

Studying the Inductive Biases of RNNs with Synthetic Variations of Natural Languages

Shauli Ravfogel¹ Yoav Goldberg^{1,2} Tal Linzen³

¹Computer Science Department, Bar Ilan University

²Allen Institute for Artificial Intelligence

³Department of Cognitive Science, Johns Hopkins University

*presented by Anastasia Shimorina
Synalp Reading Group
11 April 2019*

Recall: what is an inductive bias?

Inductive bias of an ML algorithm is a set of assumptions the learning algorithm makes to produce hypotheses for unseen instances.

E.g., in linear regression the model assumes that the output or dependent variable is related to independent variable linearly (in the weights).

Paper Overview

- Experimental paradigm for studying the interaction of the NN inductive biases with typological properties
- Exploit synthetic corpora that differ in one typological parameter

E.g., instances:

a. *The man eats the apples.* *The man eats_{sing} the apples.*

b. *The man the apples eats.* *The man the apples eats_{sing}.*

task: predict the agreement features of the verb

Task: Subject/Object Plurality

The man <verb> the apples.

Prediction of one of the arguments:

- <singular subject>
- <plural object>

Joint prediction task:

- <singular subject, plural object>

Three prediction categories:

- Singular, Plural, None (for intransitive verbs)

Model

- L and R contexts of the verb represented as biLSTM (150 hidden units)
 - The representation is fed into an MLP (2 hidden layers of size 100 and 50)
 - Independent MLPs to predict subject and object plurality
 - Sum of word embeddings and embeddings of the char n -grams
-
- Training data: 35K sentences (derived from Penn Treebank)

Generated Synthetic Sentences (Agreement)

Original

they say the broker took them out for lunch frequently .
(*subject; verb; object*)

Polypersonal agreement

they saykon the broker tookkarker them out for lunch frequently .
(*kon: plural subject; kar: singular subject; ker: plural object*)

Word order variation

SVO

they say the broker took out frequently them for lunch .

SOV

they the broker them took out frequently for lunch say .

VOS

say took out frequently them the broker for lunch they .

VSO

say they took out frequently the broker them for lunch .

OSV

them the broker took out frequently for lunch they say .

OVS

them took out frequently the broker for lunch say they .

Case systems

Unambiguous

theykon saykon the brokerkar tookkarker theyker out for lunch frequently .
(*kon: plural subject; kar: singular subject; ker: plural object*)

Syncretic

theykon saykon the brokerkar tookkarkar theykar out for lunch frequently .
(*kon: plural subject; kar: plural object/singular subject*)

Argument marking

theyker sayker the brokerkin tookkerkin theyker out for lunch frequently .
(*ker: plural argument; kin: singular argument*)

NB: Suffixes were omitted for the word order variation examples

Polypersonal Agreement

Goal: determine whether jointly predicting both object and subject plurality improves the overall performance of the model.

Prediction task	Subject accuracy	Object accuracy	Object recall
Subject	94.7 ± 0.3	-	-
Object	-	88.9 ± 0.26	81.8 ± 1.4
Joint	95.7 ± 0.23	90.0 ± 0.1	85.4 ± 2.3

Recall: model predicted Singular or Plural, but not None

Findings:

Training on an auxiliary agreement prediction task may be beneficial

Generated Synthetic Sentences (Word Order)

Original		they say the broker took them out for lunch frequently . (<i>subject; verb; object</i>)
Polypersonal agreement		they saykon the broker tookkarker them out for lunch frequently . (<i>kon: plural subject; kar: singular subject; ker: plural object</i>)
Word order variation	SVO	they say the broker took out frequently them for lunch .
	SOV	they the broker them took out frequently for lunch say .
	VOS	say took out frequently them the broker for lunch they .
	VSO	say they took out frequently the broker them for lunch .
	OSV	them the broker took out frequently for lunch they say .
	OVS	them took out frequently the broker for lunch say they .
Case systems	Unambiguous	theykon saykon the brokerkar tookkarker theyker out for lunch frequently . (<i>kon: plural subject; kar: singular subject; ker: plural object</i>)
	Syncretic	theykon saykon the brokerkar tookkarkar theykar out for lunch frequently . (<i>kon: plural subject; kar: plural object/singular subject</i>)
	Argument marking	theyker sayker the brokerkin tookkerkin theyker out for lunch frequently . (<i>ker: plural argument; kin: singular argument</i>)

NB: Suffixes were omitted for the word order variation examples

Word Order

- 6 word orders
- Unchanged
- Flexible (uniform selection of orders for each sentence)

While permuting:

- Move clausal subjects and clausal complements as well
- Keep non-core dependents in the original position wrt the verb

Task: Joint prediction

OVS: *them <verb> out frequently the broker for lunch saykon they .*

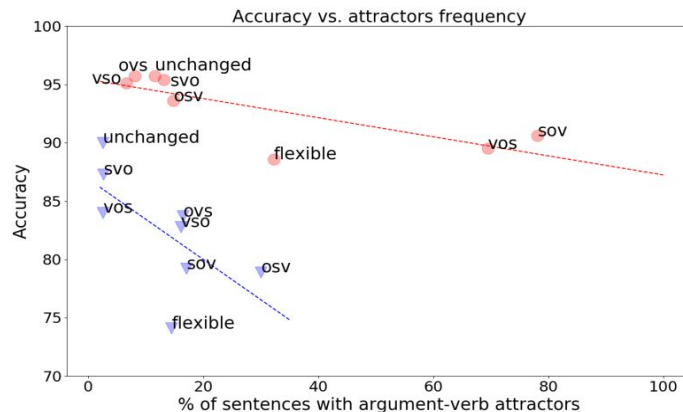
Word Order

Goal: inductive biases favoring certain word orders?

