Emergence of OV/VO language universals in neural language models

Takuma Tanaka[1*]

1 Graduate School of Data Science, Hikone, Shiga 522-8522, Shiga University, Japan

* tanaka.takuma@gmail.com

Abstract

Language typology has confirmed a considerable number of language universals, *i.e.*, the crosslinguistic correlation between two characteristics such as word order. For example, OV (object–verb order) languages tend to have postpositions (Post), whereas VO (verb–object order) languages tend to have prepositions (Pre). One of the hypotheses to explain language universals is that languages with frequently observed combinations of word orders are easily learnable. To test this hypothesis, the present study examined the learnability of neural language models of synthetic languages made from an English corpus. BERT and GPT-2 were trained with synthetic corpora consisting of combinations of VO/OV, post-nomial relative clause (NRel)/pre-nomial relative clause (RelN), and Pre/Post word orders. The results indicated that the validation loss was the lowest for the original corpus with VO-NRel-Pre and the second lowest for its complete opposite, OV-RelN-Post. Although OV-RelN-Post is the most distant from the original, it is a frequently observed word-order combination. These findings suggest that VO-NRel-Pre and OV-RelN-Post are word orders that are easy to learn and, consequently, exhibit high evolutionary fitness.

Introduction

Although there are thousands of languages spoken in a wide variety of societies and environments, linguists have noticed striking regularities among them. Since Greenberg [1], language typology has confirmed a considerable number of language universals, most of which are crosslinguistic correlations between two characteristics. The most intensively investigated correlation is that between the word order of verbs and their objects and the word order of other elements. OV (object–verb order) languages tend to have postpositions, prenominal genitives, adverb–verb order, clause-final complementizers, auxiliary verbs following verbs, and negative auxiliaries following verbs, whereas VO (verb–object order) languages tend to have prepositions, post-nomial relative clauses, verb–adverb order, clause-initial complementizers, auxiliary verbs preceding verbs, and negative auxiliaries preceding verbs [1–4].

Because no crosslinguistic correlation arises if the features are mutually independent, linguistic typologists have proposed hypotheses to explain this phenomenon. One hypothesis is that the frequently observed combinations of word orders are easily learnable. Hawkins [5] argued that the frequently observed combinations exhibit Cross-Category Harmony. In light of Cross-Category Harmony, the word orders correlated with OV and VO are summarized as the heads following and preceding their dependents, respectively. In other words, the consistency in the order of the heads and their dependents reduces cognitive load, and, consequently, languages with consistency are more likely to appear. Theoretical frameworks such as branching direction theory [4] and dependency length minimization [6] point in the same direction. Another hypothesis is that crosslinguistic correlation is the result of historical evolution reflecting the properties of ancestral languages [7, 8]. According to this hypothesis, the observed language universals are not universal tendencies, but rather lineage-specific traits. These two hypotheses are not mutually exclusive. Their potential influence on language evolution should be measured individually.

Hence, we need to experimentally examine the learnability of languages with various word-order combinations. Even if there is a strong history-dependency, the rarity of a combination can be partly explained by the difficulty of learning it. The learnability of word orders has been experimentally examined [9–11]. These experiments investigated the performance of human subjects learning artificial languages and found that frequent word orders are easily learnable. However, some studies have reported that infrequent word orders are not necessarily less learnable [12]. Although valuable findings have been obtained, excluding the influence of the native and second languages remains difficult for such studies.

To avoid the difficulties arising from using human participants, some previous studies have used neural networks. Recent advances in neural network models have allowed language models such as BERT [13] and GPT-2 [14] to achieve human-level performance in translation, question answering, and inference. Impressive performance across the entire field of natural language processing suggests the similarity in language processing between neural language models and humans. Under the assumption that their processing is similar on a not-superficial level, the learnability of a language can be examined by measuring the performance of a neural language model on this language. On the basis of this assumption, neural network models have been used in linguistic studies [15]. Artificial languages to be given as input to neural network models have been developed [16]. Using this artificial language, Kuribayashi *et al.* [17] showed that frequently observed word orders are more learnable. However, algorithmically generated texts of an artificial language may not be as rich as those of natural languages. To address this issue, a previous study [18] transformed the texts of natural languages into a synthetic language with a different word order, which can be more ethologically valid. However, that study focused on the effect of adding suffices to mark cases and agreements and left the correlation of word orders for future work.

