SimLex-999 for Modern Greek

Leonidas Mylonadis[♦], **Jelke Bloem**[♣♠]

♦ Amsterdam Center for Language and Communication, University of Amsterdam
♣ Institute for Logic, Language and Computation, University of Amsterdam
♠ Data Science Centre, University of Amsterdam
leonidas.mylonadis@student.uva.nl, j.bloem@uva.nl

Abstract

Human judgements of word similarity have been a core benchmark for the intrinsic evaluation of word embedding models, and continue to be used for assessing the capabilities of large language models. While word similarity benchmarks have been collected for a range of languages, none existed for Greek. We develop a Modern Greek variant of the SimLex-999 word similarity dataset by gathering similarity judgements from 90 native speakers of Greek. We then use this as a benchmark for intrinsically evaluating several Greek language models.

Keywords: semantic similarity, Modern Greek, evaluation, benchmark, lexical semantics

1. Introduction

Word embeddings are representations that organise word concepts into multi-dimensional vector space. They represent a word's meaning in a distributional manner by relating each word to all other words in whose context they appear. Word and subword embeddings are the foundation of modern language models due to their efficiency in expressing semantic relations, and due to their information density. Various applications of word and document embeddings rely on the assumption that more similar words and document are closer to each other in embedding space.

The SimLex-999 dataset (Hill et al., 2015) was the first gold standard evaluation metric designed to explicitly benchmark models for semantic similarity. To create this dataset, human participants were asked to rate pairs of words based on how similar they are. The dataset is then used as a benchmark by correlating the cosine similarity of embeddings of word pairs in the dataset with human similarity ratings of those word pairs. This contrasts with earlier benchmarks, such as WordSim353 (Finkelstein et al., 2001), where word association was quantified. Association refers to the likelihood of a concept occurring in the same context as another concept, such as coffee – mug. By contrast, semantic similarity refers to the relation of concepts which are alike in nature, such as mug – cup.

The field of Natural Language Processing is highly focused on English, which limits the options for evaluating the performance of language models on other languages. With over 13.3 million native speakers (Eberhard et al., 2025), the Greek language has importance as an official language of Greece, Cyprus and one of the official languages of the European Union. It is a language with a rich history that has been documented for over 3,400 years (Papantoniou and Tzitzikas, 2024), a far longer period of time than English and other Germanic languages.

A Modern Greek version of WordSim353 is available (Outsios et al., 2020), and a word relatedness dataset exists for Ancient Greek as well, the AGREE benchmark (Stopponi et al., 2024). However, we are not aware of any word similarity benchmark for Greek. Distinguishing between similarity and relatedness has been shown to benefit a range of language understanding tasks in NLP (see Vulić et al., 2020 for an overview), and is necessary for theoretical studies on lexical semantics.

We introduce a Greek version of the SimLex-999 dataset[1], including Greek word pairs that parallel the English ones and that have been rated for similarity by native speakers of Greek, following Resnik's (1995) methods for collecting semantic similarity judgments. We then use our new benchmark to evaluate the embedding representations of several Greek language models.

2. Related Work

The AGREE benchmark for Ancient Greek (Outsios et al., 2020) consists of a custom set of word pairs derived from common vocabulary in Ancient Greek texts and was rated by experts in Ancient Greek. Raters were instructed to rate for relatedness, not similarity. However, Modern Greek has changed to be very different from Ancient Greek both in grammar and vocabulary.

Previous comparisons of Greek word embedding models have relied on the Modern Greek version of WordSim353. Outsios et al. (2020) created this dataset by translating the English version. However, it appears that it was not re-rated by native speakers of Greek, instead using the English word association ratings with Greek words. They also

[1] https://github.com/bloemj/Simlex999-Greek

Proceedings of SIGUL 2026 Joint Workshop with ELE, EURALI, and DCLRL @ LREC 2026, pages 119–125
11–12 May 2026.

used it to evaluate FastText and Word2Vec-based models, with a FastText model trained on a Greek web corpus slightly outperforming Word2Vec.

