the cosine similarity, we performed a Procrustes alignment of the vector space represented by the intermediate word embeddings to the mature vector space. An illustration of these learning trajectories is provided in Fig. 4, which shows the cosine similarities of the intermediate models to the mature one for the English texts of the *English to German* corpus. Each of the learning trajectories is colored on a gradient from blue-to-red based on word frequencies in the corpus. These evolutions are consistent with the ones from the first model of AoE [5], but are more fine-grained with smoother evolutions.

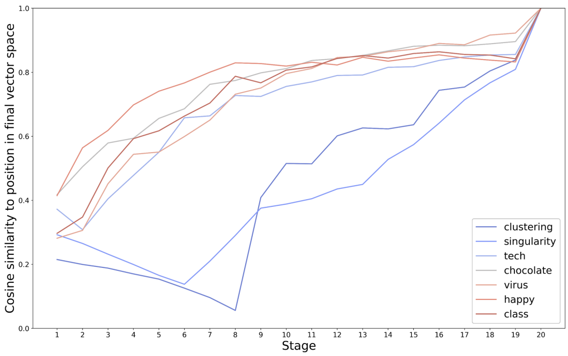

Fig. 4 Example of learning trajectories for the English to German corpus

Via these illustrations, we observed that some words, such as “tech” and “singularity”, have noticeably steeper learning trajectories. Others, such as “happy” and “chocolate”, have relatively good cosine similarities from the earliest stages, suggesting that the intermediate model’s representations of those terms are closer to the mature model representation. In terms of AoA, we can consider “happy” as having a low age of acquisition, with “clustering” being acquired later. In comparison to the AoE trajectories, the ones we generated showed a monotonic increase, which is expected from the fact that the Word2Vec model trained at a certain stage uses all the documents on which the previous intermediate stages were trained, in addition to its own portion.

Similarly, we explore the learning trajectories for words in different languages (see Fig. 5). While some common words, namely “dog” and “red”, appear to have similar trajectories in the four languages, we can observe differences. Namely, in Spanish, the word for “class” (i.e., “clase”) seems to be learned far more quickly than in other languages. Consequentially, the AoA score for the Spanish word “clase” is somewhat lower (3.84) than its translations in other languages (English “class”: 4.95, French “classe”: 4.92, German: no equivalent in word list). Similarly, the Spanish AoA score for “virus” is 8.16, while the English word list has it at 9.5 and the German word list at 9.65. The process of learning words differs from language to language, especially in the case of specialized terms. These are a few randomly chosen examples; however, the presence of differences in the trajectories modeled by AoE that are also reflected in AoA word lists suggests that our trajectories resemble aspects of human word acquisition and capture, at least partially, differences between word learning in different languages.

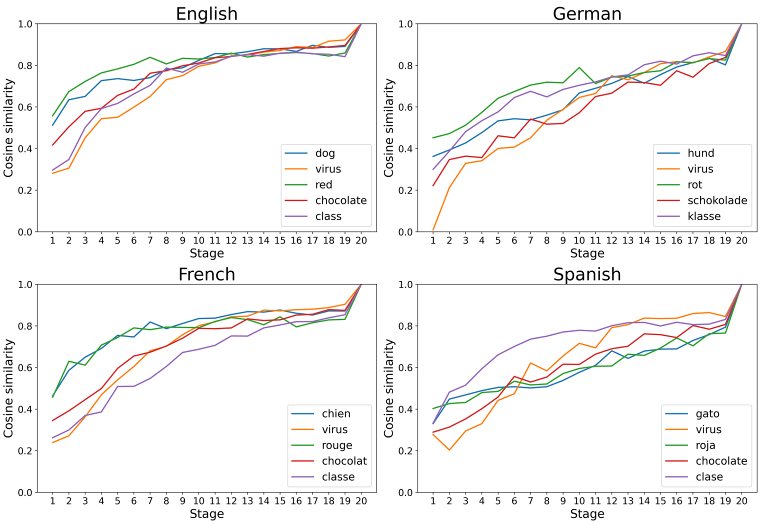

Fig. 5 Learning trajectories for different languages

From these learning trajectories, we extracted several features that described both the relations between a word and the rest of the vocabulary and the learning process for that word. These features can be split into two groups:

