Simulating the Emergence of Dependency Length Minimization Using Neural-network based Learning and Communication

Yuqing Zhang, Tessa Verhoef, Gertjan van Noord, Arianna Bisazza



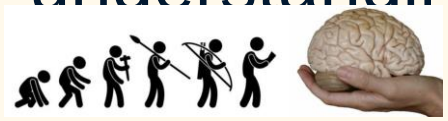

Who am I?

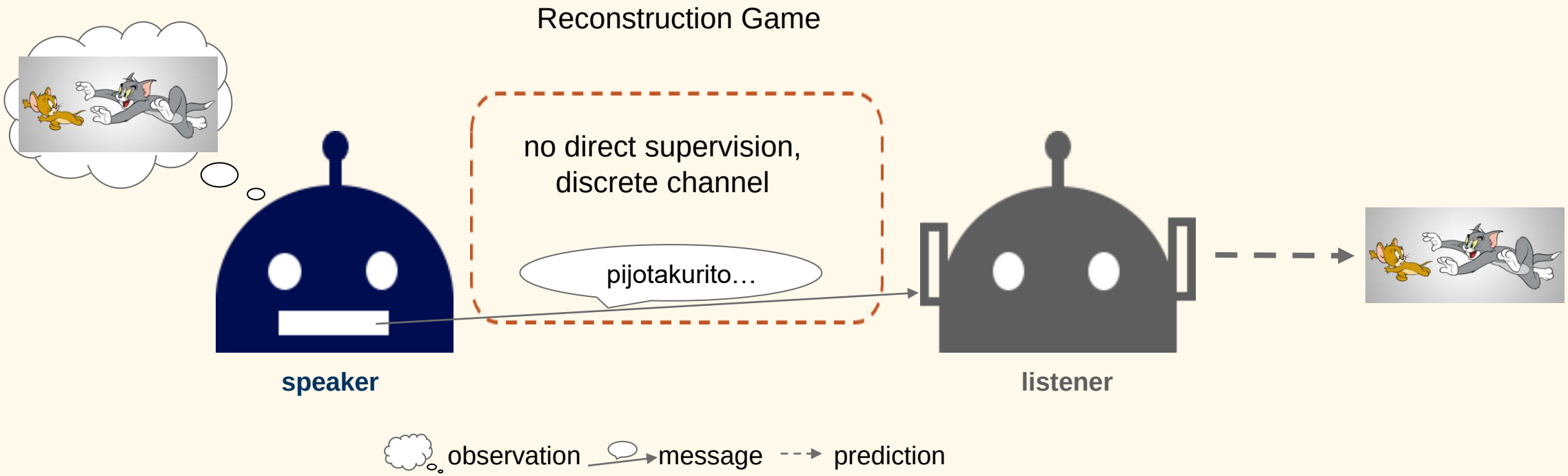
- Second-year PhD student of the computational linguistics group

How internal (e.g. cognitive) and/or external (e.g. language contact, communication) pressures shape the emergence of language universals

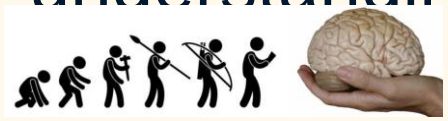
Neural network agent based simulation (frequently used in multi-agent emergent communication field)

