Emergence of linguistic universals in neural agents via artificial language learning and communication

Lian, Y.

Chapter 3

A New Framework for Neural-Agent Artificial Language Learning and Communication: NeLLCom

Artificial learners often behave differently from human learners in the context of neural agent-based simulations of language emergence and change. A common explanation is the lack of appropriate cognitive biases in these learners, which we have explored in **Chapter 2**. Besides human-like cognitive biases, it has also been proposed that more naturalistic settings of language learning and use could lead to more human-like results (Mordatch and Abbeel, 2018; Lazaridou and Baroni, 2020; Kouwenhoven et al., 2022; Galke et al., 2022). In this chapter we will explore such settings. Concretely, we ask:

RQ-B Does a human-like word-order/case-marking trade-off emerge in communicative neural agents?

We continue our investigation with the same test case as in **Chapter 2**, namely the word-order/case-marking trade-off, a widely attested language universal that has proven particularly hard to simulate. As shown in the previous chapter, the three factors we initially introduced were not sufficient to consistently lead to a human-like linguistic system exhibiting the trade-off. These findings suggested the training objective needs to be balanced with a measure of

communicative success. To this end, we propose a new Neural-agent Language Learning and Communication framework (NeLLCom), where the artificial language learning paradigm is adopted from human experiments. Within the NeLLCom framework, pairs of speaking and listening agents first learn a miniature language via supervised learning, and then optimize it for communication via reinforcement learning. We succeed in replicating the trade-off with the new framework without hard-coding specific biases in the agents.

Chapter adapted from:
Yuchen Lian, Arianna Bisazza, and Tessa Verhoef. 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

3.1 Introduction

The success of deep learning methods for natural language processing has triggered a renewed interest in agent-based computational modeling of language emergence and evolution processes (Lazaridou and Baroni, 2020; Chaabouni et al., 2022). An important challenge in this line of work, however, is that such artificial learners often behave differently from human learners (Galke et al., 2022; Rita et al., 2022; Chaabouni et al., 2019a).

One of the proposed explanations for these mismatches is the difference in cognitive biases between human and neural-network (NN) based learners. For instance, the neural-agent iterated learning simulations of Chaabouni et al. (2019b) and **Chapter 2** (Lian et al., 2021) did not succeed in replicating the trade-off between word-order and case marking, which is widely attested in human languages (Sinnemäki, 2008; Futrell et al., 2015) and has also been observed in miniature language learning experiments with human subjects (Fedzechkina et al., 2017). Instead, those simulations resulted in the preservation of languages with redundant coding mechanisms, which the authors mainly attributed to the lack of a human-like least-effort bias in the neural agents.

Besides human-like cognitive biases, it has been proposed that more natural settings of language learning and use could lead to more human-like patterns of language emergence and change (Mordatch and Abbeel, 2018; Lazaridou and Baroni, 2020; Kouwenhoven et al., 2022; Galke et al., 2022). In this work, we follow up on this second account and investigate whether neural agents that *strive to be understood* by other agents display more human-like language preferences.

To achieve that, we design a Neural-agent Language Learning and Communication (NeLLCom) framework that combines Supervised Learning (SL) with Reinforcement Learning (RL), inspired by Lazaridou et al. (2020) and Lowe et al. (2020). Specifically, we use SL to teach our agents predefined languages characterized by different levels of word order freedom and case marking. Then, we employ RL to let pairs of speaking and listening agents talk to each other while optimizing communication success (also known as *self-play* in the emergent communication literature).

We closely compare the results of our simulation to those of an experiment with a very similar setup and miniature languages involving human learners (Fedzechkina et al., 2017), and show that a human-like trade-off can indeed appear during neural-agent communication. Although some of our results differ from those of the human experiments, we make an important contribution towards developing a neural-agent framework that can replicate language universals without the need to hard-code any ad-hoc bias in the agents. We release the NeLLCom framework[1] to facilitate future work simulating the emergence of different language universals.

3.2 Background

3.2.1 Word order *vs.* case marking trade-off

A research focus of linguistic typology is to identify *language universals* (Greenberg, 1963), i.e. patterns occurring systematically among the large diversity of

[1] All code and data are available at `https://github.com/Yuchen-Lian/NeLLCom`

natural languages. The origins of such universals are object of long-standing debates. The trade-off between word order and case marking is an important and well-known example of such a pattern that has been widely attested (Comrie, 1989; Blake, 2001). Specifically, languages with more flexible constituent order tend to have rich morphological case systems (e.g. Russian, Tamil, Turkish), while languages with more fixed order tend to have little or no case marking (e.g. English or Chinese). Additionally, quantitative measures also revealed that the functional use of word order has a statistically significant inverse correlation with the presence of morphological cases based on typological data (Sinnemäki, 2008; Futrell et al., 2015).

Various experiments with human participants (Fedzechkina et al., 2012, 2017; Tal and Arnon, 2022) were conducted to reveal the underlying cause of this correlation. In particular, Fedzechkina et al. (2017), who highly inspired this work, applied a miniature language learning approach to study whether the trade-off could be explained by a human learning bias to reduce production effort while remaining informative. In their experiment, two groups of 20 participants were asked to learn one of two predefined miniature languages. Both languages contained optional markers but differed in terms of word order (fixed *vs.* flexible). After three days of training, both groups reproduced the initial word order distribution, however the flexible-order language learners used case marking significantly more often than the fixed-order language learners. Moreover, an asymmetric marker-using strategy was found in the flexible-order language learners, whereby markers tended to be used more often in combination with the less frequent language. Thus, most participants displayed an inverse correlation between the use of constituent order and case marking *during* language learning, which the authors attributed to a unifying information-theoretic principle of balancing effort with robust information transmission.

3.2.2 Agent-based simulations of language evolution

Computational models have been used widely to study the origins of language structure (Kirby, 2001; Van Everbroeck, 2003; De Boer, 2006; Steels, 2016).

In particular, Lupyan and Christiansen (2002) were able to mimic the human acquisition patterns of four languages with very different word order and case marking properties, using a simple recurrent network (Elman, 1990).

