Exploring Word Analogies in Embedding Spaces using Numerical Analogy

Baijia Qin[1] Yunhui Qi[1] Bartholomäus Wloka[2] Yves Lepage[1]
[1]EBMT/NLP lab, Graduate School of IPS, Waseda University, Japan
[2]Centre for translation Studies, University of Vienna, Austria
hyakuga@akane.waseda.jp, qi_101@toki.Waseda.jp,
bartholomaeus.wloka@univie.ac.at, yves.lepage@waseda.jp

Abstract

This paper proposes a new interpretation of the fact that word analogies in word embedding spaces are scarcely true parallelograms by using the notion of numerical analogy. Since numerical analogy is defined only on positive numbers, the embedding space is transformed to place all vectors in the first orthant while preserving their relative structure. This transformation is the first contribution of this paper; the second is the investigation of the deviation of word analogies from the parallelogram model using the notions of orderings and analogical powers present in numerical analogy. A third contribution is the proposal of a method to solve word analogies using numerical analogy, which is on par in performance with the most classical method, albeit under easier conditions.

1 Introduction

Analogies between words, like *Man* : *Woman* :: *King* : *Queen,* can be discovered in word embedding models because these models have been shown to encode some regular semantic information that is reflected in the relations between word vectors [1]. This analogical structure of word embedding spaces has hence been used to assess the quality of word embeddings [2]. Further research has led to the investigation of the properties of analogies in word embedding spaces [3].

In word embedding spaces, analogies are typically handled by mapping them to linear vector operations combined with a nearest-neighbor search [4]. However, linearization constrains the representation and interpretation of analogical relations in high-dimensional embedding spaces. The adoption of numerical analogy can offer a more principled approach by parameterizing analogical relationships via a so-called *analogical power* that captures non-linear structures. However, this requires that all numbers are positive. To investigate the potential of numerical analogy for explaining analogy in word embeddings, this paper makes the following three contributions:

- We propose a method to transform a word embedding space into one where all word vectors have positive values on all dimensions. Theoretically, this is achievable for word embedding spaces output by the Skip-Gram model with Negative Sampling (SGNS), because all vectors fall in an orthant of the n-dimensional space [5]. This result is also supported by the equivalence of the SGNS model with the factorization of the PPMI matrix of word co-occurrences [6].
- We propose the use of the notion of numerical analogy [7, 8] between scalars to cast a new perspective on the fact that word analogies in word embedding spaces are not really parallelograms [1]. This is possible because numerical analogy is a generalization of different kinds of analogy between numbers, including arithmetic analogy.
- We propose a method using numerical analogy to solve word analogies and show that it outperforms the usual **3CosAdd** method in the context of N-shot experiments.

2 Data Used

Word embedding models To validate our contributions, we employ pre-trained word embeddings from two widely used SGNS-based models: *Word2Vec* and *fastText*. Both of these models are trained using the *skip-gram* with negative sampling (SGNS) framework, which is known to produce embedding spaces with favorable geometric prop-

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

erties for analogy tasks. Our experiments are conducted on English and Japanese[1)] data.

Analogy datasets We use the Google Analogy Test Set, introduced in [1]. It is widely adopted for evaluating the analogy-solving capabilities of word embedding models. This dataset contains 19,544 analogy questions, consisting of 8,869 semantic and 10,675 syntactic (morphological) relations. The questions span 14 distinct relation types, including 9 morphological and 5 semantic categories. We use the original English version and a Japanese version[2)] of the dataset (see Table 5 in Appendix A for examples).

3 Transformation of Word Embedding Spaces

Because the use of numerical analogy requires that all components be positive, we first propose a transformation that aligns the centroid of the embedding space with the all-ones vector and contracts all word vectors toward it, ensuring that they lie within the positive orthant, thus eliminating any negative component while preserving the space structure.

