

Exploring Readability Assessment and Text Simplification for Science Documents within a Multilingual Educational Context

Suna Şeyma Uçar¹

¹HiTZ Basque Center for Language Technologies - Ixa NLP Group, University of the Basque Country UPV/EHU

Abstract

The pedagogical approaches promoted by the implementation of the Basic Education curriculum in the Basque Autonomous Community (BAC), in particular, integrated treatment of languages and problem-based learning, require substantial dedication on the part of the teachers for the joint preparation of teaching material. In this context, this project aims to make advanced use of Artificial Intelligence to develop a multilingual text characterisation system that help teachers in the task of collection and analysis of texts. For this purpose, different types of characterisation methods related to text classification and simplification will be studied, implemented and evaluated. We will work with Basque, Spanish and English, keeping up with the languages of the educational context of the BAC. In addition, teachers will be involved throughout the process. They will contribute their experience to the design and development of the project and measure its suitability and effectiveness. The project will enable the transfer and adaptation of research from educational computing and natural language processing to education. To achieve these goals, we will propose a methodology for teachers to follow so that they can adjust and assess materials according to the student profiles in the classroom.

Keywords

readability assessment, text simplification, multilingual educational context, science documents

1. Introduction

The Secondary Education curriculum of the Basque Autonomous Community (BAC) (DECRETO 236/2015, 22 December) promotes the implementation of methodological approaches that encourage meaningful, active and conscious learning, which favours the development of both the autonomy of students and their ability to work in groups. It also emphasises interdisciplinary and multilingual approaches that promote the transfer of knowledge and skills. Within this context, it advocates the combination of the Integrated Treatment of Languages (ITL) [1] and Problem-Based Learning (PBL) [2].

Previous research [3] shows that a large majority of teachers are in favour of implementing these methodologies in schools. However, there are several obstacles that hinder this process. Among the main challenges are the difficulty to coordinate among teachers [4], which makes it very hard to implement joint programmes involving several disciplines. Furthermore, the interdisciplinarity and multilingualism inherent in these approaches, combined with the facilitating

Doctoral Symposium on Natural Language Processing from the Proyecto ILENIA, 28 September 2023, Jaén, Spain.

✉ sunaseyma.ucar@ehu.eus (S. Ş. Uçar)

ORCID 0000-0002-8100-7970 (S.Ş. Uçar)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

role of the teacher, results in the need to compile a wide range of texts and activities on specific topics and characteristics. Several publishers have joined this perspective and have published textbooks based on PBL. However, there are only a few interdisciplinary initiatives, for example Eki from Ikastolen Elkartea¹, which make it possible to combine ITL and PBL. This scarcity of material is mainly due to the need to adapt to the different language profiles, competences and interests of the learners, which implies an unprecedented level of flexibility. Thus, there is a clear deficiency in the efficient and appropriate implementation of these methodologies [5].

In this context, technologies based on Artificial Intelligence (AI) will be proposed, since as stated in the Spanish Strategy for R+D+i in AI [6, p.29] *“El uso de sistemas inteligentes permitirá transformar la educación española a partir de diferentes tecnologías, garantizando una formación inclusiva, renovada y adaptada a las necesidades de estudiantes y docentes en función de las preferencias, conocimiento y la evolución individual del estudiante.”*² More specifically, the current work aims to investigate methods of text/resource characterisation in order to simplify this task for teachers involved in joint efforts. Accordingly, two scenarios will be designed, implemented and evaluated in a multilingual educational context: Automatic Readability Assessment (ARA) and Text Simplification (TS).

ARA is used to determine the level of difficulty a written text might pose to a reader. Our ultimate goal is to develop an ARA model to assist teachers in determining the suitability of a text for a specific group of students in a multilingual educational setting. In this project, the behavior of the ARA models for Basque, Spanish, and English (for non-native learners) is of particular interest. We focus on building Machine Learning (ML) and Deep Learning (DL) models that can predict the readability of Science, Technology, Engineering and Mathematics (STEM) texts for secondary school students.

TS is the task of adjusting a text according to the needs of the reader without making drastic changes in the meaning. With the help of Large Language Models (LLMs) our main goal is to simplify texts in Basque, Spanish and English that are appropriate for classroom use and evaluate the performance of LLMs with experts.