Hence, this paper reports the learnability of synthetic languages made from an English corpus with various combinations of word orders by neural language models. Typologically, English is an SVO language (comprising 42% of the world languages), with mostly VO-correlated characteristics. However, English can also be easily transformed into an SOV language (comprising 45% of the world languages), with OV-correlated characteristics such as Japanese [19, 20]. Specifically, verb–object, noun–relative clause, and preposition–noun orders are independently flipped to make synthetic corpora that have the same vocabulary and the same number of words. Word-order parameters are grouped as follows: flipping the verb–object order also flips the adverb–verb, auxiliary verb–verb, and negative auxiliary–verb orders; flipping the noun–relative clause order also flips the noun–genitive order; and flipping the preposition–noun order also flips the complementizer–complement clause order. The order of demonstrative, numeral, adjective, and noun, which is also an intensively studied subject in language typology [21], is therefore preserved. Restricting the number of parameters facilitates the implementation and reduces the error of transformation, particularly in incomplete sentences and those containing interjections. Neural language models were trained for eight training sets, and the validation loss was measured. Both the loss for the masked language model, which predicts the masked words in a sentence, and the causal language model, which predicts the next word, measure the learnability of the language. Comparing the prediction loss for corpora in which only the word order is changed allows us to measure the learnability of languages as a function of word order only. In other words, the effect of Cross-Category Harmony proposed by Hawkins [5] on neural language models is investigated. The present approach is advantageous because synthetic languages can be more ethologically valid than algorithmically generated artificial languages, and neural language models are not influenced by the effects of native and second languages. Examining the learnability of synthetic languages by language models, including neural language models, can explain crosslinguistic correlation free from historical, biological, and psychological factors.

This paper is organized as follows. The Methods section describes the materials and methods used in this paper. The corpus and the method of transformation, the parameters of neural language models, and the measure of distance between sentences are presented. The Results section reports the experimental results. Results with various neural language models and token sizes are shown. The relative independence of the loss from the distance from the original corpus is also shown. The frequent word order-combinations exhibit lower loss, *i.e.*, high learnability. In the Discussion section, the results are compared with those from previous studies, and future perspectives are given.

Methods

The training and validation sets were made from English v12 of OntoNotes 5.0 [22] as follows. Three transformation parameters were defined. The first is a parameter for flipping the order of verbs and their objects. Flipping the positions of the first verb and the following clauses in VP and leaving them intact are referred to as OV and VO, respectively. OV also flips adverbs and verbs, auxiliary verbs and verbs, and negative auxiliaries and verbs, which is typical for OV languages [4]. Also, for OV, inversions in SINV and SQ are converted to SOV. The second is a parameter for flipping the order of nouns and their relative clauses (RelN) and preserving them (NRel). RelN also flips nouns and genitives, nouns and their modifiers in NP, and adjectives and their following clauses in ADJP. The third is a parameter for flipping adpositions and nouns (Post) and preserving their order (Pre). Post flips prepositions and nouns in PP, NP, and ADJP. Post also flips complementizers and complement clauses, which are also frequently found in OV languages [4]. The combination of VO, NRel, and Pre, which is the word order in the original English corpus, corresponds to V_2 & N_2 & Pr in Hawkins (1983) [5], whereas the combination of OV, RelN, and Post corresponds to V_3 & N_3 & and Po, which shows the smallest Cross-Category Harmony deviations.

The sentences were preprocessed as follows. All capitals were converted to their lower cases to dispense with the need for consistent capitalization of proper nouns and words at the beginning of sentences. Abbreviated verbs and auxiliaries such as “n’t,” “’ve,” and “’re” were replaced with their corresponding unabbreviated forms. Sentences containing rare characters such as those from foreign words were omitted. Also, complicated sentences that were difficult to transform were omitted. Punctuation marks were separated from words with white spaces. Sentences containing VP without verbs and overly short sentences (fewer than four words) were also omitted. The number of sentences decreased from 819 579 to 810 360 in the training set and from 111 898 to 110 672 in the validation set. Of these, 1688 and 220 sentences, respectively, were omitted for reasons other than shortness.