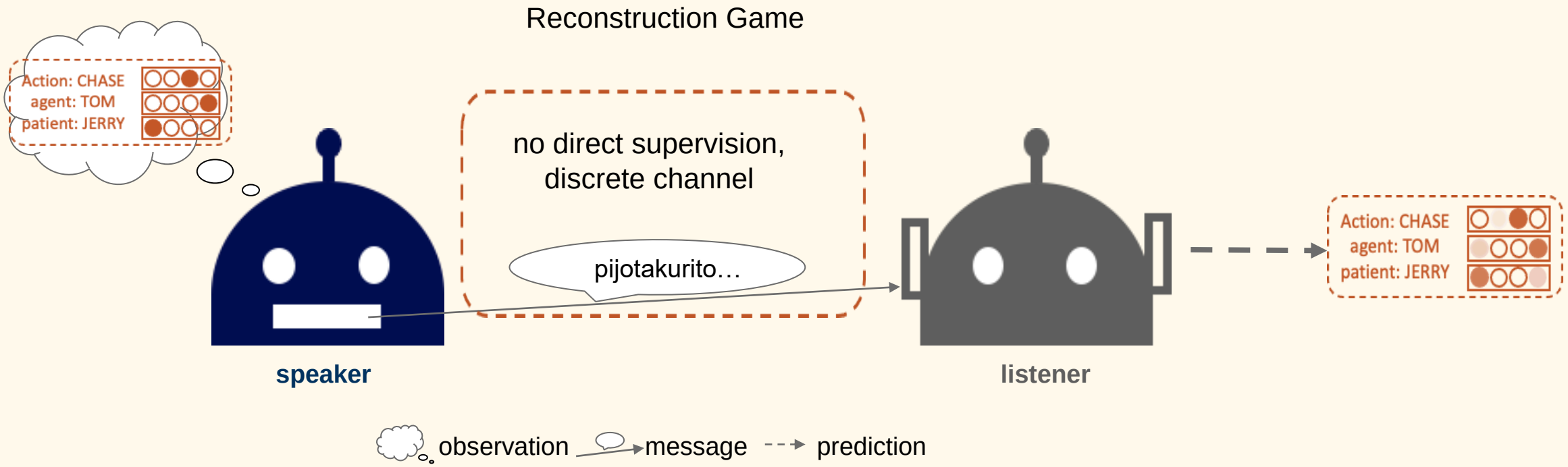
What is emergent communication?

- Experimental framework for a better understanding of human language emergence & evolution 
- Standard setup: the communication game

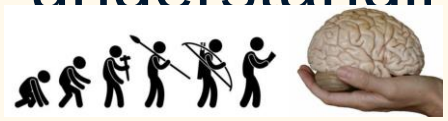


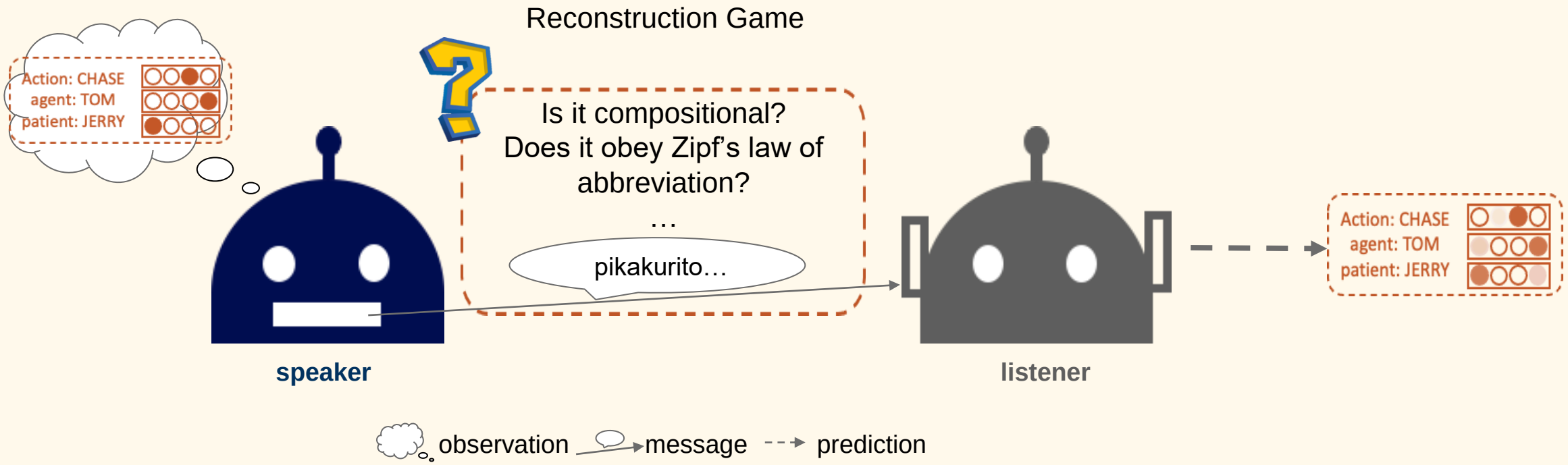
What is emergent communication?

- Experimental framework for a better understanding of human language emergence & evolution 
- Standard setup: the communication game



What is emergent communication?