In summary, this project aims at proposing solutions to facilitate the teachers' work in bringing assorted science materials in different languages together with the help of Natural Language Processing (NLP). The project is challenging in that it requires the multilingual adaptation and extension of NLP approaches for specific educational contexts. It also represents a direct social contribution to the educational context of the BAC, as the creation of models that facilitate the work of teachers in their daily work is envisaged.

2. Related Work

Previous research in ARA focused on mathematical equations which calculated the readability of a text based on linguistic features such as sentence length, number of syllables per word and number of paragraphs. In more recent studies [7], their efficiency was questioned as they

¹<https://www.ikaselkar.eus/>

²“The use of intelligent systems would allow transforming Spanish education from different technologies, guaranteeing an inclusive, innovative and tailored training adapted to the needs of students and teachers according to the preferences, knowledge and individual evolution of the student”.

ignore an important number of aspects present in a text. Recent advances in the computational capacity and development of ML approaches made it possible to build more reliable models based on a more varied set of linguistic features.

Feature-based algorithms were explored after statistical models in ARA, which were treated as a form of regression problem [8], classification task [9], or ranking problem [10] with a growing tendency to use Support Vector Machine (SVM) classifiers for text classification. More recently, DL approaches have gained popularity, for example, predictions of textual embeddings such as the Hierarchical Attention Network (HAN) and Bidirectional Encoder Representations from Transformer (BERT) models have been used as additional features in SVM models and evaluated in WeeBit and Newsela [11]. Azpiazu and Pera [12] used multi-attentive Recurrent Neural Networks (RNNs) on the VikiWiki dataset and obtained an accuracy of 84.7%. Lee and Vajjala [13] worked on a neural pairwise ranking model and obtained a zero-shot cross-lingual ranking accuracy of over 80% for Spanish when trained on English data from Newsela.

For evaluating ARA, various metrics have been used. Heilman et al. [14] evaluated their statistical and feature-based ML model with 10-fold cross validation and they used root mean square error (RMSE), Pearson's correlation coefficient, and accuracy to report their results. Vajjala and Meurers [9] evaluated their regression model with 10-fold cross validation and report accuracy. Similarly, Quispesaravia et al. [15] also selected 10-fold cross validation for evaluation and report precision, recall and F measure. DL models also use cross-validation as a common evaluation method. Meng et al. [16] and Martinc [17] employed 5-fold cross-validation to evaluate their DL models.

One of the main challenges of ML is freely available high-quality data. For Basque there are quite limited resources when it comes to ARA. Gonzalez-Dios et al. [18] compiled a corpus of Basque texts in two levels: simple and complex. For Spanish one of the mostly utilized corpus in ARA is VikiWiki by Azpiazu and Pera [12] and the Spanish version of the NewsEla corpus by Lee and Vajjala [19]. For English more data is available, recently WeeBit [7], OneStopEnglish [20] and Newsela [21] have been widely used in English ARA research.

Regarding TS, research has been conducted to enhance text readability for various reader profiles, ranging from learners with aphasia [22] to non-native speakers [23]. Initially, TS systems primarily relied on rule-based approaches [24]. Later, data-driven methods [25] emerged. To simplify Spanish texts, a hybrid approach combining rule-based and statistical methods was adopted by Bott et al. [26]. An approach called SimpleTT was proposed to extract simplification rules by Feblowitz and Kauchak [27]. More recently, neural approaches, including LLMs, have also been explored by Alva-Manchego et al. [28, 29]. LLMs have been employed in TS tasks, such as mBart, T5, and mT5 on the Wiki-Large dataset for English and the NewsEla corpus for Spanish in the work of Štajner et al. [30]. While recently we see prevalent work in science domain, such as CLEF 2023 [31], it has primarily explored the TS task on sentence level. Notably, Wu and Huang [32] explored sentence level TS task on science domain with T5 and GPT-4 models. In a similar vein, Engelmann et al. [33] employed T5, PEGASUS and ChatGPT on short texts obtained from scientific publications for the shared task. In a different approach, Rets et al. [34] conducted a manual simplification experiment where teachers simplified academic texts for their students. The authors of the study compiled a list of ten actions followed by the teachers during the process. TS has also been conducted in low-resourced languages such as Basque [35, 36, 37]. However, the exploration of TS in the science domain remains unexplored

in Basque, Spanish and English.