The combinations of VO/OV, NRel/RelN, and Post/Pre yielded eight parallel corpora with the same numbers of sentences and words. As shown in Table 1, all

Table 1. Word order and block-interchange distance

Order	Example	d_1	d_2
VO-NRel-Pre	this is the malt that lay in the house that jack built .	0	6
VO-NRel-Post	this is the malt lay the house jack built that in that .	3	6
VO-RelN-Pre	this is that lay in that jack built the house the malt .	2	6
VO-RelN-Post	this is lay jack built that the house in that the malt .	3	6
OV-NRel-Pre	this the malt that in the house that jack built lay is .	2	6
OV-NRel-Post	this the malt the house jack built that in lay that is .	4	4
OV-RelN-Pre	this that in that jack built the house lay the malt is .	3	6
OV-RelN-Post	this jack built that the house in lay that the malt is .	4	4

corresponding sentences were made by reordering without adding or deleting words. The corpora were tokenized by SentencePiece with white spaces as an obligatory token separator [23]. Performance was then measured by cross-entropy loss, which is the logarithm of the perplexity.

The language models used in this paper were BERT [13] and GPT-2 [14], both of which are Transformer-based language models [24]. BERT is a masked language model, whereas GPT-2 is a causal language model. To reduce GPU memory usage, the dimensionality of the feed-forward layer in BERT was set to 768, and the number of layers in GPT-2 was set to eight. The other parameters were set to their default values. The training was performed for 50 epochs with a batch size of 16. The language models and data loader were implemented using HuggingFace Transformers [25].

The distance of a synthesized sentence from the original sentence was measured by the block-interchange distance [26]. The block-interchange distance is defined by the minimal number of block interchanges, with which two substrings, or blocks, are swapped to make the synthesized sentence from the original sentence. Table 1 shows the block-interchange distance from the original sentence, d_1, and that from the reverse of the original sentence, d_2. The minimum of d_1 and d_2 was used as the distance d. The average of d for all sentences in a corpus was used as the distance of the corpus from the original corpus.

Results

Eight parallel corpora were made by independently flipping OV/VO, NRel/RelN, and Post/Pre. The numbers of words and tokens in corresponding sentences were kept unchanged. The training was performed for 50 epochs, during which, the validation loss was compared between corpora to measure the learnability of each one. This was considered a fair comparison because all the corpora were permutations of the same set of tokens. The original corpora with VO-NRel-Pre and OV-RelN-Post were expected to exhibit low validation loss because they are frequently found combinations.

Fig 1 shows the time course of the validation losses with BERT. Each curve corresponds to one run for each corpus with 5000 tokens. The loss was the lowest for the original corpus (VO-NRel-Pre) and the second lowest for OV-RelN-Post. As expected, the frequently found combinations exhibited the lowest validation losses, indicating that the prediction is easiest. The scatterplots in Fig 2 show the validation losses of BERT at the 10th, 20th, 30th, 40th, and 50th epochs for 10 runs each. VO-NRel-Pre and OV-RelN-Post were consistently the best and second best, respectively. The corpora with low validation loss at an early stage exhibited low validation loss at a late stage. However, the absolute difference in the validation losses among the corpora decreased at a late stage. This suggests that the difference in

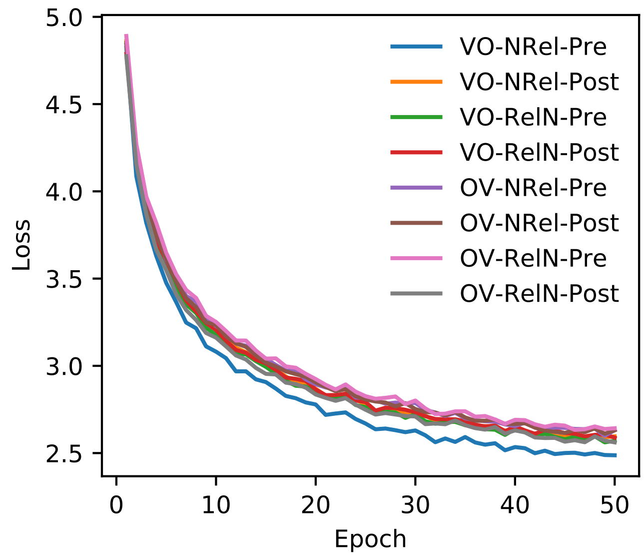

Fig 1. Time course of validation losses for BERT with 5000 tokens.