Modern deep learning methods have also been used to simulate patterns of language emergence and change (Chaabouni et al., 2019a,b, 2020, 2021; Lian et al., 2021; Lazaridou et al., 2018; Ren et al., 2020). Despite several interesting results, many report the emergence of languages and patterns that significantly differ from human ones. For example, Chaabouni et al. (2019a) found an anti-efficient encoding scheme that surprisingly opposes Zipf's Law, a fundamental feature of human language. Rita et al. (2020) obtained a more efficient encoding by explicitly imposing a length penalty on speakers and pushing listeners to guess the intended meaning as early as possible. Focusing on the word-order/case-marking trade-off, Chaabouni et al. (2019b) implemented an iterated learning setup inspired by Kirby et al. (2014) where agents acquire a language through SL, and then transmit it to a new learner, iterating over multiple generations. The trade-off did not appear in their simulations. Lian et al. (2021) (**Chapter 2**) extended the study by introducing several crucial factors from the language evolution field (e.g. input variability, learning bottleneck), but no clear trade-off was found. To our knowledge, no study with neural agents has successfully replicated the emergence of this trade-off so far.

3.3 NeLLCom: Language Learning and Communication Framework

This section introduces the Neural-agent Language Learning and Communication (**NeLLCom**) framework, which we make publicly available. Our goal differs from that of most work in emergent communication, where language-like protocols are expected to **arise from sets of random symbols** through interaction (Havrylov and Titov, 2017; Bouchacourt and Baroni, 2018; Lazaridou et al., 2018; Chaabouni et al., 2019a, 2022). We are instead interested in observing how **a given language with specific properties changes** as the result of learning and use. Specifically, in this work, agents need to learn

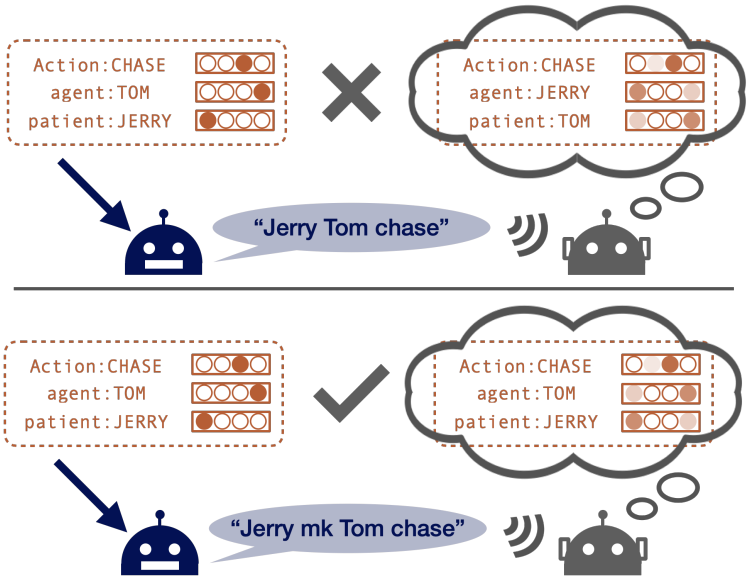


Figure 3.1: A high-level overview of the meaning reconstruction game.

miniature languages with varying word order distributions and case marking rules. While this can be achieved by a standard SL procedure, we hypothesize that **human-like regularization patterns** will only appear when our agents strive to be understood by other agents. We simulate such a need via RL, using a measure of communication success as the optimization objective.

Similar SL+RL paradigms have been used in the context of communicative AI (Li et al., 2016; Strub et al., 2017; Das et al., 2017). In particular, Lazaridou et al. (2020) and Lowe et al. (2020) explore different ways of combining SL and RL to teach agents to communicate with humans in natural language. A well-known problem in that setup is that languages tend to *drift* away from their original form as agents adapt to communication. In our context, we are specifically interested in studying how this drift compares to human experiments of artificial language learning. Our implementation is partly based on the EGG toolkit[2] (Kharitonov et al., 2019).

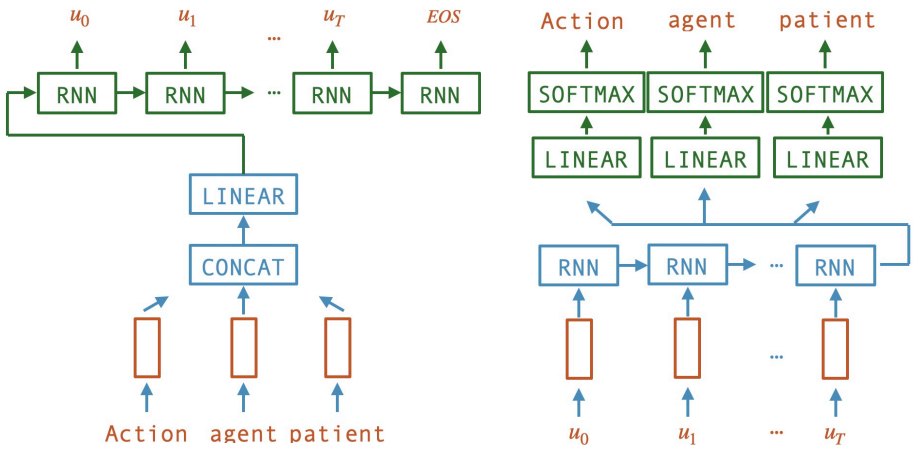
[2] https://github.com/facebookresearch/EGG

3.3.1 The Task

NeLLCom agents communicate about a simplified world using pre-defined artificial languages (see Figure 3.1). Speaking agents convey a meaning m by generating an utterance u, whereas listening agents try to map an utterance u to its respective meaning m. The meaning space includes agent-patient-action triplets, such as *dog-cat-follow, dog-mouse-follow*, defined as triplets $m = \{A, a, p\}$, where A is an action, a the agent, and p the patient. Utterances are variable-length sequences of symbols taken from a fixed-size vocabulary: $u = [w^1, ..., w^I]$, $w^i \in V$. Evaluation is conducted on meanings unseen during training.

3.3.2 Agent Architectures

Both speaking and listening agents contain an encoder and a decoder, however their architectures are mirrored as the meanings and sentences are represented differently (see Fig. 3.2).



(a) Speaking Agent **(b)** Listening Agent

Figure 3.2: Agents architecture.