Corpora play a crucial role in TS tasks for model development and evaluation purposes. In the context of Basque, Gonzalez-Dios et al. [18] compiled 200 articles from science and technology magazines, while Bott et al. [38] collected 200 news articles in Spanish tailored for individuals with disabilities. Gonzalez-Dios et al. [39] introduced IrekiaLF_es, which is an open corpus for Spanish TS, covering original texts and their corresponding easy-to-read versions from the Basque Government. The Simple English Wikipedia (SEW), which comprises simplified English texts, introduced by Coster and Kauchak [40] is widely utilized as well. The NewsEla corpus by Xu et al. [41] consisting of 1,130 articles, each having five simplified versions is also often employed. In this work, particular emphasis should be placed on the evaluation aspect. TS tasks involve both automatic evaluation and human evaluation. Automatic evaluation methods typically rely on metrics borrowed from machine translation, such as Bilingual Evaluation Understudy (BLEU) [42], which measures the extent to which n-grams in the translated text match those in the reference, or Translation Edit Rate (TER) [43], which calculates the number of edits required for a translation. On the other hand, human evaluation considers factors such as fluency, simplicity, and adequacy [44].

Human evaluation studies involving LLMs utilize binary questions and interviews, as explored by Bhat et al. [45]. To assess grammaticality and meaning preservation, Narayan and Gardent [46] employed human participants who were asked to rate both original and simplified sentences. Vu et al. [47] conducted human evaluation using a 5-point Likert-scale to assess fluency, adequacy, and simplicity. Similarly, Štajner [30] utilized a 5-point Likert-scale to evaluate grammaticality, meaning preservation, and simplicity.

3. Automatic Readability Assessment

In this study, we aim to evaluate the effectiveness of NLP techniques for ARA of educational materials in the Obligatory Secondary Education (ESO) system in the BAC. The ESO system comprises four grades and caters to students aged between 12 and 16. In BAC, the majority of classrooms have two official languages - Basque, which is the minority language and the primary language of instruction, and Spanish, which is the majority language - along with English as a foreign language.

Our primary objective is to develop an ARA model for a multilingual context that can assist teachers in determining the suitability of a text for a specific group of secondary education students. We specifically focus on predicting the readability level of STEM subject texts in Basque, Spanish and English, which is an area that has received little research attention. In particular, we aim to accomplish this with science documents.

To develop an NLP-based model, we need both efficient learning algorithms and annotated science text corpora. However, to the best of our knowledge, there are no publicly available, domain-specific graded corpora for science texts in Basque, Spanish and English for secondary education. Therefore, the first step of our research involves compiling annotated document-level corpora for the three languages. The corpus compilation process is planned to yield a science text corpus with four ESO levels in three languages, with the documents that have already been categorized into their respective ESO levels. Given the nature of our context, the first two

corpora will consist of texts for native speakers while the English corpus will comprise texts created for non-native learners. The aim is to collect open, context specific corpus.

In terms of learning techniques, we will investigate the performance of ML approaches using the SVM algorithm³ and DL approaches using transformer architectures. Once the experiments determine the best model for classification, an evaluation will be carried out to measure the usefulness of the model in a real scenario.

4. Text Simplification

In the last months we have witnessed a sudden rise of AI tools such as ChatGPT⁴ and BingChat⁵ which are massively used.⁶ It seems that this AI technology based on language processing has become mainstream. These tools are widely used by the education community whether they are students or teachers, among others, due to their user-friendly design and free use. However, we do not know how they work and if these products are successful and reliable enough when it comes to classroom use. In fact, there are no studies that evaluate their performance with teachers yet. Similarly, open alternatives such as LLaMA [48], Alpaca [49] or Vicuna [50] could play also a role in this environment if they are appropriately promoted. To study these aspects, we will focus on the use of LLMs in science domain TS. We select science dissemination articles as our corpus since they might be challenging for classroom settings where different student profiles exist.