- Experimental framework for a better understanding of human language emergence & evolution 
- Standard setup: the communication game



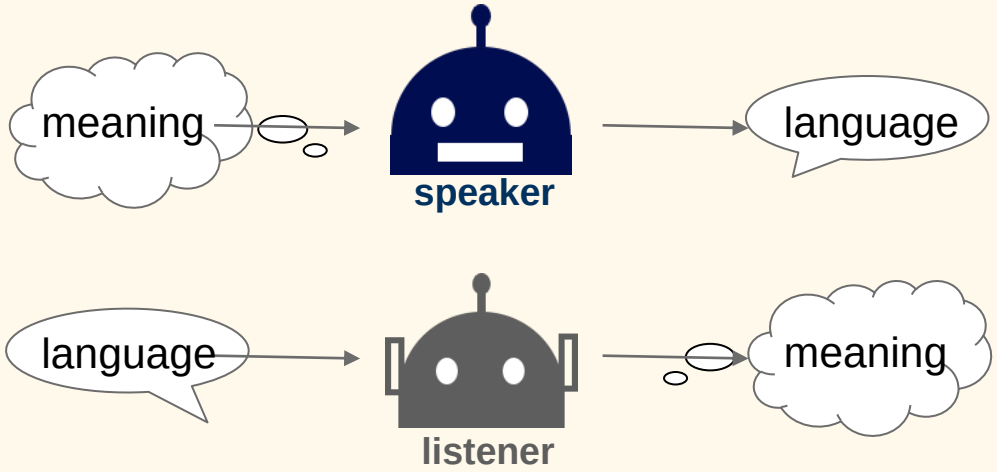
Simulating the emergence of a human-like language with statistical agents is hard…

(Lowe et al., 2017; Lowe et al., 2020)

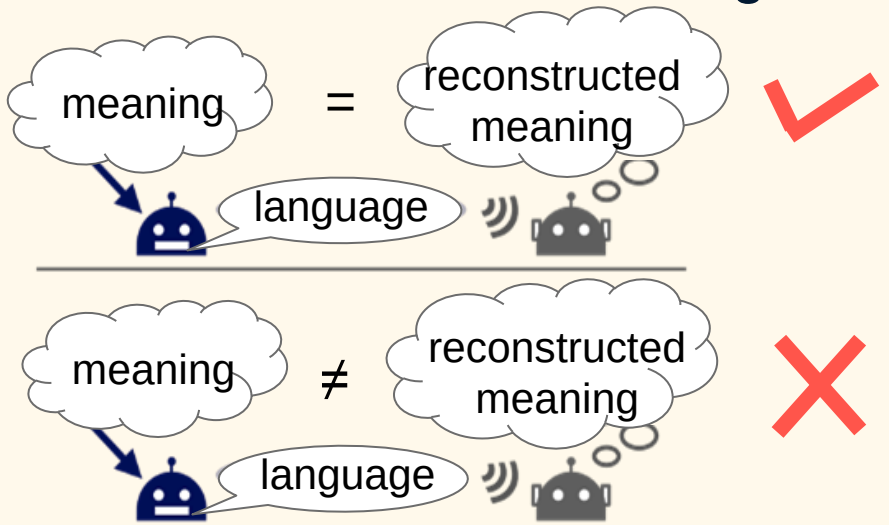
Initial supervision + communication

- Neural-agent Language Learning and Communication (NeLLCom): A New Framework for Studying the Emergence of Language Universals with Neural Agents (Lian, Bisazza & Verhoef, 2023)