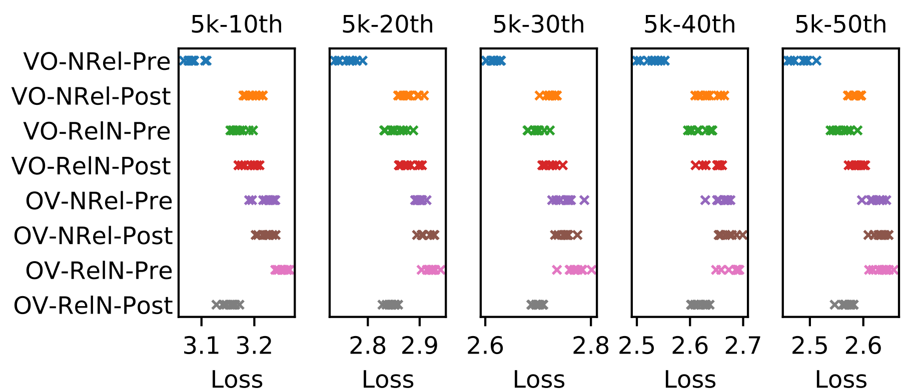

Fig 2. Validation losses for BERT at 10th, 20th, 30th, 40th, and 50th epochs with token size of 5000 for 10 runs each.

learnability is prominent in the early stage of learning. Fig 3 shows the validation losses at the 10th and 50th epochs for the token sizes of 10 000 and 20 000. VO-NRel-Pre and OV-RelN-Post consistently exhibited low validation loss, although VO-RelN-Pre also exhibited low validation loss at the 50th epoch.

Because these results were obtained with BERT (a masked language model), GPT-2 (a causal language model) was examined to test generality. Similar to BERT, GPT-2 exhibited the lowest validation losses for VO-NRel-Pre and OV-RelN-Post (Fig 4). Although BERT and GPT-2 have different structures and training strategies, this result suggests that these language models have common preferences regarding word-order correlation.

However, validation loss may reflect similarity to the original corpus, but not its learnability. Specifically, OV-RelN-Post reverses the order of most elements, sparing only the orders of subjects and verbs and of adjectives and nouns. If OV-RelN-Post sentences are the almost complete reverse of VO-NRel-Pre sentences, they may exhibit almost the same high level of learnability. Because BERT does not distinguish a sentence from its reverse, this possibility must be examined. To exclude this

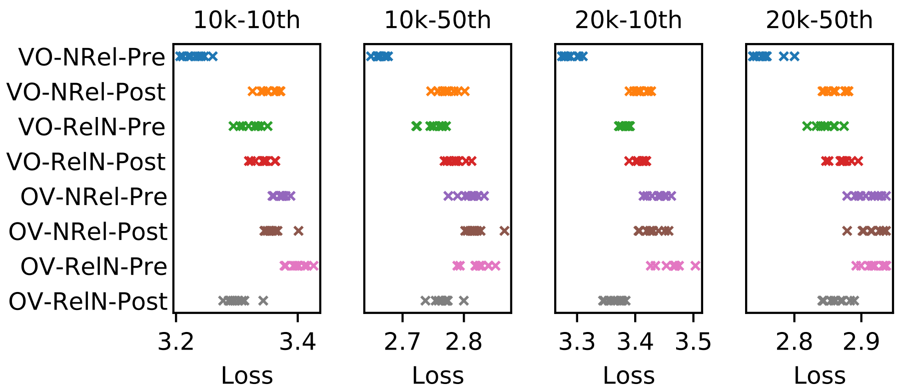

Fig 3. Validation losses for BERT at 10th and 50th epochs with token sizes of 10 000 and 20 000 for 10 runs each.