Concretely, our objective is to define a methodology or scenario that helps teachers to harness the potential of current language technology-based AI. Specifically, we aim to use LLMs in TS tasks and assess their performance in collaboration with teachers and educational content creators.

Starting from scientific dissemination articles, we will analyse various private and open LLMs in order to find the best approaches when working with Basque, Spanish and English documents. This involves determining the best prompts for each model and language, for example. In order to determine the appropriate type of prompt, it is crucial to understand the process through which teachers engage in text simplification. A study conducted by Rets et al. [34] examined this aspect by conducting an experiment involving 24 English as a foreign language (EFL) teachers. These teachers were provided with two academic texts and tasked with simplifying them for their students. As a result, the study generated a comprehensive list of strategies commonly employed by teachers during the text simplification process. We aim to use these list as our guide to define the best prompts for text simplification.

Once the prompts are defined, we will conduct a series of evaluations in order to set the best strategies to simplify texts to be used in the classroom. For that, an evaluation methodology will be defined and presented as part of this task.

³One of the most used algorithm in ARA.

⁴<https://chat.openai.com/>

⁵<https://www.bing.com/chat>

⁶According to the latest data, ChatGPT has more than 100 million users.

5. Conclusions

This project aims to provide a valuable tool for teachers who want to find, adjust and diversify their teaching materials in accordance with their student profile. By applying state-of-the-art NLP techniques, this project seeks to facilitate the creation and adaptation of science domain educational resources in different languages, especially those that are found in the BAC curriculum (Basque, Spanish and English). The involvement of teachers in the design and evaluation of the project will ensure that it meets their needs and expectations, as well as those of their students.

6. Acknowledgements

This work was partly supported by: University of the Basque Country UPV/EHU (PIF20/154 UPV/EHU 2020), Basque Government (IT1437-22 and IT1570-22) and Spanish Ministry of Science and Innovation (PID2021-127777OB-C21).

I would like to express my gratitude to my thesis supervisors Itziar Aldabe, Ana Arruarte and Nora Aranberri who have provided me with valuable guidance and help throughout the process of the work I have completed so far.

References

- [1] O. Guasch Boyé, et al., Reflexión interlingüística y enseñanza integrada de lenguas, *Textos de Didáctica de la Lengua y la Literatura* (2008).
- [2] P. Morales Bueno, V. Landa Fitzgerald, *Aprendizaje basado en problemas* (2004).
- [3] A. Habók, J. Nagy, In-service teachers' perceptions of project-based learning, *SpringerPlus* 5 (2016) 1–14.
- [4] Å. Haukås, Teachers' beliefs about multilingualism and a multilingual pedagogical approach, *International Journal of Multilingualism* 13 (2016) 1–18.
- [5] G. Wikan, T. Mølster, B. Faugli, R. Hope, Digital multimodal texts and their role in project work: Opportunities and dilemmas, *Technology, Pedagogy and Education* 19 (2010) 225–235.
- [6] Grupo de trabajo en I+D+i en Inteligencia Artificial; Secretaría General de Coordinación de Política Científica, Technical Report, *Estrategia Española de I+D+i en Inteligencia Artificial*, 2019.
- [7] S. Vajjala, D. Meurers, On improving the accuracy of readability classification using insights from second language acquisition, in: *Proceedings of the seventh workshop on building educational applications using NLP*, 2012, pp. 163–173.
- [8] L. Feng, M. Jansche, M. Huenerfauth, N. Elhadad, A comparison of features for automatic readability assessment (2010).
- [9] S. Vajjala, D. Meurers, On the applicability of readability models to web texts, in: *Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*, 2013, pp. 59–68.