Phase 1: Supervised Learning



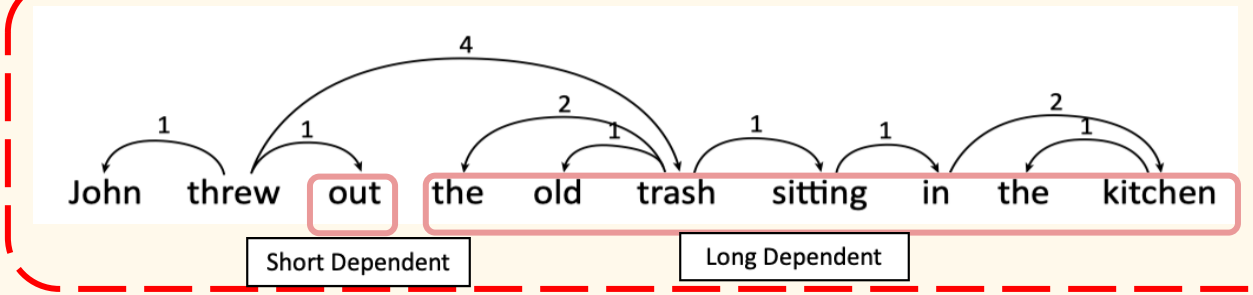
Phase 2: Communication Learning



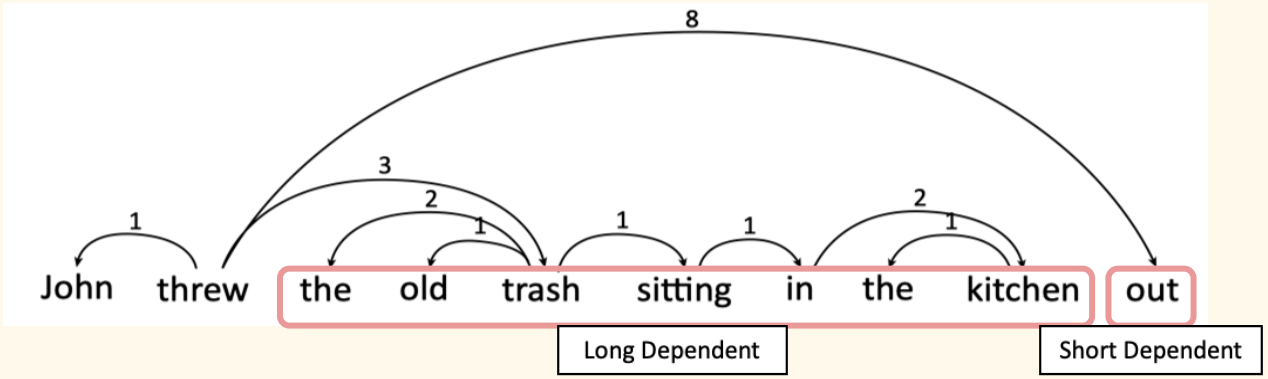
- How a predefined language w/ certain properties evolves after learning & communication

What are language universals?

- Typological studies show **some linguistic features occur more frequently than expected by chance**.
- Our case study- **Dependency Length Minimization (DLM)**
  - Verb-initial languages/verb-medial languages(e.g. English)
    - ✦ short-before-long preferences