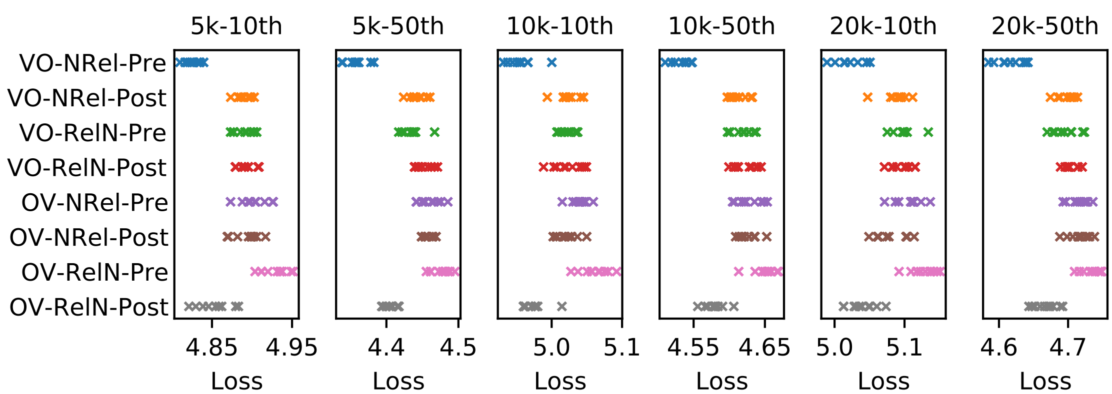

Fig 4. Validation losses for GPT-2 at 10th and 50th epochs with token sizes of 5 000, 10 000, and 20 000 for 10 runs each.

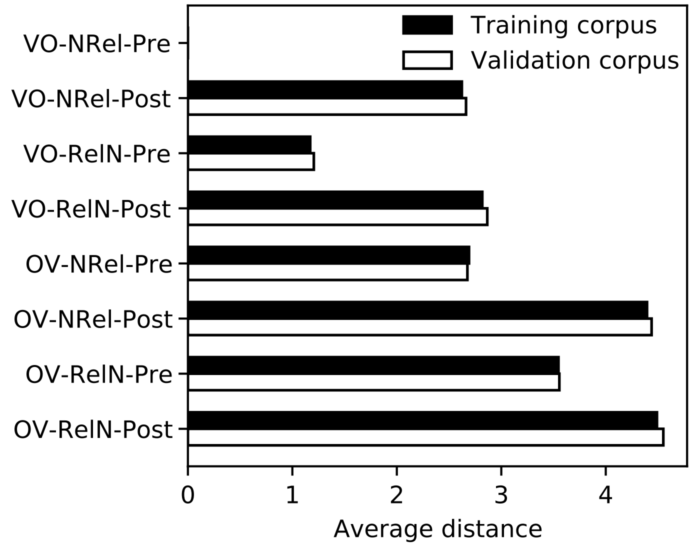

Fig 5. Average distance d of sentences in each corpus from those in original (VO-NRel-Pre).

possibility, the distance of synthesized from original sentences was measured. This distance was defined by the minimal number of block interchanges needed to transform the synthesized sentence to the original and its reverse. The average distances d of the sentences in each corpus from those in the original corpus are shown in Fig 5. The average distance of the OV-RelN-Post corpus was the greatest, although the validation loss was the lowest among the synthetic corpora. Remarkably, VO-RelN-Pre, which exhibits a low validation loss in Fig 3, was the corpus with the minimal distance to the original. Hence, the validation loss is dependent on, but not completely determined by, the distance from the original sentence or the reverse, suggesting that the validation loss can measure learnability.

Discussion

This paper examined whether neural language models can predict words in a text more easily for frequently observed word-order combinations. Specifically, instead of the corpora of an artificial language, corpora synthesized from a natural language corpus were used to ensure ethological validity. In addition, the corpora were synthesized by reordering the words without additions or deletions to ensure the fairness of the comparison. The validation losses for the language models were lowest for VO-NRel-Pre, the original text, and second lowest for OV-RelN-Post, which is frequently observed. The validation loss for OV-RelN-Post was greater than that for VO-NRel-Pre. This may have been because OV-RelN-Post is indeed more difficult to learn than VO-NRel-Pre, or because there were transformation errors. If OV-RelN-Post is difficult to learn in itself, this result is consistent with a previous study [18]. The difficulty of learning OV-RelN-Post may be explained by the greater dependency length of head-final languages such as Japanese, Korean, and Turkish [6]. This may be a reason for the fact that the syntactic change from SOV to SVO is more frequent than that from SVO to SOV [27]. If the higher validation loss for OV-RelN-Post is an artifact, this may be because the relative clause was not placed properly in the sequence of demonstrative, numeral, adjective, and noun. Placing prepositional phrases between articles and nouns, as in German, may improve the validation loss for OV-RelN-Post. Also, adding case markers to distinguish subjects from objects might improve the validation loss, as suggested in a previous study [18].