Word pairs in WordSim353 are rated highly if the concepts are associated, even if they semantically dissimilar concepts. SimLex-999 (Hill et al., 2015) overcomes this deficit by elaborating on the distinction between the two relations in its instructions. Raters are explicitly asked to rate on the basis of similarity and are given examples of both similarity and association to assist with the distinction.

SimLex has been replicated in a range of other non-English languages (for a recent overview of available language editions, see Brans and Bloem, 2024) and the multilingual MultiSimLex (Vulić et al., 2020) has been created as an extension with more items. As of yet, we are not aware of any versions of the SimLex-999 dataset for Modern Greek, or other word similarity datasets.

2.1. Language Models

While multilingual language models are increasingly popular, monolingual models still outperform multilingual ones on various NLP tasks for mid-resource languages (e.g. the examples outlined by Vlantis and Bloem, 2025). Monolingual models in several architectures have been trained for Modern Greek. A recent overview of available models and the state of Greek NLP in general is provided by Papantoniou and Tzitzikas (2024).

Outsios et al. (2018) introduced a Word2Vec-based (Mikolov et al., 2013) embedding model trained on about 3B tokens of Greek web content, though no evaluation was provided. A Greek FastText (Bojanowski et al., 2017) model is also available, trained on Common Crawl and Wikipedia data, which supports subword tokens and thus avoids out-of-vocabulary issues.

Transformer-based models are also available. GREEK-BERT (Koutsikakis et al., 2020) is a BERT-based model for modern Greek, with the same layers and layer sizes as BERT, trained on 29 GB of Greek text. It was shown to outperform Multilingual BERT (Devlin et al., 2019) and XLM-RoBERTa (Conneau et al., 2020) on the named entity recognition, part-of-speech tagging and natural language inference tasks. No intrinsic evaluation was performed. GreekDeBERTa-v3 (Koukouvinis, 2024) was also trained on the same dataset, but with the DeBERTa approach, including the Replaced Token Detection training objective instead of Masked Language Modeling. PaloBERT (Alexandridis et al., 2021) is a RoBERTa-variant specifically trained on social media data, with 458,293 texts included in the training data.

Aside from these encoder models, there is also an encoder-decoder model, GreekBART (Evdaimon et al., 2024), a BART variant trained on 77GB of Greek text. This model was shown to outperform XLM-RoBERTa and the aforementioned Greek-BERT on summarization and natural language inference tasks, though it was outperformed by Greek-BERT on a sentiment analysis task.

As for decoder models, gpt2-greek (Radford et al., 2019) is openly available as a small GPT-2 type model (12 layers, 768-dimensional, like BERT), trained on 5GB of Greek data.

The Greek language is also supported by some recent larger generative decoder models. Meltemi-7B (Voukoutis et al., 2024) was the first large open decoder model that appeared after the rise in popularity of generative LLMs, based on Mistral-7B. More recently, Llama-Krikri-8B-Instruct (Roussis et al., 2025) outscaled and outperformed it on natural language understanding and generation benchmarks. The model is a monolingual Greek instruction-tuned model based on Llama-3.1-8B. The multilingual EuroLLM-9B (Martins et al., 2025) also supports Greek.

In this work, we perform the first intrinsic evaluation of Greek word embedding models that incorporates transformer models, and the first that uses similarity scores. As large generative decoder models are not typically used for word embedding extraction, we leave their intrinsic evaluation for future work.

3. Methodology

3.1. Translation of Dataset

The existing SimLex-999 word pairs were translated to Modern Greek following the guidelines laid out in the Multi-SimLex paper (Vulić et al., 2020).

1. All pairs in the translated set must be unique

2. Translating two words from the same English pair into the same word in the target language is not allowed. For example, it is not allowed to translate car and automobile to the same Spanish word coche.

3. The translated pairs must preserve the semantic relations between the two words when possible. When multiple translations are possible, the translation that best conveys the semantic relation between the two words found in the original English pair is selected.