- [10] M. Xia, E. Kochmar, T. Briscoe, Text readability assessment for second language learners, in: Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications, Association for Computational Linguistics, San Diego, CA, 2016, pp. 12–22. URL: <https://aclanthology.org/W16-0502>. doi:10.18653/v1/W16-0502.
- [11] T. Deutsch, M. Jasbi, S. Shieber, Linguistic features for readability assessment, in: Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications, Association for Computational Linguistics, Seattle, WA, USA → Online, 2020, pp. 1–17. URL: <https://aclanthology.org/2020.bea-1.1>. doi:10.18653/v1/2020.bea-1.1.
- [12] I. M. Azpiazu, M. S. Pera, Multiattentive recurrent neural network architecture for multilingual readability assessment, Transactions of the Association for Computational Linguistics 7 (2019) 421–436.
- [13] J. Lee, S. Vajjala, A neural pairwise ranking model for readability assessment, arXiv preprint arXiv:2203.07450 (2022).
- [14] M. Heilman, K. Collins-Thompson, M. Eskenazi, An analysis of statistical models and features for reading difficulty prediction, in: Proceedings of the third workshop on innovative use of NLP for building educational applications, 2008, pp. 71–79.
- [15] A. Quispesaravia, W. Perez, M. Sobrevilla Cabezudo, F. Alva-Manchego, Coh-Matrix-Esp: A complexity analysis tool for documents written in Spanish, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), European Language Resources Association (ELRA), Portorož, Slovenia, 2016, pp. 4694–4698. URL: <https://www.aclweb.org/anthology/L16-1745>.
- [16] C. Meng, M. Chen, J. Mao, J. Neville, Readnet: A hierarchical transformer framework for web article readability analysis, in: J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, F. Martins (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2020, pp. 33–49.
- [17] M. Martinc, S. Pollak, M. Robnik-Šikonja, Supervised and unsupervised neural approaches to text readability, Computational Linguistics 47 (2021) 141–179.
- [18] I. Gonzalez-Dios, M. J. Aranzabe, A. D. de Ilarraza, H. Salaberri, Simple or complex? assessing the readability of basque texts, in: Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical papers, 2014, pp. 334–344.
- [19] J. Lee, S. Vajjala, A neural pairwise ranking model for readability assessment, arXiv preprint: 2203.07450 (2022). URL: <https://arxiv.org/abs/2203.07450>. doi:10.48550/ARXIV.2203.07450.
- [20] S. Vajjala, D. Meurers, Readability assessment for text simplification: From analysing documents to identifying sentential simplifications, ITL-International Journal of Applied Linguistics 165 (2014) 194–222.
- [21] W. Xu, C. Callison-Burch, C. Napoles, Problems in current text simplification research: New data can help, Transactions of the Association for Computational Linguistics 3 (2015) 283–297.
- [22] J. Carroll, G. Minnen, Y. Canning, S. Devlin, J. Tait, Practical simplification of english newspaper text to assist aphasic readers, in: Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology, Association for the Advancement of Artificial Intelligence, 1998, pp. 7–10.
- [23] G. Paetzold, L. Specia, Unsupervised lexical simplification for non-native speakers, in:

- Proceedings of the AAAI Conference on Artificial Intelligence, volume 30, 2016.
- [24] A. Siddharthan, An architecture for a text simplification system, in: Language Engineering Conference, 2002. Proceedings, IEEE, 2002, pp. 64–71.
 - [25] S. E. Petersen, M. Ostendorf, Text simplification for language learners: a corpus analysis, in: Workshop on speech and language technology in education, Citeseer, 2007.
 - [26] S. Bott, H. Saggion, D. Figueroa, A hybrid system for spanish text simplification, in: Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies, 2012, pp. 75–84.
 - [27] D. Feblowitz, D. Kauchak, Sentence simplification as tree transduction, in: Proceedings of the second workshop on predicting and improving text readability for target reader populations, 2013, pp. 1–10.
 - [28] F. Alva-Manchego, C. Scarton, L. Specia, Data-driven sentence simplification: Survey and benchmark, *Computational Linguistics* 46 (2020) 135–187. URL: <https://aclanthology.org/2020.cl-1.4>. doi:10.1162/coli_a_00370.
 - [29] S. Štajner, Automatic text simplification for social good: progress and challenges, *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (2021) 2637–2652.
 - [30] S. Štajner, K. C. Sheang, H. Saggion, Sentence simplification capabilities of transfer-based models, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, 2022, pp. 12172–12180.
 - [31] L. Ermakova, S. Bertin, H. McCombie, J. Kamps, Overview of the clef 2023 simpletext task 3: Simplification of scientific texts (2023).
 - [32] S.-H. Wu, H.-Y. Huang, A prompt engineering approach to scientific text simplification: Cyut at simpletext2023 task3 (2023).
 - [33] B. Engelmann, F. Haak, C. K. Kreutz, N. N. Khasmakhi, P. Schaer, Text simplification of scientific texts for non-expert readers, *arXiv preprint arXiv:2307.03569* (2023).
 - [34] I. Rets, L. Astruc, T. Coughlan, U. Stickler, Approaches to simplifying academic texts in english: English teachers’ views and practices, *English for Specific Purposes* 68 (2022) 31–46.
 - [35] I. Gonzalez-Dios, M. J. Aranzabe, A. Díaz de Ilarraza, A. Soraluze, Detecting apposition for text simplification in basque, in: *Computational Linguistics and Intelligent Text Processing: 14th International Conference, CICLing 2013, Samos, Greece, March 24-30, 2013, Proceedings, Part II 14*, Springer, 2013, pp. 513–524.
 - [36] M. J. Aranzabe, A. D. De Ilarraza, I. Gonzalez-Dios, Transforming complex sentences using dependency trees for automatic text simplification in basque, *Procesamiento del lenguaje natural* 50 (2013) 61–68.
 - [37] I. Gonzalez-Dios, M. J. Aranzabe, A. Díaz de Ilarraza, The corpus of basque simplified texts (cbst), *Language Resources and Evaluation* 52 (2018) 217–247.
 - [38] S. M. Bott, H. Saggion, Spanish text simplification: An exploratory study (2011).
 - [39] I. Gonzalez-Dios, I. Gutiérrez-Fandiño, O. M. Cumbicus-Pineda, A. Soroa, Irekialfes: a new open benchmark and baseline systems for spanish automatic text simplification, in: *Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2022)*, 2022, pp. 86–97.
 - [40] W. Coster, D. Kauchak, Simple english wikipedia: a new text simplification task, in: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics:*

- Human Language Technologies, 2011, pp. 665–669.
- [41] W. Xu, C. Callison-Burch, C. Napoles, Problems in Current Text Simplification Research: New Data Can Help, *Transactions of the Association for Computational Linguistics* 3 (2015) 283–297. URL: https://doi.org/10.1162/tacl_a_00139. doi:10.1162/tacl_a_00139. arXiv:https://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00139/1566780/tacl_a_00139.pdf.
 - [42] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, Bleu: a method for automatic evaluation of machine translation, in: *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, pp. 311–318.
 - [43] M. G. Snover, N. Madnani, B. Dorr, R. Schwartz, Ter-plus: paraphrase, semantic, and alignment enhancements to translation edit rate, *Machine Translation* 23 (2009) 117–127.
 - [44] S. Nisioi, S. Štajner, S. P. Ponzetto, L. P. Dinu, Exploring neural text simplification models, in: *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: Short papers)*, 2017, pp. 85–91.
 - [45] S. Bhat, H. A. Nguyen, S. Moore, J. Stamper, M. Sakr, E. Nyberg, Towards automated generation and evaluation of questions in educational domains, in: *Proceedings of the 15th International Conference on Educational Data Mining*, 701, volume 704, 2022, p. 2022.
 - [46] S. Narayan, C. Gardent, Hybrid simplification using deep semantics and machine translation, in: *The 52nd annual meeting of the association for computational linguistics*, 2014, pp. 435–445.
 - [47] T. Vu, B. Hu, T. Munkhdalai, H. Yu, Sentence simplification with memory-augmented neural networks, arXiv preprint arXiv:1804.07445 (2018).
 - [48] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al., Llama: Open and efficient foundation language models, arXiv preprint arXiv:2302.13971 (2023).
 - [49] R. Taori, I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, P. Liang, T. B. Hashimoto, Stanford Alpaca: An Instruction-Following Llama Model, <https://github.com/tatsu-lab/stanford-alpaca>, 2023.
 - [50] W.-L. Chiang, Z. Li, Z. Lin, Y. Sheng, Z. Wu, H. Zhang, L. Zheng, S. Zhuang, Y. Zhuang, J. E. Gonzalez, I. Stoica, E. P. Xing, Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90% ChatGPT Quality, <https://vicuna.lmsys.org>, 2023.