Rotation To align the centroid $\mathbf{c}$, i.e., the average of all vectors of the embedding space, with the all-ones vector $\mathbf{u} = (1, 1, \ldots, 1)$, we follow the core idea of the *Aguilera–Pérez rotation algorithm* [9], a method for rotating objects in n-dimensional space by decomposing a general rotation into a sequence of elementary two-dimensional rotations (called a *main n-dimensional rotation* in [10]). This method iteratively applies $n-1$ elementary two-dimensional rotations, each represented by a sparse orthogonal matrix, to adjust successive coordinate pairs towards $\mathbf{u} = (1, 1)$ in two-dimensional subspaces. Scanning all n dimensions logically requires $n-1$ elementary rotations. Applying a complete series of $n-1$ elementary rotations tends to align the centroid $\mathbf{c}$ with the all-ones vector $\mathbf{u}$ by rotating the entire space toward the positive orthant. Since a single application of the $n-1$ rotation is generally not sufficient, an iterative application of a series of elementary rotations is required to gradually improve the alignment. The effectiveness of this iterative application is shown in Table 1. As the number of iterations increases, the cosine tends toward 1, indicating that the average vector is progressively getting aligned with the unit vector.

Table 1: Cosine between $\mathbf{c}$ and $\mathbf{u}$ during iterative rotations.

# of iter.	0	1	10	100	1000	5000
$\cos(\mathbf{c}, \mathbf{u})$	0.1	0.66	0.83	0.94	0.99	1.0

Contraction Aligning the centroid of the embedding space with the all-ones vector is not sufficient to ensure that all word vectors become non-negative. To guarantee that all vectors are within the positive orthant, we introduce a contraction step that progressively pulls each vector toward the all-ones vector. This is achieved through Spherical Linear Interpolation (*Slerp*) [11], a technique for interpolating between two unit vectors along the shortest path on the unit sphere. We note $\Omega = \arccos(\mathbf{x} \cdot \mathbf{y})$ the angle between two given unit vectors $\mathbf{x}$ and $\mathbf{y}$, and $\alpha \in [0, 1]$ an interpolation parameter. With this, *Slerp* is defined as:

$$\mathrm{Slerp}(\mathbf{x}, \mathbf{y}; \alpha) = \frac{\sin((1-\alpha)\Omega)}{\sin \Omega}\mathbf{x} + \frac{\sin(\alpha\Omega)}{\sin \Omega}\mathbf{y}. \quad (1)$$

The resulting vector lies on the great-circle arc between $\mathbf{x}$ and $\mathbf{y}$, preserving unit norm and angular structure. In our implementation, each word vector is first normalized, and the angle Ω with respect to $\mathbf{u}$ is computed. This angle is then nonlinearly compressed to $\Omega' \in [0, \rho_n]$, where $\rho_n = \arccos(\sqrt{(n-1)/n})$ denotes the boundary of the positive spherical cap in n-dimensional space. We then define $\alpha = \Omega'/\Omega$, and apply the above *Slerp* formula to obtain a contracted direction that remains within the positive orthant. Finally, the contracted vector is rescaled to match the original norm of the input.

Table 2: Statistics before and after contraction using *Slerp*.

	Before	After
Vocabulary size	1,198,203	1,198,203
Number of word vectors with negative components	1,198,203	0
Ave. percentage of negative components per word vector	49.21%	0.00%

Stability of the transformation Table 2 shows that the resulting embedding space satisfies the positivity constraint, i.e., no component of any vector is negative.

To assess the preservation of the neighboring structure of the embedding space, we perform two tasks: *Nearest Neighbors* and *3CosAdd Analogy*. The first task assesses whether each word maintains consistent semantic neighborhoods across the two spaces [12]. The second task

1) https://fasttext.cc/docs/en/crawl-vectors.html
2) https://github.com/taishi-i/awesome-japanese-nlp-resources

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

evaluates whether the transformation affects vector arithmetic in analogy solving [4]. For the *Nearest Neighbors* task, we randomly select a target word and retrieve its top-n nearest neighbors based on cosine similarity from both the original model and the transformed model. For the *3CosAdd Analogy* task, we randomly select a word analogy $A : B :: C : D$ from the analogy dataset, compute the predicted vector $\mathbf{v} = \vec{B} + \vec{C} - \vec{A}$, and retrieve its top-$n$ nearest word neighbors from both models. We repeat both tasks 5,000 times. Table 3 reports average stability scores: for fastText, scores remain above 0.95 across all n, with near-perfect preservation at $n = 1$; for Word2Vec, *Nearest Neighbors* drops to around 0.75, whereas *3CosAdd Analogy* remains strong. This gap likely reflects the limited ability of Word2Vec to capture morphological regularities, leading to reduced neighborhood consistency.