4. If it is not possible to use a single-word translation in the target language, then a multi-word expression (MWE) can be used to convey the nearest possible semantics given the above points. Several loanwords were replaced in favour of Greek-specific alternatives. For example, ρακί (raki) was used instead of τζίν (gin)

(also to differentiate between gin and jean, also τζίν in Greek).

Ambiguity and polysemy were preserved as much as possible in translation (e.g. rough – άγριο) although in some cases this was not possible (e.g. cell – κελί (prison cell) instead of κύτταρο (biological cell)).

Most challenging were several pairs of words in English which would have translated to the identical word in Greek (woman / wife – γυναίκα / γυναίκα), thereby violating rule 2. In these cases, the decision was made to replace one word in the pair that would provide the same distinction without duplicating the word pair elsewhere in the set (in this case, woman / spouse – γυναίκα / σύζηγος).

The translation was lead by a native speaker Greek linguist, and was cross-referenced against suggestions from a small subset of native Greek speakers. We did not otherwise evaluate the effectiveness of the translation guidelines as it is an established procedure and our main goal is to produce and apply the benchmark.

3.2. Questionnaire Design

The instructions to participants were translated from those used for the English SimLex-999, which were designed to elaborate the difference between similarity and association relationships. The questionnaire was designed and administered on the Qualtrics data collection platform.

20 of the word pairs were chosen at random to form the "consistency set". The ratings of these word pairs were used to determine whether each individual rater was systematically rating pairs higher, lower, or inconsistently. The set was formed of 14 noun, 2 adjective and 4 verb pairs.

The participants were allocated randomly into 11 tranches. Each tranche was asked to rate 90-91 word pairs as well as the 20 word pairs in the consistency set. The randomisation was executed by Qualtrics' randomisation functionality.

Within each tranche, users were presented with 11 sections relating to noun pairs, each showing a random set of 7 word pairs from that tranche. Participants were then shown 4 verb sections, each showing 6 word pairs in a randomised order, and finally 3 adjective sections, each showing 4 word pairs at random.

3.3. Response Collection

The survey was circulated and distributed through private channels and through acquaintances of the authors. Personal data of participants was not collected, meaning it was not possible to determine the geographical or social distribution of respondents. Participants were only asked to confirm that they were over the age of sixteen, spoke Greek as a native language and did not have any learning or language disorders to the best of their knowledge.

4. Benchmark Creation

A total of 119 responses were recorded. A total of 33 responses were incomplete, either not answering any questions or exiting the survey at an early stage (<5% progress). However, out of those, 10 returned and re-did the survey at a later stage.

16 responses were excluded for incorrectly answering one of the six checkpoint questions, which were multiple choice questions where participants were asked to distinguish a synonym pair from non-synonym pairs. 3 responses were further excluded due to answering consistently with 0 ratings. This leaves us with 90 valid responses.

4.1. Ratings

Across the whole data set, each word pair was rated an average of 5.0 times, ranging from 1 rating to 15. Inter-rater agreement was measured by measuring Krippendorff's α (for ordinal data) across the consistency set of word pairs. Across all ratings it was 0.10, indicating low inter-rater agreement, with ratings of low similarity (0 and 1) showing the highest agreement (0.16 and 0.12 respectively). This overall low rating is due to the high number of raters in the consistency set and disagreement between adjacent categories. Rating distributions for specific pairs do show normal distributions around a mean, e.g. σχολείο-βαθμός (school-grade) was rated 7x 0, 11x 1, 19x 2, 12x 3, 6x 4, 0x 5 and 0x 6.

4.2. Adjustments & Exclusions

As per the original SimLex-999 paper, the ratings of the word pairs in the consistency set were used as a basis to determine whether any particular rater was scoring higher or lower on average. If a respondent's ratings of words in the consistency set were found to be systematically one or more points above or below other raters for each of those words, their ratings were adjusted. In total, 17 adjustments were made: seven respondents had their ratings systematically increased by 1 point, nine had their ratings reduced by 1 point, and one respondent had their ratings reduced by 2 points.

Additionally, as with the English SimLex-999 study, the answers of any respondents who answered incorrectly to any of the 6 checkpoint questions were excluded. This amounted to a total of 16 exclusions. Respondents who left any of the 6 unanswered were kept in.