Total dependency length: 14



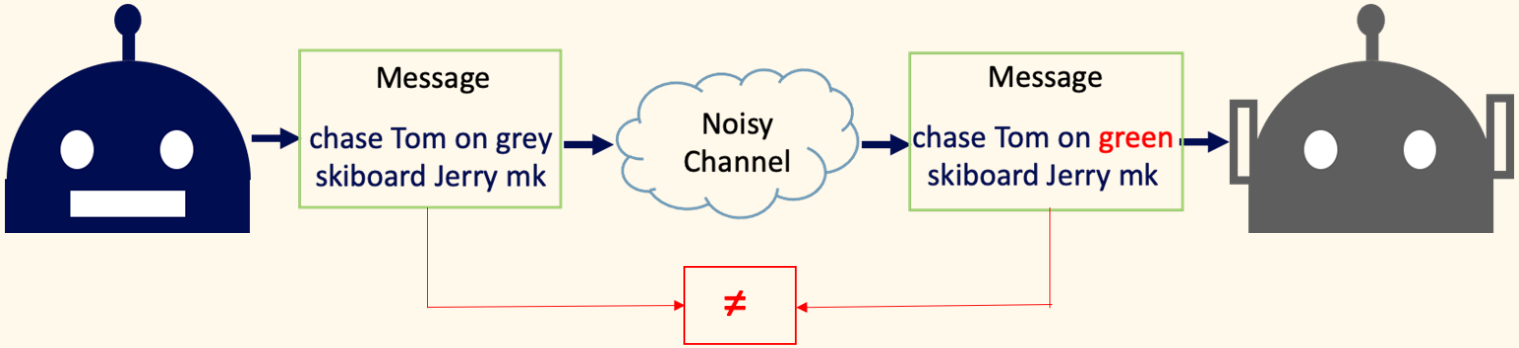
Total dependency length: 20

Outline

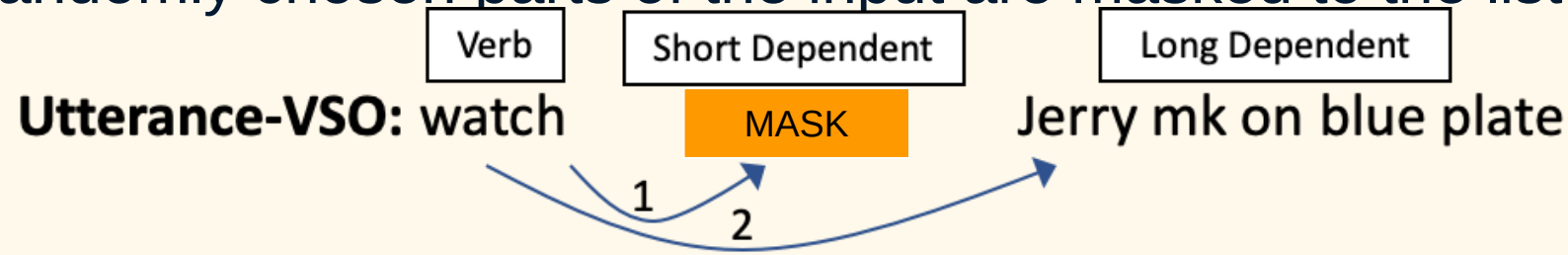
1. Emergent communication
2. Language universals (e.g. dependency length minimization)
3. **More realistic settings**
   - **Noisy communication**
   - More realistic distributions in predefined languages
   - Incremental sentence processing
4. Results & summary

Noisy communication

- Real-life communication: transmission errors, limited listener attention…



- Word dropout
  - randomly chosen parts of the input are masked to the listener

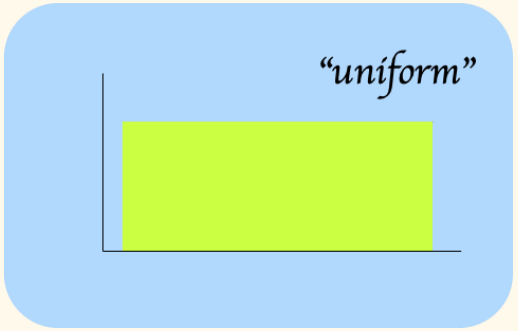


    - ✦ dropout rate δ ~ Bernoulli(p), p=[0, 0.1, 0.2]
  - both training & testing

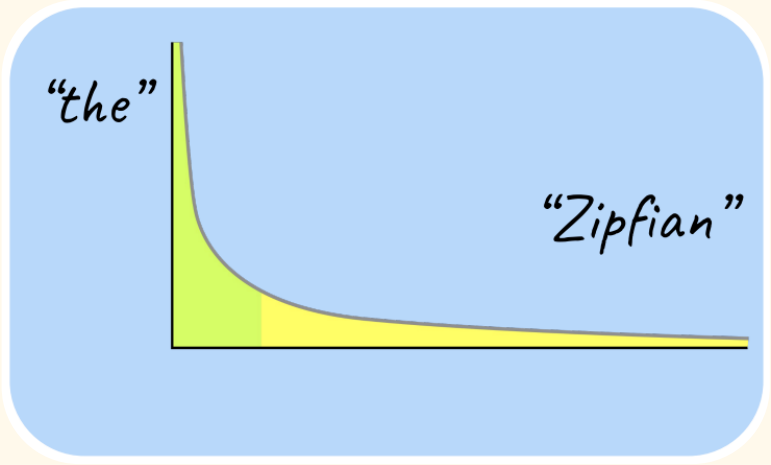
Outline

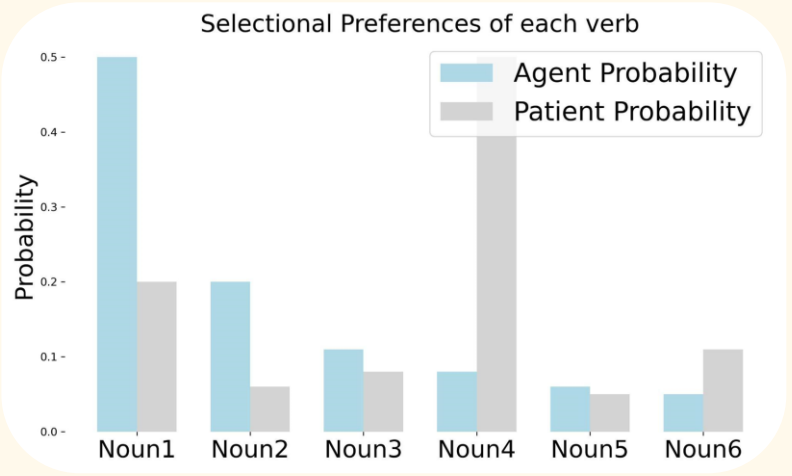
1. Emergent communication
2. Language universals (e.g. dependency length minimization)
3. **More realistic settings**
   - Noisy communication
   - **More realistic distributions in predefined languages**
   - Incremental sentence processing
4. Results & summary