Speaker: linear-to-sequence

In a speaker network (S), the encoder receives the hot-vector representations of A, a, and p, and projects them to latent representations or embeddings. The order of these three elements is irrelevant. The concatenation of the embeddings followed by a linear layer becomes the latent meaning representation,[3] based on which the Recurrent Neural Network (RNN) decoder generates a sequence of symbols.[4]

Listener: sequence-to-linear

The listener network (L) works in the reverse way: its RNN encoder takes an utterance as input and sends its encoded representation to the decoder, which tries to predict the corresponding meaning. Specifically, the final RNN cell is fed to the decoder, which passes it through three parallel linear layers, for A, a, and p, respectively. Finally, each of the three elements is generated by a softmax layer.

Unlike the agents of Chaabouni et al. (2019b) and **Chapter 2** (Lian et al., 2021), our agents can only behave as either speaker or listener, but not both. Chaabouni et al. (2019b) achieved this by tying input and output embeddings, however they reported only a minor effect on the results. As another difference, we represent meanings as unordered attribute-values instead of sequences, which we find important to avoid any ordering bias in the meaning representation. We note that the framework is rather general: in future studies, it could be adapted to different meaning spaces and different artificial languages, as well as different types of neural sequence encoders/decoders.

[3] We opted for simple architectural choices whenever possible. Adding a non-linearity to the meaning encoder did not affect the results.

[4] We implement the speaker's decoder and the listener's encoder as one-layer Gated Recurrent Units (GRU) (Chung et al., 2014) following previous work on language emergence (Chaabouni et al., 2020; Dessì et al., 2019). The latter paper, in particular, reports slower convergence with LSTM than GRU, and a lack of success at adapting Transformers to their setup.

3.3.3 Supervised Language Learning

SL is a natural choice to teach agents a specific language. This procedure requires a dataset D of meaning-utterance pairs $\langle m, u \rangle$ where u is the gold-standard generated for m by a predefined grammar (see grammar details in Section. 3.4.1). The learning objectives differ between speaker and listener agents.

Speaker

Given D, speaker's parameters θ_S are optimized by minimizing the cross-entropy loss:

$$Loss_{(S)}^{sup} = -\sum_{i=1}^{I} \log p_{\theta_S}(w^i | w^{<i}, m) \tag{3.1}$$

where w^i is the i^{th} word of the gold-standard utterance u. Notice that SL implies a teacher forcing procedure (Goodfellow et al., 2016), meaning that at each timestep the gold history $w^{<i}$ is used to predict the next word w^i and update the network weights accordingly.

Listener

Given D, listener's parameters θ_L are optimized by minimizing the cross-entropy loss:

$$Loss_{(L)}^{sup} = -(\log p_{\theta_L}(a|u) + \log p_{\theta_L}(p|u) + \log p_{\theta_L}(A|u)) \tag{3.2}$$

3.3.4 Optimizing Communication Success

While SL may be sufficient to (perfectly) learn a given meaning-to-signal mapping and vice versa, we are interested in whether and how such language changes as a result of repeated usage. Following a long-established practice of simulating emergent communication with humans and computer agents in language evolution (Steels, 1997, 2016; Selten and Warglien, 2007; Galantucci and Garrod, 2011), and more recently also in the computational linguistics literature (Bouchacourt and Baroni, 2018; Lazaridou et al., 2018, 2020; Lowe

et al., 2020; Havrylov and Titov, 2017; Evtimova et al., 2018), we simulate communication with a meaning reconstruction game where a speaker S learns to convey meanings m to a listener L using utterances $\hat{u}$ in the language it has learned by SL. The goal for both agents is to maximize a shared reward evaluated by the listener's prediction. For this phase, we adopt the classical policy-based algorithm REINFORCE (Williams, 1992). Specifically, we optimize:

$$Loss_{(S,L)}^{comm} = -r^L(m, \hat{u}) * \sum_{i=1}^{I} \log p_{\theta_S}(w^i | w^{<i}, m) \tag{3.3}$$

where $r^L(m, \hat{u})$ is defined as the cross-entropy loss between input meaning m and listener's prediction:

$$r^L(m, \hat{u}) = \sum_{e \in m=\{a,p,A\}} \log p_{\theta_L}(e|\hat{u}) \tag{3.4}$$

3.3.5 Combining Supervision and Communication

We adopt the simplest possible way of combining SL and RL, which is to first train the agents by SL until convergence and then continue training them by RL to maximize the communicative reward.[5] While more sophisticated combination techniques were proposed recently (Lowe et al., 2020; Lazaridou et al., 2020), we find this simple SL+RL sequence to work well in our context, and leave an exploration of other techniques to future work.

Crucially, using communication success as task reward rather than forcing agents to imitate given training pairs $\langle m, u \rangle$ allows agents to depart from the initially learnt grammar, as long as the new language remains understandable by other agents. This principle is well studied in the framework of Rational Speech Act (RSA) (Goodman and Frank, 2016) which implemented utterance understanding from a social cognition aspect. If a language is suboptimal for an agent, e.g. in terms of efficiency or ambiguity, we expect it to change throughout multiple communication rounds. Note that the listener's role can also be

[5] This procedure corresponds to *reward fine-tuning* in Lazaridou et al. (2020) and to *sup2sp* in Lowe et al. (2020).

interpreted as that of a *speaker-internal* monitoring system that predicts the chance of a message to be understood by a listener *before* uttering it (Ferreira, 2019).

3.3.6 Evaluation

Accuracy

During evaluation, both types of agents generate their predictions by greedy decoding. Accuracy is computed at the whole utterance or meaning level. Specifically, **listening accuracy** is 1 if all of A, a, and p are correct, otherwise it is 0. Speaking accuracy is evaluated in two ways: (i) Regular **speaking accuracy** is 1 only if the generated utterance is identical to the one in the dataset. (ii) **'Permissive' speaking accuracy** considers the fact that our grammars admit multiple utterances for the same meaning: for each test sample, we generate all correct candidates (i.e., with or without marker; OSV and SOV for the flexible-order language). Permissive accuracy is 1 if the generated utterance matches any of the candidates. As long as the utterance is acceptable, matching an arbitrary choice of order or marking for a given meaning does not matter. Hence, the discussion in this section is based on permissive speaking accuracy.