- **Mature Model Features:** the cosine similarities between the word embeddings of a term and other words in the vocabulary. These include the 1st, 2nd and 3rd highest cosine similarities to words in the vocabulary and their average, as well as the number of words that have a cosine similarity of at least 0.3 to the term and their average cosine similarity.
- **Learning Trajectory Features:** the 19 intermediate model cosine similarities, their average and its 1-complement, the index of the first intermediate model that achieves a cosine similarity above a certain threshold (from 0.3 to 0.7 in 0.05 increments) and the slope of the best fitting line on the plots shown in Fig. 4 and its inverse value.

Through these features, we aimed to capture a combination of vocabulary knowledge and information about the learning trajectories. These features were then used as predictor variables in order to train regression models to predict AoE word scores.

2.3 Regression Models

For each word, 39 features were generated from the learning trajectories and the mature word embeddings. Of these features, 9 are continuous (being cosine similarities) and the remainder are ordinal. Performing a variance inflation factor analysis of multicollinearity, using a threshold of 5 would reduce these features to 6. However, we found that our models, which are non-linear, perform better when multicollinearity-based pruning of features was not used. For standardizing the input features, we utilized z-score normalization prior to training the models.

Given the limited number of features generated, as well as the relatively small number of data points (i.e., 1,493 to 30,121 terms), we elected to evaluate the models using Random Forest Regression and Support Vector Regression (SVR). For Random Forest Regression, we used 50 estimator trees. For SVR, we found that the best results were produced using a radial basis function kernel, with $\varepsilon = 0.2$, $C = 1$, and with γ set to inverse of the number of features multiplied by the variance of the feature matrix.

3 Results

We measured the performance across 10 cross-validation folds and report both the mean absolute error and the mean R^2 coefficient for the test splits. For each of the three corpora, namely English-German (en-de), English-French (en-fr), and English–Spanish (en-es), we performed four experiments: one per language and one per document ordering criteria (i.e., arbitrary ordering and ordered by their Flesch Reading Ease). These results are provided in Table 1; consistently throughout all experiments, ordering ensures a more predictive model than the consideration of texts in a random order.

Table 1 Cross-validation results for predicting AoA scores

Corpus	Language	Ordering	Random Forest		Support Vector Regressor	
			MAE	R^2	MAE	R^2
EN-DE	English	Arbitrary	1.95	0.34	1.94	0.35
		Sorted	1.87	0.39	**1.85**	**0.40**
	German	Arbitrary	1.67	0.27	1.84	0.18
		Sorted	**1.67**	**0.28**	1.84	0.19
EN-ES	English	Arbitrary	1.97	0.33	1.97	0.34
		Sorted	1.88	0.39	**1.87**	**0.40**
	Spanish	Arbitrary	1.53	0.16	1.56	0.14
		Sorted	1.44	0.25	**1.41**	**0.27**
EN-FR	English	Arbitrary	2.02	0.31	2.02	0.31
		Sorted	1.90	0.37	**1.89**	**0.38**
	French	Arbitrary	1.82	0.12	1.75	0.14
		Sorted	1.67	0.21	1.65	**0.24**

The first observation is that the ordering the documents by their English Flesch Readability Score seems to bring an improvement of performance in all cases. This strengthens our hypothesis that the Readability Score as measured on the English document offers a reasonable proxy for its foreign-language counterpart. Additionally, English results are consistent between the three corpora and do not appear to be correlated to the size of each corpus in terms of the number of documents (see Fig. 2).

The AoA word lists differed in the range of possible AoA scores. Hence, comparing the results between languages using the mean absolute error does not provide a good estimate of model performance. The R^2 coefficient, on the other hand, shows that the English models have a much better performance, while the other languages tend yield results in the 0.24–0.28 range. One immediate explanation for this might be that the English word list is much larger than the others, which translates into more sample points for training the regression models. Additionally, the English word list is the most normally distributed of the four (see Fig. 3), which may also help explain the better performance of the models trained on the English data. While the German and Spanish results are similar, the French results are slightly lower. These results may be attributed to there being words in the French word list and their relatively non-normal distribution.