Initially, the survey previews counted towards randomisation count, meaning that the survey began by ignoring certain tranches and word pairs. This caused the aforementioned imbalance in tranche allocation. It was adjusted and re-balanced one week after the survey had launched.

5. Overview

Overall, the average rating across the dataset was 2.43. Nouns had the highest average rating of 2.45 and verbs had a lower one of 2.39. This contrasts with English SimLex-999 results: with values rescaled from 0-6, nouns were rated the lowest (2.69) on average, while verbs were rated the highest (2.83). While the precise reason for this difference is uncertain, it may be influenced by Modern Greek's lack of the infinitive form of the verb. Without an infinitive, all the verbs in the word set were translated into the first-person singular present tense form. It could be argued that this influences the sense of similarity between verbs, with infinitives being easier to consider as similar to one another compared to conjugated forms.

Word pairs that expressed complete synonymity were rated highest on average (e.g. θυμός – οργή / anger – rage), as expected, ranging from 4.9 to 6. Several other types of relations followed: hypernymy (categories and exemplars, e.g. όργανο – συκώτι / organ – liver); words expressing a similar purpose or concept (e.g. μέρος – περιοχή / place – area) and words being paraphrased (e.g. ακρογιαλιά – ακτή / shore – coast). Cohyponymy, i.e. word pairs representing two examples of the same implicit category were the most varied, some being rated very highly (χέρι – πόδι / hand – foot; 4.5) while others were rated very low (αυτοκίνητο – ποδήλατο / car – bicycle; 1.67).

Interestingly, antonyms were not rated the lowest out of all. Antonyms were rated between 1 and 3.8, as some cases were more difficult to distinguish from similarity (βασιλιάς – βασίλισσα / king – queen; 3.8). Instead, words that had no similar features but were merely associated (e.g. ρούχα – ντουλάπα / clothes – wardrobe; 1.4) were rated the lowest of all, ranging from 0.6 to 3.7, implying that on the whole, raters understood the distinction between similarity and relatedness or association.

Compared to the English SimLex dataset, we did indeed see a similar distribution, with synonyms rated the highest on average, and different levels of hypernymy following. The different levels mentioned refer to non-immediate levels of relation as categorised by WordNet.

As a sanity check, we also computed the correlations between the ratings of Greek SimLex-999 and those of other languages. In Table 1, we see that Greek ratings show good correlations with those of other languages. The English correlation is strongest, followed by Italian. This makes sense as our benchmark was based on English SimLex-999, and Greek and Italian have had more language contact than Greek and German. This type of correlation analysis has also been performed to assess cross-language similarity in MultiSimLex (Vulić et al., 2020).

Dataset Comparison	Correlation
Greek - English	0.7044
Greek - Italian	0.6549
Greek - German	0.6223
Greek - Dutch	0.6194
English - German	0.7478

Table 1: Cross-Language Similarity Correlation. English-German value provided for reference.

Model	Layer	Corr.
FASTText	-	**0.443**
GREEK-BERT	1	0.404
PaloBERT	1	0.036
GreekDeBERTav3-base	3	0.024
GreekBART	3	0.387
GPT2-Greek	12	-0.045
mBERT	4	0.088
XLM-RoBERTa	1	0.129

Table 2: Comparison of models on Greek SimLex-999. Layer 0 is the input layer and correlations are Spearman's rank correlations.

6. Model Benchmarking

We evaluated the Greek FastText model as a static word embedding model. For transformer models, we evaluated two Greek encoder models, two multilingual encoders, GreekBART encoder-decoder and Greek GPT-2 as a small decoder model. Models were evaluated by computing the Spearman correlation coefficient between the human similarity ratings of word pairs and the cosine similarity of the embeddings of the words in the pairs.

For the transformer models, words were embedded without any context, but with a [CLS] and [SEP] token if applicable. Subword pooling was performed for subtokenized items. This follows the procedure of Bommasani et al. (2020). Then, we extracted embeddings from all layers and present the results of the best-performing layer, following Brans and Bloem (2024). For GreekBART, we extracted embeddings from the model's encoder.