Table 3: Stability of transformed word embedding spaces.

Top-n	Nearest Neighbors		3CosAdd Analogy	
	Word2Vec	fastText	Word2Vec	fastText
1	0.752	0.961	0.970	0.990
3	0.764	0.957	0.923	0.970
5	0.770	0.955	0.901	0.963

4 Numerical Analogy in Word Embedding Models

Numerical analogy as a generalization of arithmetic analogy Numerical analogy is built on the notion of generalized means. For any real $p \in (-\infty, +\infty)$ and strictly positive real numbers $x_1, \ldots, x_N$, generalized mean is defined as

$$m_p(x_1, \ldots, x_N) \stackrel{\text{def}}{=} \lim_{r \to p} \left(\frac{1}{N} \sum_{i=1}^{N} x_i^r \right)^{1/r} . \quad (2)$$

Given $a, b, c, d > 0$, an analogy $a : b ::^p c : d$ in power p holds if $m_p(a, d) = m_p(b, c)$. By varying p, this framework subsumes arithmetic ($p = 1$), geometric ($p \to 0$), harmonic ($p = -1$), and infinitely many other analogies. Although 24 permutations exist, symmetry reduces them to three orderings: `a:b::c:d`, `a:c::d:b` and `a:d::b:c` [13]. Numerical analogy has two fundamental properties: (i) for any four strictly positive real numbers $a < b < c < d$ (ordering `a:b::c:d`), there exists a unique analogical power $p \in \mathbb{R}$ such that $a : b ::^p c : d$; (ii) for any three strictly positive real numbers a, b, c and a given analogical power $p \in \mathbb{R}$, there exists a unique solution d if $b^p + c^p - a^p \geq 0$.

Analogical powers and orderings in word analogies. The common conception about word analogies holds that they correspond to parallelograms in word embedding spaces [1] and thus to arithmetic analogies on each dimension. Expressed in the generalized framework of numerical analogy, it means that, at each dimension, the analogical ordering should be `a:b::c:d` and the analogical power should be 1.

This conception is challenged by the observed distribution of analogical orderings and powers obtained by the use of numerical analogy introduced above, in transformed word embedding spaces obtained using the transformation described in Section 3. Figure 1 in Appendix B presents such distributions across all dimensions for nine analogies drawn from three distinct analogy categories.

These distributions do not exhibit strikingly distinctive patterns that allow for clear separation. Firstly, the different colors in Figure 1 in Appendix B show that all three possible orderings are encountered. Still, the ordering `a:b::c:d`, which corresponds directly to the placement sequence of the four terms, accounts for more than 50% of the cases in all nine analogies. Secondly, the values of analogical powers depart highly from a unique value of 1, as expected for arithmetic analogy. In fact, they span an interval of $[-70, 70]$, with most values concentrating around zero.

Aggregation of powers The above observations indicate that word analogies do not correspond to arithmetic analogy at each dimension. Accordingly, for each dimension i, scalar components are treated as forming a numerical analogy under a given ordering, with analogical powers inferred from example analogies within the same category. Orderings and powers are computed from one or multiple examples, aggregated by averaging, and each dimension is weighted by the number of supporting examples when computing cosine similarity. When no support is available, $p = 1$ and the weight is set to 1, reducing the 0-shot case to *3CosAdd*. Experiments show that no N-shot configuration yields higher accuracy than the 0-shot case of *3CosAdd*.

Aggregation of solutions Instead of aggregating by dimension, which is disproved by observation and shown to be unsuccessful in experiments, aggregation is performed by solutions. For each few-shot example k, the predicted vector $d^{(k)}$ is solved dimension-wise according to its actual ordering, using the appropriate permutation branch among

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Table 4: Experiment results in solving analogies using numerical analogy.