In any case, it is noteworthy that the validation losses for the corpora closer to the original were in most cases greater than that for OV-RelN-Post. This suggests that language models exhibit a preference analogous to language universals. In the present experiment, although OV-RelN-Post was slightly less easily learnable than VO-NRel-Pre, it was more easily learnable than the others. These results suggest that the frequent word-order combinations are more learnable than their partially modified combinations. Less frequent word-order combinations might be less easily learnable for language models because of their greater dependency length.

Along with other studies [16–18, 28], the present study indicates the effectiveness of the neural network modeling approach to language typology. This study examined the relationships among OV/VO, RelN/NRel, and Post/Pre word-order parameters. By contrast, Ravfogel *et al.* [18] examined the relationships between SVO/SOV/VOS/VSO/OSV/OVS and polypersonal agreement and case systems. To this end, they added suffices, whereas the present study did not. Thus, their study and the present study are complementary.

Together with previous studies, the present study indicates that some language universals originate from the common properties of the human brain and neural networks. Masked language models such as BERT and causal language models such as GPT-2 are quite successful in natural language processing and exhibit human-like generalization. However, these models do not have any special mechanism to realize human-like induction bias. Presumably, information processing systems with diverse internal structures can exhibit preferences similar to language universals as long as they have sufficient capacity.

However, this does not mean that all the typological features can be explained by efficient information processing systems, including neural language models. Because the SOV word order emerges in gesture-production experiments [29], embodied artificial intelligence might explain a wider range of language universals. In a different cognitive experiment, native speakers of left- and right-branching languages were found to recall initial and final stimuli better, respectively [30]. The same bias might be found for multimodal neural networks depending on the languages used in training. Recent advances in speech generation with neural network models may shed light on the relationship between the phonetic property and meaning of words.

The connection between the frequency of languages and their learnability by human subjects and neural language models has been criticized by cultural evolution and language typology itself. Rafferty *et al.* [31] called attention to the fact that easily learnable languages are not necessarily prevalent because there may be a vast number of less easily learnable languages. Indeed, the difference in validation loss may not be sufficiently large to prevent infrequent combinations from emerging. Thus, the present study does not preclude the cultural evolution explanation. However, among several options, such as the eight corpora examined in the present study, easily learnable language can be dominant. Although Dunn *et al.* [8] argued that the co-occurrence of language features is better fitted by evolutionary models, learnability as evolutionary fitness can play a role in language evolution. Studies using synthetic corpora can measure the fitness of a language with a combination of features. It has been suggested that word orders such as OV and VO are not categorical, but rather, gradient [32]. Further study is needed to examine the gradient variation in word order. To respond to these criticisms, it is worth exploring the variety of learnable and generatable languages using neural network models. More specifically, this could be examined by investigating whether a language-like structure emerges if a neural network needs to linearize the latent representation to a sequence of a small number of tokens. If the manner of linearization were similar to known human languages, learnability would be proven to be essential in the evolution of language. This would also be a test for the string-context mutual segmentation hypothesis on the origin of human language [33].

Acknowledgements

This work was supported by the Japan Society for the Promotion of Science KAKENHI under grant No. JP22K18526.

References

1. Greenberg JH. Some universals of grammar with particular reference to the order of meaningful elements. In: Universals of language. MIT Press; 1963. p. 73–113.

2. Dryer MS. SVO languages and the OV:VO typology. Journal of linguistics. 1991;27(2):443–482.

3. Dryer MS. The Greenbergian word order correlations. Language. 1992;68(1):81–138.

4. Dryer MS. The branching direction theory of word order correlations revisited. Universals of language today. 2009; p. 185–207.

5. Hawkins JA. Word order universals. No. 3 in Quantitative analyses of linguistic structure. Academic Press; 1983. Available from: `https://ci.nii.ac.jp/ncid/BA03206662`.

6. Futrell R, Mahowald K, Gibson E. Large-scale evidence of dependency length minimization in 37 languages. Proceedings of the National Academy of Sciences. 2015;112(33):10336–10341.

7. Maslova E. A dynamic approach to the verification of distributional universals. Linguistic Typology. 2000;4(3):307–333.

8. Dunn M, Greenhill SJ, Levinson SC, Gray RD. Evolved structure of language shows lineage-specific trends in word-order universals. Nature. 2011;473(7345):79–82.

9. Tabullo A, Arismendi M, Wainselboim A, Primero G, Vernis S, Segura E, et al. On the learnability of frequent and infrequent word orders: An artificial language learning study. Quarterly Journal of Experimental Psychology. 2012;65(9):1848–1863.