Order	Subject		Object		
	% Attractors	Accuracy	% Attractors	Accuracy	Recall
Unchanged	11.56	95.7 ± 0.23	2.55	90.0 ± 0.1	85.4 ± 2.37
SVO	13.16	95.4 ± 0.41	2.6	87.3 ± 0.23	80.0 ± 2.61
SOV	78.12	90.6 ± 0.37	17.04	79.2 ± 0.78	63.3 ± 4.62
VOS	69.50	89.5 ± 0.54	2.57	84.0 ± 0.39	77.8 ± 3.68
VSO	6.65	95.1 ± 0.12	16.09	82.8 ± 0.7	70.0 ± 1.91
OSV	14.81	93.6 ± 0.23	30.00	78.9 ± 0.17	63.5 ± 4.59
OVS	8.13	95.7 ± 0.37	16.42	83.7 ± 0.32	72.8 ± 1.58
Flexible	32.24	88.6 ± 0.43	14.44	74.1 ± 0.70	60.2 ± 3.24

Word Order: Results

- Accuracy is negatively correlated with the freq of attractors
- Inductive bias favoring dependencies with recent elements
- Flexible WO: human learner would be also stuck



RNN Recency Bias

Two compatible generalisations for intransitive verbs:

- The subject is the most recent core element before the verb.
- The subject is the first core constituent of the clause.

More controlled experiment: withhold direct objects for SOV and VOS during training

- A. Object of the opposite plurality
- B. Object of the same plurality as the subject
- C. Without object, but with a noun of the opposite plurality (*The gap between winners and losers will grow.*)

	Object (attractor)	Object (non attractor)	Non-object attractor
SOV	60.3 ± 3.7	92.8 ± 0.3	79.2 ± 3
VOS	48.3 ± 2.3	94.0 ± 2.3	83.1 ± 1.1

Word Order: Findings

- Strong recency bias of RNNs
- RNNs were able to distinguish between core and non-core dependents
- Subject agreement is more difficult to model in SOV than in SVO (cf. typological studies)
- Weak support for typological word order distributions (best performance on OVS)

Generated Synthetic Sentences (Case Marking)

Original		they say the broker took them out for lunch frequently . (<i>subject; verb; object</i>)
Polypersonal agreement		they saykon the broker tookkarker them out for lunch frequently . (<i>kon: plural subject; kar: singular subject; ker: plural object</i>)
Word order variation	SVO	they say the broker took out frequently them for lunch .
	SOV	they the broker them took out frequently for lunch say .
	VOS	say took out frequently them the broker for lunch they .
	VSO	say they took out frequently the broker them for lunch .
	OSV	them the broker took out frequently for lunch they say .
	OVS	them took out frequently the broker for lunch say they .
Case systems	Unambiguous	theykon saykon the brokerkar tookkarker theyker out for lunch frequently . (<i>kon: plural subject; kar: singular subject; ker: plural object</i>)
	Syncretic	theykon saykon the brokerkar tookkarkar theykar out for lunch frequently . (<i>kon: plural subject; kar: plural object/singular subject</i>)
	Argument marking	theyker sayker the brokerkin tookkerkin theyker out for lunch frequently . (<i>ker: plural argument; kin: singular argument</i>)

Syncretic: modelled after Basque; Argument marking: only the plurality is indicated (syncretism in Russian)

Case Marking: Results

Case system	Flexible word order		VOS		OVS	
	Subject A	Object A/R	Subject A	Object A/R	Subject A	Object A/R
Unambiguous	99.2 \pm 0.5	98.7 \pm 0.2 /98.0 \pm 0.5	98.9 \pm 0.2	99.5 \pm 0.1 /99.1 \pm 0.1	99.5 \pm 0.2	98.6 \pm 0.3 /98.4 \pm 0.6
Syncretic	99.3 \pm 0.2	93.6 \pm 0.4 /88.9 \pm 1.7	99.1 \pm 0.2	97.1 \pm 0.2 /95.0 \pm 1.1	99.4 \pm 0.1	97.8 \pm 0.2 /97.4 \pm 1.2
Argument marking	96.0 \pm 0.3	86.1 \pm 0.9 /79.7 \pm 4.9	96.9 \pm 0.1	93.6 \pm 0.1 /89.8 \pm 2.4	99.6 \pm 0.1	96.8 \pm 0.1 /95.5 \pm 0.5

Table 5: Accuracy (A) and recall (R) in predicting subject and object agreement with different case systems.

Case Marking: Findings

- Flexible+argument marking gives comparable results on unmodified English
- Support the observation that languages with explicit case marking tend to allow a more flexible word orders

Conclusions

- Multitask training on polypersonal prediction improved performance
- High variation in performance across word orders
- This variation is not correlated with the freq of WO in the world languages
- Exhibit a recency bias
- Overt case making dramatically improved plurality prediction performance

Case Suffixes

	Singular	Plural
Subject	-kar	-kon
Object	-kin	-ker
Indirect Object	-ken	-kre