The results are presented in Table 2. We observe that the static FastText embeddings slightly outperform GREEK-BERT as the best transformer model. Figure 1 shows that performance peaks in the early layers. This performance is in line with that for other languages, such as Dutch BERTje which reaches a correlation of 0.421 on Dutch SimLex (Brans and Bloem, 2024). English BERT achieves

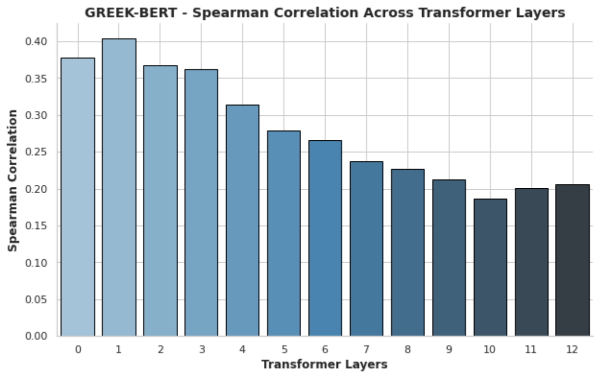

Figure 1: Spearman Correlation Across Transformer Layers in GREEK-BERT.

a correlation of 0.476 (Ehrmanntraut et al.).

7. Conclusion

We have introduced the Greek SimLex-999 resource, enabling a new kind of language model evaluation for Modern Greek. We also included a first intrinsic evaluation focused on encoder models. Future work could use this dataset to evaluate the embeddings of larger generative decoder models, as explored by Brans and Bloem (in press) for Dutch. It is also worth exploring whether instruction-tuned models can be prompted to generate Greek word similarity ratings, a task suggested by Trott (2024).

8. Ethical considerations and limitations

Word embedding evaluation based on intrinsic evaluation has various limitations, clearly outlined by Gladkova and Drozd (2016). In particular, performance on an intrinsic benchmark does not always correlate with performance on particular extrinsic tasks.

Despite the existence of MultiSimLex (Vulić et al., 2020), we only adapted the original SimLex-999 to Greek. MultiSimLex has more word pairs and we would have needed more participants than we were able to get. In future work, our dataset can be extended with the MultiSimLex items, rated following the same procedure. To facilitate this, we already rated this dataset using the 0-6 rating scale of MultiSimLex rather than the 0-10 rating scale of SimLex-999.

We are not sure to what extent our dataset is representative of the judgments of all native speakers of Greek, as we do not know where our raters were from. Our study was approved by the ethical committee of (redacted for anonymity), but without the option of collecting personally identifiable information. This potentially limits the relevance of our dataset for e.g. communities of Greek speakers outside of Greece who may not have been reached by our recruitment efforts.

References

Georgios Alexandridis, Iraklis Varlamis, Konstantinos Korovesis, George Caridakis, and Panagiotis Tsantilas. 2021. A survey on sentiment analysis and opinion mining in Greek social media. *Information*, 12(8):331.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Rishi Bommasani, Kelly Davis, and Claire Cardie. 2020. Interpreting pretrained contextualized representations via reductions to static embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4758–4781, Online. Association for Computational Linguistics.

Lizzy Brans and Jelke Bloem. 2024. SimLex-999 for Dutch. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 14832–14845, Torino, Italia. ELRA and ICCL.

Lizzy Brans and Jelke Bloem. in press. Multi-SimLex for Dutch: Benchmarking embedding- and prompt-based model performance on semantic similarity. In *Proceedings of the Fifteenth International Conference on Language Resources and Evaluation (LREC 2026)*.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. pages 4171–4186.

David M. Eberhard, Gary F. Simons, and Charles D. Fennig. 2025. Ethnologue: Languages of the world. Online version, accessed October 08, 2025.

Anton Ehrmanntraut, Thora Hagen, Leonard Konle, and Fotis Jannidis. Type-and token-based word embeddings in the digital humanities. In *CHR 2021: Computational Humanities Research Conference*, pages 16–38. CEUR Workshop Proceedings.