More realistic distributions in predefined languages



Languages used in artificial language learning experiments with human participants or neural agents





Human languages

- We model **selectional preferences of verbs** via skewed **conditional word distributions** in our artificial language design.
  - strength of association/dependency of one action with one agent or patient

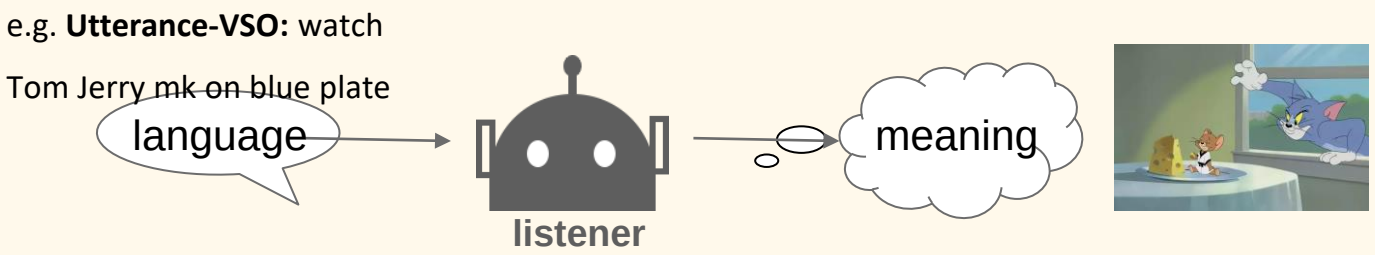
Outline

1. Emergent communication
2. Language universals (e.g. dependency length minimization)
3. **More realistic settings**
   - More realistic distributions in predefined languages
   - Noisy communication
   - **Incremental sentence processing**
4. Results & summary

Incremental sentence processing

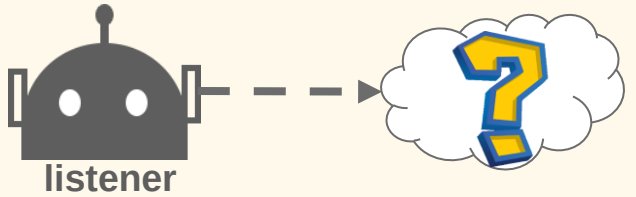
- **Trained listeners** process **varying-length prefixes** of the utterances.

Supervised learning phase:



Test phase:

prefix length=1: chase
prefix length=2: chase Tom
prefix length=3: chase Tom on
prefix length=4: chase Tom on grey
prefix length=5: chase Tom on grey skiboard
prefix length=6: chase Tom on grey skiboard Jerry
prefix length=6: chase Tom on grey skiboard Jerry mk