Utterance Length and Production Preferences

In principle, RNN can generate sequences of variable length. In practice, this is achieved by fixing a maximum message length (10 words in our setup) and truncating the sequence when the first symbol ⟨EOS⟩ is generated. We noticed, however, that during communication our speaking agents do not always end their message with ⟨EOS⟩, but rather duplicate their final words to fill up the maximum utterance length after generating a well-formed initial message. As long as the first part of the utterance perfectly matches one of the structures admitted by the grammar, we truncate the utterance at the last word before duplication. On average, this affects 15% of the utterances by epoch 60.

Speaker-generated utterances for the unseen meanings (240 in total) are then

classified into five types: SOV without marker, SOV with marker, OSV without marker, OSV with marker and uncategorized (other). Computed properties:

$$\%\text{SOV} = (\text{SOV}_{\text{mk}} + \text{SOV}_{\text{no\_mk}})/\text{Total} \tag{3.5}$$

$$\%\text{OSV} = (\text{OSV}_{\text{mk}} + \text{OSV}_{\text{no\_mk}})/\text{Total} \tag{3.6}$$

$$\%\text{with\_mk} = (\text{SOV}_{\text{mk}} + \text{OSV}_{\text{mk}})/\text{Total} \tag{3.7}$$

$$\%\text{no\_mk} = (\text{SOV}_{\text{no\_mk}} + \text{OSV}_{\text{no\_mk}})/\text{Total} \tag{3.8}$$

Uncertainty Measure

This measure taken from Fedzechkina et al. (2017) captures the uncertainty about the role of the two entities expressed in an utterance, which is experienced by a listener with perfect knowledge of the grammar. It is formalized as the conditional entropy of grammatical function assignment (GF) given sentence form (s.form):

$$H(\text{GF}|\text{s.form}) = -\sum_{\text{GFs}} \sum_{\text{s.forms}} p(\text{s.form}, \text{GF}) * log_2 p(\text{GF}|\text{s.form}) \tag{3.9}$$

According to the constraints of each grammar, possible sentence forms are

$$\text{s.forms} = \{\text{SOV}, \text{OSV}\}$$

and function assignments

$$\text{GFs} = \{\text{N}_1\text{N}_2\text{V},\ \text{N}_1\text{mkN}_2\text{V},\ \text{N}_1\text{N}_2\text{mkV}\}$$

Initial language uncertainties are as in Fedzechkina et al. (2017): 0 for fix+op and 0.33 for flex+op.

3.4 Experimental Setup

We use NeLLCom to replicate the results of Fedzechkina et al. (2017), who taught human subjects miniature languages with varying order distributions.

language	word order	case marking	candi. utterance
fix+op	100% SOV	66.7% on OBJ	*Tom Jerry chase* *Tom Jerry mk chase*
flex+op	50% SOV, 50% OSV	66.7% on OBJ	*Tom Jerry chase* *Tom Jerry mk chase* *Jerry Tom chase* *Jerry mk Tom chase*

Table 3.1: The two miniature grammars used in this study, along with meaning-utterance $\langle m, u \rangle$ examples.

Subjects watched short videos of two actors performing simple transitive events (e.g. *a chef hugging a referee*) accompanied by spoken descriptions in the novel language.[6] We adopt the same setup, with two notable differences: (i) our agents do not take videos or images as input, but triplets of symbols representing agent, patient and action, respectively (see Section. 3.3); (ii) descriptions are not spoken but written, and words are represented by dummy strings (such as *noun-1*, *verb-2*, etc.) instead of English-like sounding nonce words. Thus, we abstract away from the problem of (i) mapping visual input to structured meaning representations and (ii) mapping continuous audio signals to discrete word representations, respectively. Dealing with these interfaces is necessary when working with humans, but not with neural agents. Moreover, none of them are a core aspect of our investigation.

3.4.1 Miniature Languages

Following Fedzechkina et al. (2017), we consider two head-final languages: one with fixed order and optional case markers (**fix+op**), and one with flexible order and optional case markers (**flex+op**). Optional marking means that 2/3 of all objects are followed by a special mark (the token *mk*), whereas subjects are never marked. Possible constituent orders are SOV and OSV: the fixed-order language uses always SOV, while the flexible-order one uses both with a

[6] Sentence learning was preceded by a noun learning phase which we do not model in our experiments. For more details on the human training process, see Fedzechkina et al. (2017).

probability of 50-50%. The two languages are illustrated in Table 3.1.

In fix+op, order is informative and sufficient to disambiguate grammatical functions. Case marking is therefore a redundant cue. In flex+op, order is uninformative therefore marking –when present– is important to recover the meaning. The hypothesis that language learning and use create biases towards efficient communication systems (Gibson et al., 2019; Fedzechkina et al., 2012) yields two predictions: fix+op is expected to become less redundant (by a decrease of case marking) whereas flex+op should become more predictable (by an increase of marking *or* a more consistent order).

3.4.2 Meaning Space

The meaning space used by Fedzechkina et al. (2017) included 6 entities and 4 actions, resulting in a total of $6\times(6-1)\times4=120$ possible meanings (an entity cannot be agent and patient at the same time). While suitable for human learners, such a space is too small to train neural agents (Zhao et al., 2018; Chaabouni et al., 2020). In preliminary experiments, we found that our learners converge well with a meaning space size of 720 (10 performers and 8 actions in our languages, resulting in a total of $10\times(10-1)\times8=720$ possible meanings).

To test the agents' ability to convey new meanings, we split our dataset into 66.7% training and 33.3% testing. We also ensure that each entity and action of the meaning space appears at least once in the training set. To prevent the agents from memorizing spurious correlations between a meaning and a particular order or marking choice, we regenerate a new utterance per meaning (according to the same grammar) after each epoch of SL.

3.4.3 Datasets

Each word in a language corresponds uniquely to an entity or an action in the meaning space, leading to vocabulary size $|V|= 8+10+1$(marker) $=19$. After the train/test split, we check that each entity and each action in the test set appears at least once in the training. If that's not the case, we randomly swap the meaning-utterance pairs containing unseen entities with random ones from

the training set. An end-of-sentence ⟨EOS⟩ token is appended to each utterance and padding is used to deal with variable utterance lengths.