For the SVR models with radial basis functions, extracting feature importance directly is not possible because the data is projected into another dimensional space. For the Random Forest Regressors, feature importance can be extracted by measuring the impurity (i.e., the Gini importance); however, this method has been shown to be biased towards features with high cardinalities [15]. Thus, a better alternative for our case was to use permutation importance.

While we did find variance in terms of the order of the top features, the most important ones were always those in the "Learning Trajectory Features" category (see Sect. 2.2). Statistical information about the learning trajectories (i.e., slope, average) or the values of the points of the learning trajectories (i.e., the cosine similarities between intermediate models and the mature model) were found to have higher feature importance scores than the Mature Model Features, across all languages and ordering criteria. This aligned with our expectations because the learning trajectories were intended to simulate the way in which humans acquire new words in their vocabulary.

4 Conclusions

This study explores the possibility of estimating AoA scores for multiple languages, through a simulation of human word acquisition. Statistical features generated from the learning trajectories were then used to train regressors capable of predicting AoA scores. Expanding on the work done in the AoE model [5], we applied Word2Vec on incrementally increasing corpora of texts, and then generated features based on the resulting learning trajectories. AoA score regressors were trained, achieving reasonable results, with R^2 coefficients ranging from 0.27 to 0.40 on word lists for four languages: Spanish, German, French and English. The post-training feature importance analyses confirmed that the generated features from the learning trajectories were rated as being the most relevant by the regressors. Additionally, empirical observations reveal that our simulated learning trajectories captured differences in word acquisition between languages that are also present in AoA word lists, with certain words having lower AoA scores in one language (e.g., Spanish) than in the others – this corresponds to less steep learning trajectories for that particular language. Our approach can be uniformly applied for any language and has strong potential to help bridge the gap in word complexity research for non-English languages.

Our approach of automatically estimating AoE scores opens up the possibility of expanding existing word lists. Generalizing from the regression training data (i.e., the human-sourced AoA lists) allows us to estimate AoE scores for the entirety of the English, German, French and Spanish vocabularies that were present in the corpora during training (i.e., over 40,000 words for each language). Having access to more complete AoA lists can positively impact research on textual complexity and reading comprehension. Comparisons between learning trajectories of words in different languages, as shown in Fig. 5, highlight notable differences in word acquisition that could form the basis of better L2 learning systems through the creation of curriculums that take multicultural lingual differences into account.

The principal limitations of our method relate to the distributions of the scores in the AoA word lists used to train the regressors, as well as the cardinality of the AoA lists. Our results indicate that the English word list, which is normally distributed and has a large number of terms, leads to better regression results with higher R^2 coefficients. Training the language models is also a limiting factor because it is a computationally expensive process. For each language, we trained 20 Word2Vec models on up to 31,374,161 documents, for 50 epochs each. A possible avenue of research would be to explore the possibility of using smaller datasets and to find a criterion for selecting adequate documents. When choosing the "Para Crawl" dataset, we looked at the distribution of Flesch Reading Ease scores on the corpora to ensure that a sufficient range of complexity existed in the texts; however. Other methods might allow for the targeted selection of documents in order to not use the entire dataset. Another avenue of research would be to explore the use of different language models. In addition to previously used methods, namely LSA and LDA, temporal word embedding models [16, 17] can be used to model diachronic changes in vocabulary and could be applied to the cumulatively increasing language exposure corpus used to simulate human learning.

This study illustrates the potential of machine learning to inform measures of word complexity across different languages. The ability to predict word complexity enhances teachers' and researchers' capacity to develop instructional materials for a broader range of students, and for particular student abilities. For example, research on AoA scores has demonstrated processing advantages for phrases consisting of low-AoA words compared to high-AoA words [18]. Thus, texts might be modified by replacing words with low-AoA or high-AoA synonyms (e.g., "the dog ate my homework" *versus* "the dog devoured my essay"). Providing students with personalized materials is critical for learning because the readability of texts is partially influenced by the difficulty of words in relation to students' vocabulary, prior knowledge, and reading skills. Mulilingual AoE provides a potential means to enhance foreign language learning materials by focusing on the aspects that are either easier or harder to understand by students of different cultures. Because our method is applied uniformly across languages, it can be readily used in multilingual textual complexity applications and can help bring research in non-English languages to parity.