N-Shot	Top-1		Top-3		Top-5	
	Accuracy	Index	Accuracy	Index	Accuracy	Index
(3CosAdd) 0	0.836 ± 0.017	1.164 ± 0.017	0.916 ± 0.012	1.358 ± 0.040	0.942 ± 0.010	1.488 ± 0.058
1	0.118 ± 0.014	1.882 ± 0.014	0.180 ± 0.017	3.540 ± 0.046	0.222 ± 0.019	5.114 ± 0.079
5	0.642 ± 0.021	1.358 ± 0.021	0.796 ± 0.018	1.812 ± 0.054	0.836 ± 0.017	2.164 ± 0.085
10	0.760 ± 0.019	1.240 ± 0.019	0.878 ± 0.015	1.502 ± 0.045	0.900 ± 0.013	1.708 ± 0.070
50	0.856 ± 0.016	1.144 ± 0.016	0.920 ± 0.012	1.332 ± 0.039	0.936 ± 0.011	1.468 ± 0.059
100	0.852 ± 0.016	1.148 ± 0.016	0.920 ± 0.012	1.330 ± 0.039	0.940 ± 0.011	1.460 ± 0.058

the three possible orderings:

$$x_i \text{ s.t.} \begin{cases} m_{p_i}(a_i, x_i) = m_{p_i}(b_i, c_i), & \text{if } \texttt{a:b::c:d}, \\ m_{p_i}(c_i, x_i) = m_{p_i}(a_i, b_i), & \text{if } \texttt{a:c::d:b}, \\ m_{p_i}(b_i, x_i) = m_{p_i}(a_i, c_i), & \text{if } \texttt{a:d::b:c}. \end{cases} \quad (3)$$

This produces multiple predicted vectors $\{d^{(k)}\}$. For each dimension i, we count ordering frequencies, assign weights $w_i^{(k)}$, and compute the aggregated prediction as

$$d_{\text{pred},i} = \frac{\sum_{k=1}^{K} w_i^{(k)} \, d_i^{(k)}}{\sum_{k=1}^{K} w_i^{(k)}}. \quad (4)$$

Since d_{pred} rarely matches an existing embedding, we take the nearest word vector under cosine similarity. When support is missing, we normalize using residual norm completion. Let $\mathcal{M}$ denote the index set of missing dimensions. Define

$$d_{\text{pred}}^{\text{no\_none}} = (v_1, \ldots, v_d) \quad \text{with } v_i = 0, \ \ \forall i \in \mathcal{M}, \quad (5)$$

with norm $\|v^{\text{no\_none}}\|_2 = \rho < 1$. The residual length is then $r = 1 - \rho^2$. If $k = |\mathcal{M}| > 0$, we set $v_i = \sqrt{r/k}$ for all $i \in \mathcal{M}$. Otherwise, if $\rho \geq 1$, all missing entries are set to 0. This ensures that d_{pred} remains normalized, consistent with the transformed embedding space.

Evaluation We evaluate on 500 analogy questions, retrieving the Top-5 nearest candidates to d_{pred}. The N-shot setting varies by category ($N = 0$ uses *3CosAdd*). Performance is measured using *Accuracy@k* (proportion of correct answers in the Top-k) and *Index@k* (average solution rank, with $k + 1$ if missing), reported for $k = 1, 3, 5$. Table 4 shows the results on English data with FastText. All results on English and Japanese data with both FastText and Word2Vec exhibit the same trends and lead to the same conclusions. We observe a notable trend: as the number of example analogies increases, the performance of numerical analogy under the *aggregation of solutions* strategy improves steadily and gradually approaches that of arithmetic analogy. For instance, with Top-5 evaluation, arithmetic analogy achieves an Accuracy@5 of 0.942, whereas numerical analogy with 100-shot reaches 0.940 with a standard error showing a statistically insignificant difference. These results suggest that numerical analogy offers a meaningful lens for analyzing analogical structure in word embedding spaces. As the number of example analogies increases, accuracy consistently improves, indicating that arithmetic analogy in real word embedding spaces corresponds to an aggregate effect over multiple analogical instances. In this sense, numerical analogy serves not as a competing solver, but as a more expressive and interpretable approach for studying analogy in word embeddings.

5 Conclusion

We revisited word analogies in word embedding spaces by using the notion of numerical analogy. As numerical analogy requires that all components of vectors be positive, a first contribution was the proposal of an effective transformation to transform a word embedding space into such a space. Word analogies in word embedding spaces are usually interpreted as parallelograms. A consequence would be that an arithmetic analogy exists at each dimension, but we showed, as a second contribution, that analogical powers on word analogies from real data actually span a range of values, and that there is no common analogical power on the same dimension for different analogies from the same category. As a third contribution, we proposed a method that aggregates solutions obtained from the analogical powers of different examples from the same category to solve word analogies and obtained performance comparable with *3CosAdd*, albeit in N-shot experiments.