10. Martin A, Ratitamkul T, Abels K, Adger D, Culbertson J. Cross-linguistic evidence for cognitive universals in the noun phrase. Linguistics Vanguard. 2019;5(1):20180072.

11. Martin A, Holtz A, Abels K, Adger D, Culbertson J. Experimental evidence for the influence of structure and meaning on linear order in the noun phrase. Glossa: a Journal of General Linguistics. 2020;5(1):97. doi:https://doi.org/10.5334/gjgl.1085.

12. Davis E, Smith K. The learnability and emergence of dependency structures in an artificial language. Journal of Language Evolution. 2023;8(1):64–89.

13. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Burstein J, Doran C, Solorio T, editors. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers). Association for Computational Linguistics; 2019. p. 4171–4186. Available from: `https://doi.org/10.18653/v1/n19-1423`.

14. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I, et al. Language models are unsupervised multitask learners. OpenAI blog. 2019;1(8):9.

15. Everbroeck Ev. Language type frequency and learnability from a connectionist perspective. Linguistic Typology. 2003;7:1–50.

16. White JC, Cotterell R. Examining the Inductive Bias of Neural Language Models with Artificial Languages. In: Zong C, Xia F, Li W, Navigli R, editors. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Online: Association for Computational Linguistics; 2021. p. 454–463. Available from: `https://aclanthology.org/2021.acl-long.38`.

17. Kuribayashi T, Ueda R, Yoshida R, Oseki Y, Briscoe T, Baldwin T. Emergent Word Order Universals from Cognitively-Motivated Language Models. arXiv preprint arXiv:240212363. 2024;.

18. Ravfogel S, Goldberg Y, Linzen T. Studying the Inductive Biases of RNNs with Synthetic Variations of Natural Languages. In: Burstein J, Doran C, Solorio T, editors. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics; 2019. p. 3532–3542. Available from: `https://aclanthology.org/N19-1356`.

19. Tomlin RS. Basic Word Order (RLE Linguistics B: Grammar): Functional Principles. Routledge; 2014.

20. Fromkin V, Rodman R, Hyams N. An Introduction to Language. Wadsworth; 2018.

21. Dryer MS. On the order of demonstrative, numeral, adjective, and noun. Language. 2018; p. 798–833.

22. Pradhan SS, Hovy E, Marcus M, Palmer M, Ramshaw L, Weischedel R. OntoNotes: A Unified Relational Semantic Representation. In: International Conference on Semantic Computing (ICSC 2007); 2007. p. 517–526.

23. Kudo T, Richardson J. SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. In: Blanco E, Lu W, editors. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Brussels, Belgium: Association for Computational Linguistics; 2018. p. 66–71. Available from: `https://aclanthology.org/D18-2012`.

24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems. 2017;30.

25. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; 2020. p. 38–45.

26. Christie DA. Sorting permutations by block-interchanges. Information Processing Letters. 1996;60(4):165–169. doi:https://doi.org/10.1016/S0020-0190(96)00155-X.

27. Gell-Mann M, Ruhlen M. The origin and evolution of word order. Proceedings of the National Academy of Sciences. 2011;108(42):17290–17295. doi:10.1073/pnas.1113716108.

28. Wang D, Eisner J. The galactic dependencies treebanks: Getting more data by synthesizing new languages. Transactions of the Association for Computational Linguistics. 2016;4:491–505.

29. Langus A, Nespor M. Cognitive systems struggling for word order. Cognitive Psychology. 2010;60(4):291–318.

30. Amici F, Sánchez-Amaro A, Sebastián-Enesco C, Cacchione T, Allritz M, Salazar-Bonet J, et al. The word order of languages predicts native speakers' working memory. Scientific Reports. 2019;9(1):1124.

31. Rafferty AN, Griffiths TL, Ettlinger M. Greater learnability is not sufficient to produce cultural universals. Cognition. 2013;129(1):70–87. doi:https://doi.org/10.1016/j.cognition.2013.05.003.

32. Naranjo MG, Becker L. Quantitative word order typology with UD. In: Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018). vol. 155. Linköping University Electronic Press Linköping, Norway; 2018. p. 91–104.

33. Okanoya K, Merker B. Neural substrates for string-context mutual segmentation: A path to human language. In: Emergence of Communication and Language. Springer; 2007. p. 421–434.