Simulating Dependency Length Minimization using neural-network based learning and communication

Yuqing Zhang[1], Tessa Verhoef[2], Yuchen Lian[2], Gertjan van Noord[1], and Arianna Bisazza[1]

[*]Corresponding Author: yuqing.zhang@rug.nl
[1]Center for Language and Cognition, University of Groningen, Groningen, The Netherlands
[2]Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands

A fruitful approach to studying the influence of human cognitive biases and processes like language learning and use in shaping linguistic structure is to simulate them computationally (De Boer, 2006; Steels, 1997). Recent advances in machine learning and computational linguistics have yielded powerful (neural-network based) artificial learners that can deal surprisingly well with the complexity of human languages and can be used to set up increasingly realistic simulations (Chaabouni et al., 2021; Warstadt & Bowman, 2022). An important challenge in this line of work, however, is that such artificial learners still often behave differently from human learners (Chaabouni et al., 2019a; Galke et al., 2022).

Recently, Lian et al. (2023) proposed a novel framework for simulations of language learning and change with artificial languages and neural-network learners, which addresses some of these challenges. In the NeLLCom (Neural-agent Language Learning and Communication) framework, pairs of speaking and listening agents learn a pre-defined artificial language through supervised learning and then communicate with each other, optimizing a shared reward via reinforcement learning. Communication is simulated with a meaning reconstruction game where a speaker learns to convey a meaning m to a listener using the language it has learnt by supervised learning. This language use is tied to a shared goal: maximizing the communicative reward evaluated by the listener's prediction. Speakers are modeled as a linear-to-sequence structure whereas listeners work in the reverse direction, i.e., a sequence-to-linear structure. Sequential encoding/decoding is implemented by a recurrent neural network (RNN), specifically Gated Recurrent Units (Cho et al., 2014). In the original study, Lian et al. (2023) applied the framework to simulate the emergence of the word-order/case-marking trade-off and found that a human-like trade-off appears during communication without hard-coding specific biases in the agents.

In this work, we focus on another statistical language universal: dependency length minimization (DLM), a tendency to minimize the linear distance between heads and their dependents in natural languages (Futrell et al., 2020). Motivated by the contradictory patterns found in previous simulations (Chaabouni et al., 2019b; Zhao, 2022), we adopt NeLLCom to further investigate the minimal conditions that may lead to DLM in neural-network learners. Inspired by Fedzechkina et al. (2018)'s experiment with human learners, we expand Lian et al. (2023)'s original meaning space of agent-patient-action triplets by adding optional modifier phrases to agent and patient. Each modifier phrase consists of three items: adposition, adjective, and inanimate noun (e.g. *behind white door*). The meanings are descriptions of scenes that have only one long constituent (i.e. only subject or object has adpositional-phrase modification). Similar to NeLLCom's original setting, utterances are variable-length sequences of symbols taken from a fixed-size vocabulary: $u = [w_1, ..., w_l]$, $w_i \in V$. For each meaning, there are two possible utterance orderings (subject-object and object-subject). Ordering symbols representing short dependents closer to verbs leads to shorter total dependency length.

We train agents on a verb-initial language and a verb-final language, each comprising 50% long- and 50% short-dependency utterances. Additionally, we train on two control languages, each containing only short or long dependency utterances, respectively. We then conduct evaluations on meanings unseen during training. In this initial setup, speakers do not regularize towards reducing DL in production nor do the shorter-dependency languages show a learning advantage compared to their longer-dependency counterparts.

We then consider three additional factors to make the simulation more realistic: (i) introducing noise during listening (Futrell & Levy, 2017) through a word dropout technique (Gal & Ghahramani, 2016), (ii) modeling non-uniform word distributions and selectional preferences of verbs (McRae et al., 1998), i.e. strength of association of one action with one agent or patient, and (iii) testing listeners' incremental utterance processing (Kamide et al., 2003), or the extent to which an utterance's meaning can be guessed before hearing it entirely. We find that, during communication, neural learners tend to regularize towards one word order instead of one dependency length, failing to display a DLM preference in their productions. However, the proposed factors contribute to a small but consistent learning advantage of shorter dependencies for listening agents of the verb-initial language, but not for the verb-final language, which is consistent with patterns found in natural languages (Jing et al., 2022). Specifically, for the verb-initial language, we find that: 1) under noisy conditions, listeners learn the short-dependency language slightly better and faster than the long-dependency one; 2) the presence of noise affects the learning accuracy of uniform languages more severely than languages with skewed conditional word distributions; 3) the short-dependency language shows an advantage over the long one when evaluated incrementally, suggesting that the disambiguation of meanings related to local dependencies tends to occur earlier in the sentence.

Future directions include integrating incremental processing modeling into our use of the NeLLCom framework for investigating language emergence.

References

Chaabouni, R., Kharitonov, E., Dupoux, E., & Baroni, M. (2019a). Anti-efficient encoding in emergent communication. *Advances in Neural Information Processing Systems*, *32*, 6293–6303.

Chaabouni, R., Kharitonov, E., Dupoux, E., & Baroni, M. (2021). Communicating artificial neural networks develop efficient color-naming systems. *Proceedings of the National Academy of Sciences*, *118*(12), e2016569118.

Chaabouni, R., Kharitonov, E., Lazaric, A., Dupoux, E., & Baroni, M. (2019b). Word-order biases in deep-agent emergent communication. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 5166–5175). Florence, Italy: Association for Computational Linguistics.

Cho, K., Merriënboer, B. van, Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1724–1734). Doha, Qatar: Association for Computational Linguistics.

De Boer, B. (2006). Computer modelling as a tool for understanding language evolution. In *Evolutionary epistemology, language and culture: A non-adaptationist, systems theoretical approach* (pp. 381–406). Springer.

Fedzechkina, M., Chu, B., & Florian Jaeger, T. (2018). Human information processing shapes language change. *Psychological science*, *29*(1), 72–82.

Futrell, R., & Levy, R. (2017). Noisy-context surprisal as a human sentence processing cost model. In *Proceedings of the 15th conference of the european chapter of the association for computational linguistics: Volume 1, long papers* (pp. 688–698). Valencia, Spain.

Futrell, R., Levy, R. P., & Gibson, E. (2020). Dependency locality as an explanatory principle for word order. *Language*, *96*(2), 371–412.

Gal, Y., & Ghahramani, Z. (2016). A theoretically grounded application of dropout in recurrent neural networks. *Advances in neural information processing systems*, *29*.

Galke, L., Ram, Y., & Raviv, L. (2022). Emergent communication for understanding human language evolution: What's missing? In *Emergent communication workshop at iclr 2022.* virtual.

Jing, Y., Blasi, D. E., & Bickel, B. (2022). Dependency-length minimization and its limits: A possible role for a probabilistic version of the final-over-final condition. *Language*, *98*(3).

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, *49*(1), 133–156.

Lian, Y., Bisazza, A., & Verhoef, T. (2023). Communication Drives the Emergence of Language Universals in Neural Agents: Evidence from the Word-order/Case-marking Trade-off. *Transactions of the Association for Computational Linguistics*, *11*, 1033-1047.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283–312.

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of communication*, *1*(1), 1–34.

Warstadt, A., & Bowman, S. R. (2022). What artificial neural networks can tell us about human language acquisition. *Algebraic Structures in Natural Language*, 17–60.

Zhao, Y. (2022). *How to probe linguistic knowledge and bias.* Unpublished doctoral dissertation, The University of Arizona.