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

References

[1] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In Lucy Vanderwende, Hal Daumé III, and Katrin Kirchhoff, editors, **Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 746–751, Atlanta, Georgia, June 2013. Association for Computational Linguistics.

[2] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics.

[3] Louis Fournier, Emmanuel Dupoux, and Ewan Dunbar. Analogies minus analogy test: measuring regularities in word embeddings. In Raquel Fernández and Tal Linzen, editors, **Proceedings of the 24th Conference on Computational Natural Language Learning**, pp. 365–375, Online, November 2020. Association for Computational Linguistics.

[4] Anna Gladkova, Aleksandr Drozd, and Satoshi Matsuoka. Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't. In Jacob Andreas, Eunsol Choi, and Angeliki Lazaridou, editors, **Proceedings of the NAACL Student Research Workshop**, pp. 8–15, San Diego, California, June 2016. Association for Computational Linguistics.

[5] David Mimno and Laure Thompson. The strange geometry of skip-gram with negative sampling. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing**, pp. 2873–2878, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[6] Omer Levy and Yoav Goldberg. Linguistic regularities in sparse and explicit word representations. In Roser Morante and Scott Wen-tau Yih, editors, **Proceedings of the Eighteenth Conference on Computational Natural Language Learning**, pp. 171–180, Ann Arbor, Michigan, June 2014. Association for Computational Linguistics.

[7] Yves Lepage and Miguel Couceiro. Analogie et moyenne généralisée. In Jean-Guy Mailly, François Schwarzentruber, and Anaelle Wilczynski, editors, **Actes des 18es Journées d'Intelligence Artificielle Fondamentale et des 19es Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JIAF 2024 et JFPDA 2024)**, pp. 114–124, La Rochelle, France, Juillet 2024.

[8] Yves Lepage and Miguel Couceiro. L'analogie numérique revisitée et étendue, précisée, rétrécie ou agrandie. In Jean-Guy Mailly, François Schwarzentruber, and Anaelle Wilczynski, editors, **Actes des 19es Journées d'Intelligence Artificielle Fondamentale et des 20es Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JIAF 2025 et JFPDA 2025)**, Vol. (accepted, to appear), Dijon, France, Juillet 2025.

[9] Antonio Aguilera and Ricardo Pérez-Aguila. General n-dimensional rotations. In **The 12-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG, 2004)**, pp. 1–8, University of West Bohemia, Campus Bory, Plzen-Bory, Czech Republic, February 2004.

[10] K.L. Duffin and W.A. Barrett. Spiders: a new user interface for rotation and visualization of n-dimensional point sets. In **Proceedings Visualization '94**, pp. 205–211, 1994.

[11] Ken Shoemake. Animating rotation with quaternion curves. In **Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques**, SIGGRAPH '85, p. 245–254, New York, NY, USA, 1985. Association for Computing Machinery.

[12] Laura Wendlandt, Jonathan K. Kummerfeld, and Rada Mihalcea. Factors influencing the surprising instability of word embeddings. In Marilyn Walker, Heng Ji, and Amanda Stent, editors, **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)**, pp. 2092–2102, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[13] Yves Lepage. Analogy and formal languages. **Electronic Notes in Theoretical Computer Science**, Vol. 53, pp. 180–191, 2004. Proceedings of the joint meeting of the 6th Conference on Formal Grammar and the 7th Conference on Mathematics of Language.

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

A Example analogies in the Japanese and the English languages

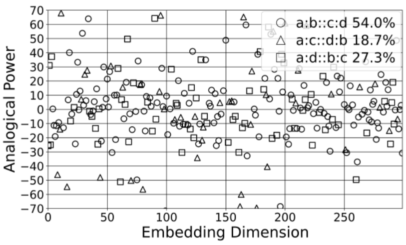
Table 5: Example analogies in both the Japanese and the English languages.