Iakovos Evdaimon, Hadi Abdine, Christos Xypolopoulos, Stamatis Outsios, Michalis Vazirgiannis, and Giorgos Stamou. 2024. GreekBART: The first pretrained Greek sequence-to-sequence model. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 7949–7962, Torino, Italia. ELRA and ICCL.

Lev Finkelstein, Evgeniy Gabrilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppin. 2001. Placing search in context: The concept revisited. In *Proceedings of the 10th international conference on World Wide Web*, pages 406–414.

Anna Gladkova and Aleksandr Drozd. 2016. Intrinsic evaluations of word embeddings: What can we do better? In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP*, pages 36–42, Berlin, Germany. Association for Computational Linguistics.

Felix Hill, Roi Reichart, and Anna Korhonen. 2015. SimLex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, 41(4):665–695.

Apostolos Koukouvinis. 2024. The large language model GreekDeBERTa. Bachelor's thesis, University of Athens.

John Koutsikakis, Ilias Chalkidis, Prodromos Malakasiotis, and Ion Androutsopoulos. 2020. GREEK-BERT: The Greeks visiting sesame street. In *11th Hellenic conference on artificial intelligence*, pages 110–117.

Pedro Henrique Martins, Patrick Fernandes, João Alves, Nuno M Guerreiro, Ricardo Rei, Duarte M Alves, José Pombal, Amin Farajian, Manuel Faysse, Mateusz Klimaszewski, et al. 2025. EuroLLM: Multilingual language models for Europe. *Procedia Computer Science*, 255:53–62.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Stamatis Outsios, Christos Karatsalos, Konstantinos Skianis, and Michalis Vazirgiannis. 2020. Evaluation of Greek word embeddings. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2543–2551, Marseille, France. European Language Resources Association.

Stamatis Outsios, Konstantinos Skianis, Polykarpos Meladianos, Christos Xypolopoulos, and Michalis Vazirgiannis. 2018. Word embeddings from large-scale Greek web content. *arXiv preprint arXiv:1810.06694*.

Katerina Papantoniou and Yannis Tzitzikas. 2024. NLP for the Greek language: A longer survey. *arXiv preprint arXiv:2408.10962*.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th international joint conference on Artificial intelligence-Volume 1*, pages 448–453.

Dimitris Roussis, Leon Voukoutis, Georgios Paraskevopoulos, Sokratis Sofianopoulos, Prokopis Prokopidis, Vassilis Papavassileiou, Athanasios Katsamanis, Stelios Piperidis, and Vassilis Katsouros. 2025. Krikri: Advancing open large language models for Greek. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 5012–5033, Suzhou, China. Association for Computational Linguistics.

Silvia Stopponi, Saskia Peels-Matthey, and Malvina Nissim. 2024. AGREE: a new benchmark for the evaluation of distributional semantic models of ancient Greek. *Digital Scholarship in the Humanities*, 39(1):373–392.

Sean Trott. 2024. Can large language models help augment English psycholinguistic datasets? *Behavior Research Methods*, pages 1–19.

Daniel Vlantis and Jelke Bloem. 2025. Intrinsic evaluation of mono-and multilingual Dutch language models. *Computational Linguistics in the Netherlands Journal*, 14:525–553.

Leon Voukoutis, Dimitris Roussis, Georgios Paraskevopoulos, Sokratis Sofianopoulos, Prokopis Prokopidis, Vassilis Papavasileiou, Athanasios Katsamanis, Stelios Piperidis, and Vassilis Katsouros. 2024. Meltemi: The first open large language model for Greek. *arXiv preprint arXiv:2407.20743*.

Ivan Vulić, Simon Baker, Edoardo Maria Ponti, Ulla Petti, Ira Leviant, Kelly Wing, Olga Majewska, Eden Bar, Matt Malone, Thierry Poibeau, Roi Reichart, and Anna Korhonen. 2020. Multi-SimLex: A large-scale evaluation of multilingual and crosslingual lexical semantic similarity. *Computational Linguistics*, 46(4):847–897.