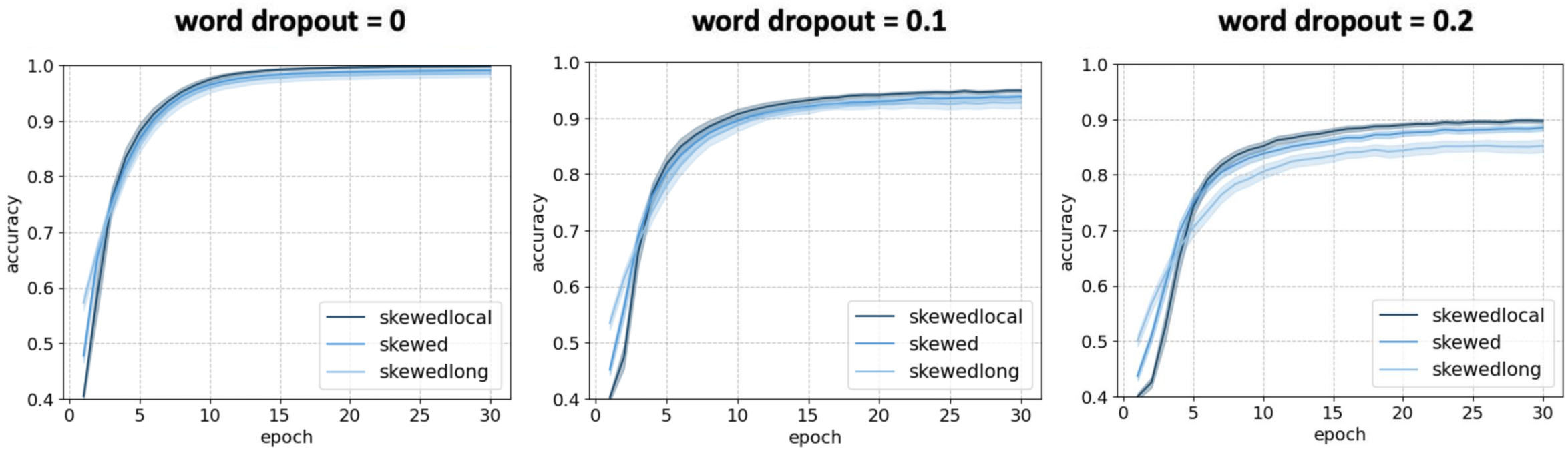
incremental listening accuracy

Outline

1. Emergent communication
2. Language universals (e.g. dependency length minimization)
3. More realistic settings
   - Noisy communication
   - More realistic distributions in predefined languages
   - Incremental sentence processing
4. **Results & summary**

Are short/local dependency languages easier to learn?

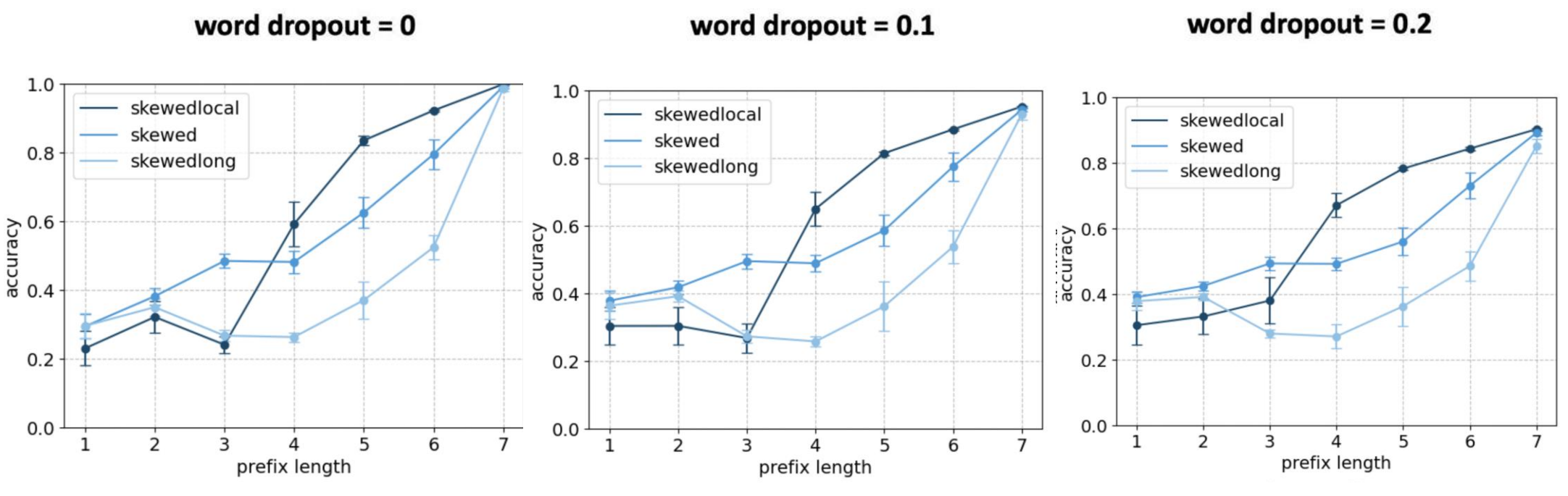
Listening Accuracy



- noise ↑ listening accuracy: (skewed_local > skewed (mixed w/ long & local) > skewed_long)
- Languages with shorter dependency length might be easier to learn for neural network agents when realistic noisy conditions are provided.

Incremental Listening Accuracy

- **Trained listeners** process **varying-length prefixes** of the utterances.