Category	Example analogy	
	Japanese	English
country - capital	中国 : 北京 :: フランス : パリ	*China* : *Beijing* :: *France* : *Paris*
name - occupation	孔子 : 哲学者 :: ピカソ : 画家	*confucius* : *philosopher* :: *picasso* : *painter*
hyponyms - misc	車 : タクシー :: 色 : 赤	*car* : *taxi* :: *color* : *red*
antonyms - binary	後 : 前 :: 下 : 上	*after* : *before* :: *down* : *up*
meronyms - substance	大気 : 酸素 :: バッグ : 革	*atmosphere* : *oxygen* :: *bag* : *leather*
adj+ness_reg	面白い : 面白さ :: 弱い : 弱さ	*interesting* : *interestingness* :: *weak* : *weakness*
verb+er_irreg	消費 : 消費者 :: 研究 : 研究者	*consume* : *consumer* :: *research* : *researcher*

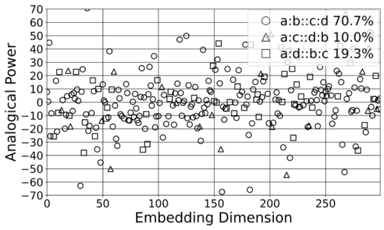
Results of selection of seven analogy categories from an analogy dataset, and for each category, and extraction of one representative analogy in both English and Japanese.

B Distribution of analogical powers and corresponding orderings

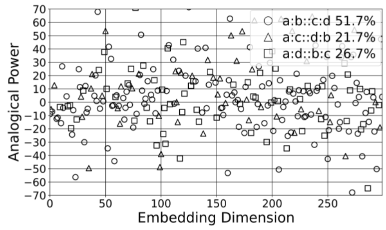
Figure 1: Distribution of analogical powers and corresponding orderings for analogies from three different categories.



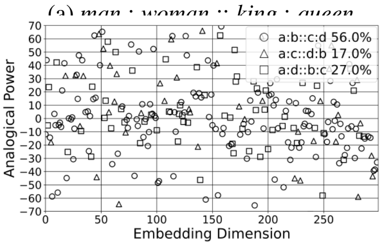
(a) *man* : *woman* :: *king* : *queen*

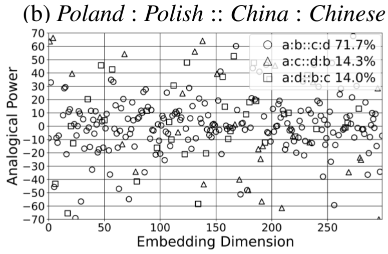


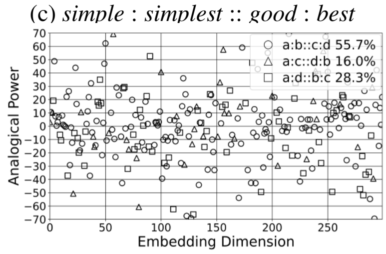
(b) *Poland* : *Polish* :: *China* : *Chinese*

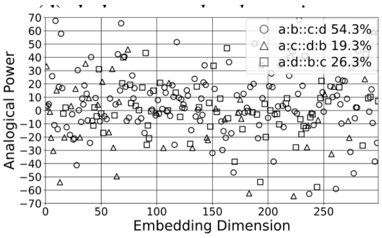


(c) *simple* : *simplest* :: *good* : *best*

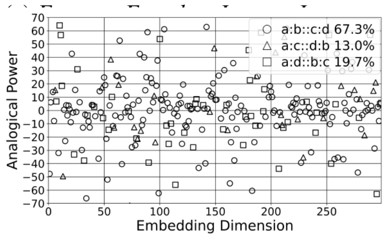




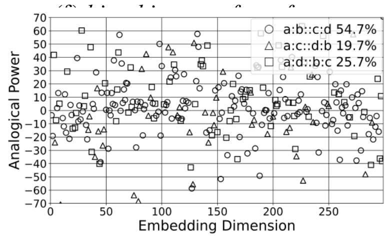




(g) *boy* : *girl* :: *he* : *she*



(h) *Greece* : *Greek* :: *Sweden* : *Swedish*



(i) *sweet* : *sweetest* :: *dark* : *darkest*

This appendix presents the distributions of analogical powers and corresponding orderings for analogies from three categories: *gender*, *nationality*, and *superlative*. For each analogy, the vertical axis represents the analogical power at each embedding dimension. Different orderings are distinguished by marker shapes, and the exact proportion of each ordering is indicated in the legend of each plot.

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).