3.4.4 Model training

Hyper-parameters were set in preliminary SL experiments: Speakers have 8-dim. embeddings and a 128-dim. GRU layer. Listeners have 32-dim. embeddings and a 32-dim. GRU layer. A default Adam optimizer (Kingma and Ba, 2015) in PyTorch (Paszke et al., 2017) is used for both SL and RL, with learning rate 0.01 and batch size 32. Each training phase lasts 60 epochs and we repeat each experiment with 20 different random seeds.

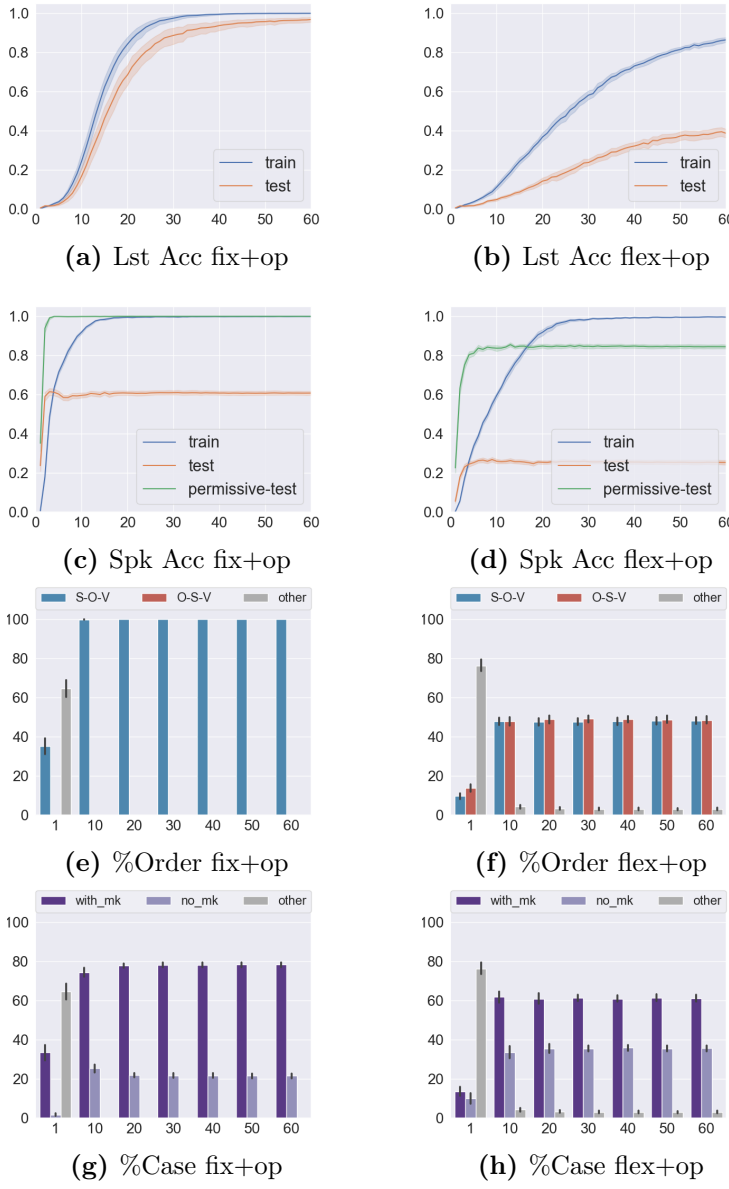
3.5 Supervised Learning Results

We start by evaluating the agents' ability to learn to speak or listen in a fully supervised way, that is, using the generated meaning-utterance pairs from a specific language as labeled data.

3.5.1 Accuracy

Fig. 3.3 shows accuracy results for both agent types, each averaged over 20 random initialization seeds. We find that our agents learn to speak and understand the fixed-order language with extremely high accuracies (Fig. 3.3a, 3.3c). By contrast, the flexible-order language reaches only 38.7% listening accuracy (Fig. 3.3b) and 84.5% permissive speaking accuracy (Fig. 3.3d) on average for the unseen test. Note this does not reflect a weakness of the learners, but the ambiguity of the language itself: namely, subject and object are not distinguishable when the marker is absent, which happens in a third of the utterances.[7] These results are consistent with the higher comprehension and production accuracy of human participants learning the fix+op vs. flex+op language in Fedzechkina et al. (2017). Specifically, their flex+op

[7]The ability of RNNs to learn fixed-order languages equally well as their flexible-order/case-marking has been demonstrated by previous studies (Lupyan and Christiansen, 2002; Chaabouni et al., 2019b; Bisazza et al., 2021), but only when case marking is consistently present.



(a) Lst Acc fix+op **(b)** Lst Acc flex+op

(c) Spk Acc fix+op **(d)** Spk Acc flex+op

(e) %Order fix+op **(f)** %Order flex+op

(g) %Case fix+op **(h)** %Case flex+op

Figure 3.3: Supervised learning results across training epochs for the fixed- (left) and flexible-order (right) language: accuracy of listening (a,b) and speaking (c,d) agents; distribution of word order (e,f) and markers (g,h) in speaker-generated utterances. All results are averaged over 20 random seeds.

group reached 96% comprehension accuracy with 6.2% grammatical mistakes, while the fix+op group reached 99% accuracy with no grammatical mistakes (see Section 3.1 in Fedzechkina et al. (2017)). Next, we inspect the properties of the language generated by speaking agents during the learning process.

3.5.2 Production Preferences

Fig. 3.3e, 3.3f show the proportion of SOV vs. OSV test utterances generated by the speaking agents across training epochs. For both languages, learners show a clear **probability-matching behavior**: in a few epochs, the order distribution becomes the same as in the input language and remains unchanged throughout the whole training. A similar pattern is visible for marking (Fig. 3.3g, 3.3h). Looking closer at fix+op (Fig. 3.3g) we notice a slightly higher production of cases than the initial 66.7%, which is even less efficient than the input language.

Taken together, these results show that our agents are good learners but do not regularize the use of the two strategies in a human-like way after SL, which is in line with the iterated supervised learning results of Chaabouni et al. (2019b) and **Chapter 2** (Lian et al., 2021). This leads us to the next phase: optimizing agents for communicative success.

3.6 Communication Learning Results

We study the effect of communication learning on communication success and language properties.