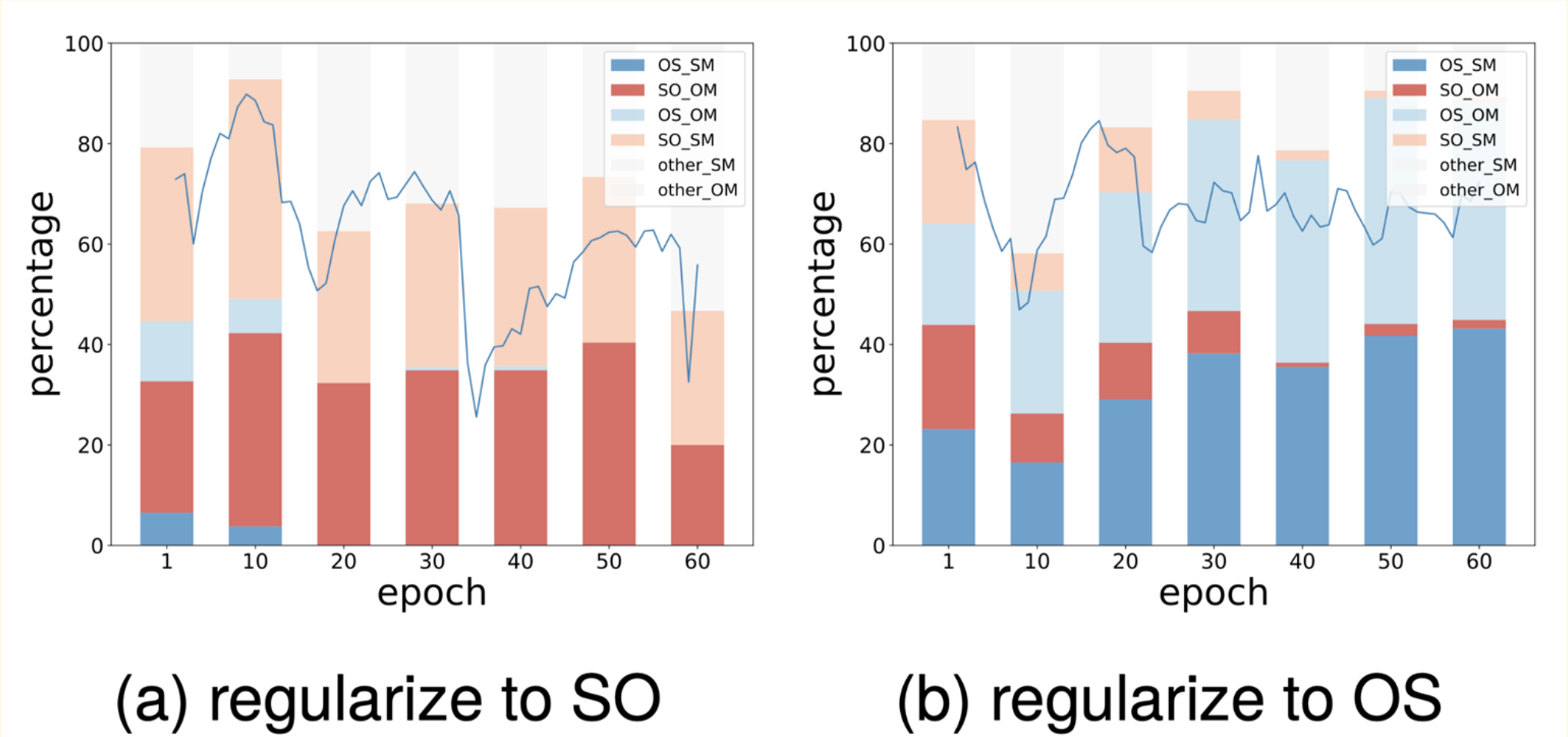


- The disambiguation of meanings of shorter dependency language tends to occur earlier in the sentence.

Will speakers produce more short-dependency utterances after communication?

Speaker's production after communication

Regularize towards one word order, but not towards shorter dependencies.



Summary

- We add three factors to the standard neural-agent language learning and communication framework to make the simulation more realistic.
  - (i) the presence of noise during listening,
  - (ii) context-sensitivity of word use through non-uniform conditional word distributions,
  - (iii) incremental sentence processing
- We simulate the emergence of DLM and find that the proposed factors can contribute to a **small but consistent learning advantage of DLM** for listeners of the verb-initial languages.

Thank you!