Acknowledgements. This research was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number TE 70 *PN-III-P1-1.1-TE-2019-2209,* ATES – "Automated Text Evaluation and Simplification", the Institute of Education Sciences (R305A180144 and R305A180261), and the Office of Naval Research (N00014-17-1-2300; N00014-20-1-2623). The opinions expressed are those of the authors and do not represent views of the IES or ONR.

References

1. Rayner, K., Duffy, S.A.: Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. Memory Cogn. **14**(3), 191–201 (1986)
2. Rosa, K.D., Eskenazi, M.: Effect of word complexity on L2 vocabulary learning. In: 6th Workshop on Innovative Use of NLP for Building Educational Applications, pp. 76–80. ACL, Portland, Oregon (2011)
3. Maddela, M., Xu, W.: A word-complexity lexicon and a neural readability ranking model for lexical simplification. arXiv preprint, arXiv:1810.05754 (2018)
4. Kuperman, V., Stadthagen-Gonzalez, H., Brysbaert, M.: Age-of-acquisition ratings for 30,000 English words. Behav. Res. Methods **44**(4), 978–990 (2012)
5. Dascalu, M., McNamara, D.S., Crossley, S.A., Trausan-Matu, S.: Age of exposure: a model of word learning. In: 30th AAAI Conference on Artificial Intelligence, pp. 2928–2934. AAAI Press, Phoenix, AZ (2016)
6. Landauer, T.K., Kireyev, K., Panaccione, C.: Word maturity: a new metric for word knowledge. Sci. Stud. Reading **15**(1), 92–108 (2011)
7. Landauer, T.K., Dumais, S.T.: A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. Psychol. Rev. **104**(2), 211–240 (1997)
8. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**(4–5), 993–1022 (2003)
9. Esplà-Gomis, M., Forcada, M.L., Ramírez-Sánchez, G., Hoang, H.: ParaCrawl: Web-scale parallel corpora for the languages of the EU. In: Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks, pp. 118–119. ACL, Dublin, Ireland (2019)
10. Flesch, R.: A new readability yardstick. J. Appl. Psychol. **32**(3), 221–233 (1948)
11. Ferrand, L., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., Brysbaert, M.: Age-of-acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. Behavior Res. Methods **40**(4), 1049–1054 (2008)
12. Alonso, M.A., Fernandez, A., Díez, E.: Subjective age-of-acquisition norms for 7,039 Spanish words. Behavior Res. Methods **47**(1), 268–274 (2015)
13. Birchenough, J.M., Davies, R., Connelly, V.: Rated age-of-acquisition norms for over 3,200 German words. Behavior Res. Methods **49**(2), 484–501 (2017)
14. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representation in vector space. In: Workshop at ICLR, Scottsdale, AZ (2013)
15. Altmann, A., Toloşi, L., Sander, O., Lengauer, T.: Permutation importance: a corrected feature importance measure. Bioinformatics **26**(10), 1340–1347 (2010)
16. Yao, Z., Sun, Y., Ding, W., Rao, N., Xiong, H.: Dynamic word embeddings for evolving semantic discovery. In: 11th ACM International Conference on Web Search and Data Mining, pp. 673–681. ACM, Marina Del Rey, CA, USA (2018)
17. Di Carlo, V., Bianchi, F., Palmonari, M.: Training temporal word embeddings with a compass. In: AAAI Conference on Artificial Intelligence, vol. 33, pp. 6326–6334, Honolulu, Hawaii, USA (2019)
18. Arnon, I., McCauley, S.M., Christiansen, M.H.: Digging up the building blocks of language: age-of-acquisition effects for multiword phrases. J. Memory Lang. **92**, 265–280 (2017)