3.6.1 Communication Success

Once a pair of agents is trained to speak/listen, they start communicating with each other to achieve a shared goal: the listener should understand the speaker, i.e. reconstruct the intended meaning. Task success is evaluated by **meaning reconstruction accuracy**, which corresponds to the listening accuracy

(Section. 3.5.1) of a listener receiving a speaker-generated utterance as input.[8]

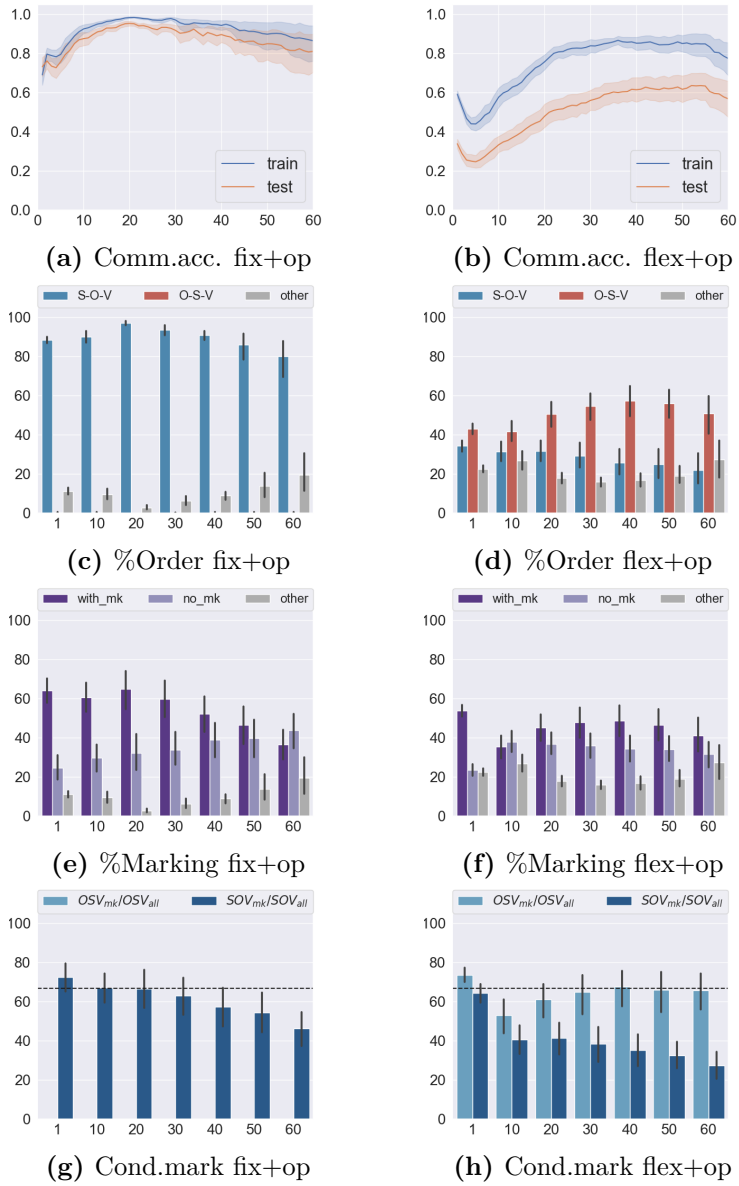
The results in Fig. 3.4a, 3.4b show that agents understand each other better after several communication rounds. More specifically, the non-ambiguous language (Fig. 3.4a) suffers from an initial drop but recovers the initial accuracy by epoch 20. The ambiguous language (Fig. 3.4b) starts from a lower communication success rate as expected but becomes more and more informative throughout communication. In particular, around epoch 40, agents recover the communication success they had achieved at the end of SL on known meanings (85.2%) while even *exceeding* it for new meanings (61.5% vs. 38.7%). These results strongly suggest the language becomes less ambiguous by interaction.

Additionally, we report a noticeable drop in average performance towards the last epochs. The individual seed results reveal that most agent pairs suffer from a collapse of their communication protocol in the final stages of RL. We attribute this issue to a known limitation of the REINFORCE algorithm related to its high gradient variance (Lu et al., 2020). Having assessed that our NeLLCom agents are able to learn a language and use it for conveying meanings, we now inspect *how* their language changes during communication.

3.6.2 Production Preferences

The proportions of word order and case markers generated by the speaking agents are shown respectively in Fig. 3.4c, 3.4d and Fig. 3.4e, 3.4f. We can see that these properties change considerably during communication learning, which was not the case during SL. The increase of communication success observed in both languages already indicates that languages tend to become more informative. The key question is whether informativity is being balanced with efficiency, in a similar way as observed in human experiments (Fedzechkina et al., 2017).

[8] Greedy decoding is used for both speaker and listener during the evaluation of communication success.



(a) Comm.acc. fix+op **(b)** Comm.acc. flex+op

(c) %Order fix+op **(d)** %Order flex+op

(e) %Marking fix+op **(f)** %Marking flex+op

(g) Cond.mark fix+op **(h)** Cond.mark flex+op

Figure 3.4: Communication learning results across training epochs for the fixed- (left) and flexible-order (right) language: meaning reconstruction accuracy (a,b); distribution of order (c,d) and markers (e,f) in speaker-generated utterances; marking conditioned on different orders (g,h). Dashed lines indicate marking in the initial dataset (66.7%). All results averaged over 20 random seeds.

Fix+op

This language is redundant as it uses both fixed order (SOV) and marking to convey argument roles. As shown in Fig. 3.4c, agents keep using SOV throughout the communication process.[9] Similarly, human experiments of language emergence have shown that participants hardly ever create innovations in languages that are already systematic (St. Clair et al., 2009; Tily et al., 2011; Fedzechkina et al., 2017). Importantly, Fig. 3.4e reveals a clear preference towards dropping markers, as evidenced by a steady increase of *no_mk* utterances (light color). This aligns with the finding in Fedzechkina et al. (2017), whereby human learners of the fixed-order language significantly reduced the use of marking over three days of training.[10] The tendency to drop case markers is often explained by a human preference for reducing redundancy and increasing efficiency. Notably, the agents in our framework did not have any manually coded efficiency bias. The maximum allowed message length was much longer than the utterances needed to get the message across and the agents were not incentivized in any way to produce shorter sentences. Thus, we explain the observed pattern as a tendency of the neural agents to make the language more systematic as long as this does not harm communicative success.

Flex+op

Recall this language is originally as efficient as fix+op (i.e. same average utterance length) but less informative due to the presence of ambiguous utterances. We can think of at least two ways in which human or human-like learners could improve it, namely: (i) keep using both orders interchangeably but use markers more systematically, or (ii) choose one order as dominant and keep using markers optionally (or not at all). Note that different pairs of speaking/listening agents may opt for different, though equally optimal strategies.

[9] The slight drop of SOV in the last epochs is due to the increase of non-classifiable (other) utterances, in turn related to the final communication collapse mentioned in Section. 3.6.1.

[10] For detailed case marking results in human production, see Section 3.3 and Fig.4 in Fedzechkina et al. (2017).

We find that NeLLCom agents increasingly produce OSV utterances (Fig. 3.4d), reaching a situation where OSV is twice as common as SOV when communication success is at its highest (epoch ∼50). At the same time, marker use fluctuates initially and then stabilizes around 55%, that is still the majority of cases but less than the initial rate (66.7%). This strongly suggests that agents are making the language more informative while reducing effort, according to strategy (ii). These results do not fully match those of Fedzechkina et al. (2017), where most subjects instead adopted strategy (i).[11] Nonetheless, our findings provide important evidence that the word-order/case-marking trade-off can emerge in neural learners without hard-coded biases.

Conditional case marking

Besides *how many* markers are used, it is important to understand *how* they are used. As Fedzechkina et al. (2017) point out, learners of a flexible-order language could reduce uncertainty by conditioning their marker use on word order (asymmetric case marking). For instance, using object marking only in SOV utterances could minimize uncertainty while maximizing efficiency. As discussed above, NeLLCom agents using flex+op tend to prefer an order over the other, however they are far from using one exclusively. Could our agents also be using markers conditionally? Fig. 3.4h shows the proportion of OSV utterances having a marker out of all OSV's (OSV_{mk}/OSV_{all}) and the same for SOV's.[12] Indeed, agents use marking decreasingly when producing SOV utterances but maintain the marker percentage in OSV utterances, which matches unexpectedly well the human tendency observed by Fedzechkina et al. (2017), Section 3.4. Whether this is due to a coincidence or to a bias (e.g. towards marking the first entity appearing in an utterance) remains for now unexplained.

[11] For detailed word order results in human production, see Section 3.2 and Fig.3 in Fedzechkina et al. (2017).

[12] For completeness, Fig. 3.4g shows conditional case marking results for fixed+op, however this is less interesting as word order is a sufficient disambiguation cue in this language.

3.7 Individual Learners' Trajectories

All results so far were averaged over multiple randomly initialized agents. Here, we look at possible variations among pairs of speaking-listening agents. We focus only on the flexible-order language, as it is more likely to undergo different optimization strategies. Fig. 3.5 shows 20 production distributions, each corresponding to a different random seed. Most agents (no. 1 to 14) regularize their productions towards the OSV order, as anticipated by the average results in Fig. 3.4. However, we also find two agent pairs that take the opposite path and produce more SOV (no. 19 and 20). The remaining four agents show no clear order preferences (no. 15, 16, 17 and 18). As for case marking, a clear preference to drop the marker from SOV utterances can be found in 15/20 pairs (no. 4, 5, 9, 10 16 are exceptions), which reflects the average trend of conditional case marking shown in Fig. 3.4h. This high degree of between-agents variability matches human results (Fedzechkina et al., 2012, 2017; Culbertson et al., 2012; Hudson Kam and Newport, 2005) where learners often adopt different strategies to reach a common optimization objective.

Uncertainty/efficiency trade-off

We explore whether the observed trajectories can be explained by a single principle: a trade-off between uncertainty and efficiency. Following Fedzechkina et al. (2017), we quantify **production effort** as the average number of words per generated utterance.[13] To quantify **uncertainty**, we use their "conditional entropy over grammatical function assignment" (H), which captures the uncertainty over the intended meaning experienced by a listener with perfect knowledge of the initial grammar. Fig. 3.6 presents uncertainty versus production effort at three time points: the initial language defined by the grammar, production after SL, and production after communication. For comparison, the human results of Fedzechkina et al. (2017) are reported in Fig. 3.6c.

In Fig. 3.6a, the tight distribution of data points (empty circles) around the

[13] Fedzechkina et al. (2017) used the number of syllables, but that correlated perfectly with the number of words.

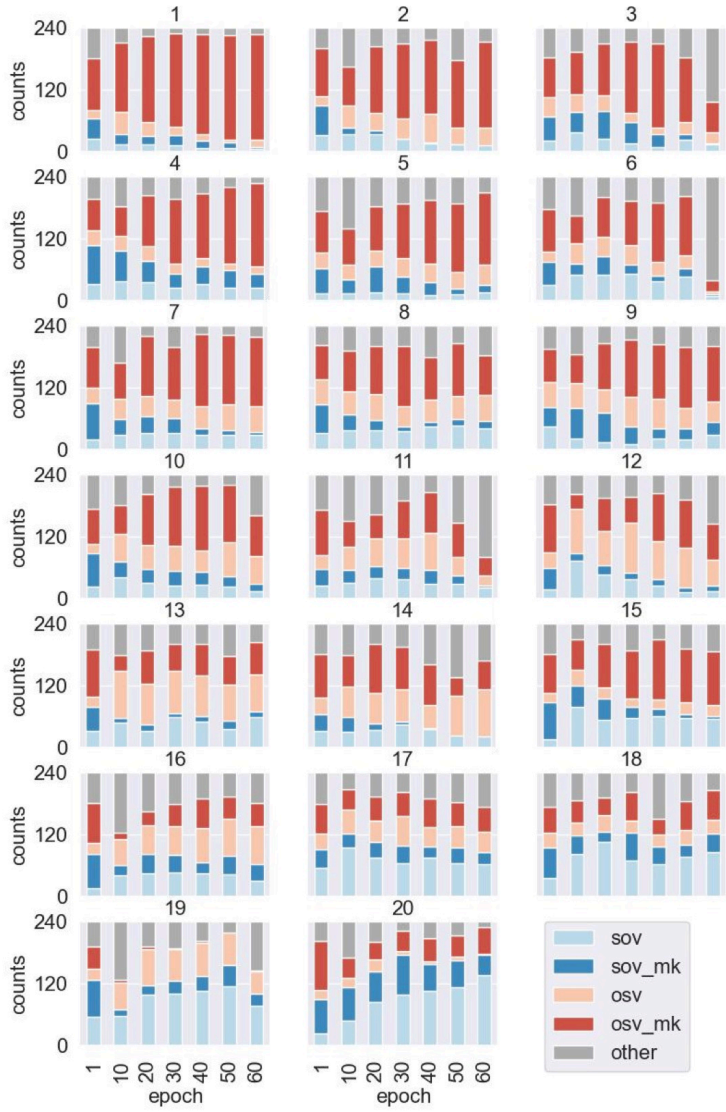
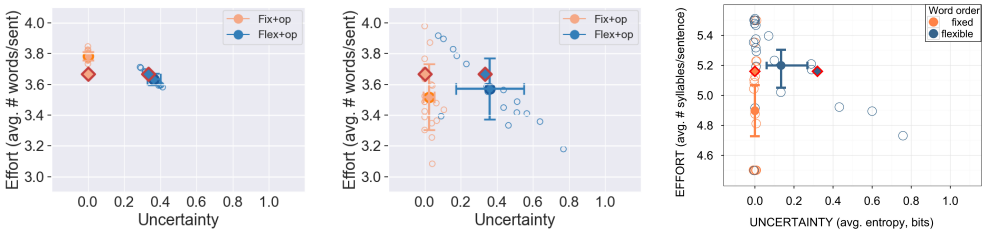


Figure 3.5: Individual production distributions (flex+op language). Utterances are categorized into 5 types, namely SOV without marker, SOV with marker, OSV without marker, OSV with marker and uncategorized (other). Color denotes word order (blue: SOV, red: OSV), shading denotes marking (dark: with marker, light: without). Subplots are manually arranged to highlight clusters of similar trajectories.



(a) Supervision Results **(b)** Communication Results **(c)** Human Results

Figure 3.6: Uncertainty (H) versus production effort: NellCom agents' results after supervised (a) and communication learning (b); human results on last day of training (c), reproduced with permission from Fedzechkina et al. (2017). Solid diamonds mark the initial uncertainty-effort value for each language. Empty circles represent the individual 20 agent pairs. Solid circles are the average of all agent pairs.

initial state (diamonds) reconfirms that SL alone does not lead to meaningful regularization. In fact, the only noticeable drift happens for fix+op in the counter-intuitive direction of increasing effort in the absence of uncertainty, as also anticipated by Fig. 3.3g. Communication results (Fig. 3.6b) show a very different picture: for both languages, average effort appears to *decrease* without noticeably increasing uncertainty. Variability among agents is also wide, as already noticed in the qualitative analysis of Section. 3.7. In fix+op, 17/20 agents produce shorter sentences. Fedzechkina et al. (2017) report effort reductions in 14/20 participants. In flex+op, the average uncertainty/effort values do not deviate much from the initial state, but individual data points reveal an unmistakable pattern, namely an inverse linear correlation between effort and uncertainty (empty blue circles in Fig. 3.6b). We closely inspect three instances: (i) The top-left data point (H=0.08, E=3.91) corresponds to agent pair no. 1 in Fig. 3.5 whose language becomes fixed-order (OSV) and fully marked, i.e. unambiguous but inefficient. (ii) The bottom-right data point (H=0.77, E=3.19) corresponds to no .19 in Fig. 3.5 where most markers are dropped (5% for OSV and 24% for SOV) but no order strongly dominates, resulting in high ambiguity. (iii) Finally, the data point at (U=0.09, E=3.43) represents the only clear outlier from the linear correlation. This agent pair, corresponding to no. 20 in Fig. 3.5, succeeds at minimizing *both* effort and

uncertainty by using SOV predominantly (76%) and reserving most markers to the less common order OSV (highly asymmetric case marking). Interestingly, no outliers are found on the other side of the line: i.e. none of the 20 agents pairs appears to increase both effort and uncertainty, just like in the human results (Fig. 3.6c).

3.8 Discussion and Conclusion

We studied the conditions in which the word-order/case-marking trade-off, a well established language universal example, could emerge in a small population of neural-network learners. We hypothesized that more naturalistic settings of language learning and use could lead to more human-like results, without the need to hard-code specific biases, such as least effort, into the agents. We then proposed a new Neural-agent Language Learning and Communication framework (NeLLCom) where pairs of speaking and listening agents learn a given language through supervised learning, and then use it to communicate with each other, optimizing a shared reward via reinforcement learning.

We used NeLLCom to replicate the experiments of Fedzechkina et al. (2017), where two groups of human participants were asked to learn a fixed- and a flexible-order miniature language, respectively, and to use it productively after training. Our results with RNN-based meaning-to-sequence and sequence-to-meaning networks confirm that SL is sufficient for perfectly learning the languages, but does not lead to any human-like regularization, in line with recent simulations of iterated learning (Chaabouni et al., 2019b; Lian et al., 2021). By contrast, communication learning leads agents to modify their production in interesting ways: Firstly, optional markers are dropped more frequently in the redundant fixed-order language than in the ambiguous flexible-order language, which matches human learning results. Moreover, one of the two equally probable word orders in the flexible-order language becomes clearly dominant and case marking starts to be used consistently more often in combination with one order than with the other. This conditional use of marking also matches human results. Some interesting differences were also observed: for instance,

NeLLCom agents showed, on average, a slightly stronger tendency to reduce effort rather than uncertainty. As another difference, several human subjects managed to 'break' the linear correlation by making the language more efficient *and* less uncertain, whereas this happened only in one of our agent pairs. Despite these differences, agents' productions show a clear correlation between effort and uncertainty, which strongly matches the core finding of Fedzechkina et al. (2017). We conclude that the word-order/case-marking trade-off as a specific realization of the efficiency/informativity trade-off can, in fact, emerge in neural network learners equipped with a need to be understood.

We made an important step towards developing a neural-agent framework that replicates patterns of human language change without the need to hard-code ad-hoc biases. Future work includes extending the current framework with iterated learning, which might lead agents to further optimize the ambiguous language and improve communication success over generations. We also plan to experiment with different neural network architectures to study the impact of architecture-specific structural biases, and with different word order universals.