What can language models teach us about human language?

Isabel Papadimitriou
Language Models

- Artificial models of human language
- Recent, huge progress
  - In many cases, a pretty good approximation of novel human language production
- But, we don’t know how they do it
- How can we use LMs to learn about language?
What is a language model?

The cat sat on the __________

Behavior is impressive

Process is a mystery

Neural network (1,000s of dimensions)
An exciting development for linguistics

Language models are:

- **Very good** –
  - We have computational models of language that capture a lot of the subtlety of language use

- **Very empirically flexible** –
  - We can control their training and examine their language system
What makes language models so good is precisely *that* we don’t understand how they work.

- Language models are not engineering products applying one linguistic theory or analysis.
What makes language models so good is precisely *that* we don’t understand how they work

- Language models are not engineering products applying one linguistic theory or analysis

Now:

- A functioning theory of **possibilities to analyze human language** [Baroni 2021]
Empirical flexibility: experiments that are impossible with humans

The cat sat on the ______

Matix operations

Totally control training data

Neural network

mat
Empirical flexibility: experiments that are impossible with humans.

The cat sat on the _____.

Analyze internal representations.

Neural network

Matrix operation

alyze internal representations

Neural network

Matrix operation
Language models are:

- **Very good** –
  - We have computational models of language that capture the subtlety of language use

- **Very empirically flexible** –
  - We can control their training and examine their language system

- **But.. very different from humans** –
  - Unconstrained language learners
  - Continuous high-dimensional space
This talk: methodologies for **bridging the gap**

Use language models to address two linguistic questions:

What makes language acquisition possible?  
How do speakers represent syntactic information?

---

What makes language acquisition possible?

- [Papadimitriou and Jurafsky 2020, Papadimitriou and Jurafsky 2023]

How do speakers represent syntactic information?

- [Papadimitriou et al 2021, Papadimitriou et al 2022]
This talk: methodologies for bridging the gap

- But.. very different from humans –
  - Unconstrained language learners
  - Continuous high-dimensional space

1) Bias language models towards theoretically-significant structural constraints
This talk: methodologies for bridging the gap

- But.. very different from humans –
  - Unconstrained language learners
  - Continuous high-dimensional space

2) Map out the representation of grammatical role in the model’s internal space
Use language models to address two linguistic questions:

What makes language acquisition possible?

Method: structural injection before LM training

How do speakers represent syntactic information?

Method: subjecthood representation analysis
What does a language learner need to start from?

Method: Inject a model with a bias that we choose

Language exposure

Recursion
Inject structural language, learn natural language

Formal structure

Neural network
Inject structural language, learn natural language

1. Formal structure

2. The cat sat on the _____

Does the structural bias help with language learning?
With structural injection, we combine theoretically-significant biases with the power of LMs

**Structural bias**

**Strong statistical learner**

LMs let us do **hypothesis testing** of different biases
Test three types of structure:

1) Recursion

The lawyer that the man that the dog bit hired was disbarred

2) Simple regular bias, repetition

And he said this, and he said that

3) Crossing dependencies

“I voted for him even though I am negatively affected by his redistribution policies” he said
Recursion

- Nesting, context-free

Hypothesis that recursion is what makes language [Hauser Chomsky Fitch 2002]

The cat sat on the mat

I think that the cat sat on the mat

You always accuse me that I think that the cat sat on the mat
Recursion

- Nesting, context-free

- Hypothesis that recursion is what makes language [Hauser Chomsky Fitch 2002]
Structural injection formal language: Nesting parentheses

- Well-nested, matching pairs
- Constituents: S strings contain S strings
Recursive bias helps language acquisition

![Graph showing perplexity on English for Random and Nesting Parentheses pre-trained models. The Random model has a higher perplexity, indicating worse performance, while the Nesting Parentheses model has a lower perplexity, indicating better performance. The y-axis represents perplexity (lower is better), and the x-axis represents pre-trained models.]
Simple regular bias

- Is a recursive bias really necessary?
- Test a structural bias that’s not very theoretically important in human language

Inject finite repetition language (regular)

\[499 \ 472 \ 300 \ 499 \ 472 \ 300 \ 309 \ 18 \ \ldots\]
Recursive bias is needed – not just any simple structure
Crossing links and dependencies bias

- Crossing dependencies arise in meaning, reference, discourse, pragmatic relationships

- Example: anaphora

"I voted for **him** even though I am negatively affected by **his** redistribution policies" he said
Crossing Dependencies

- Tokens have to **match**, but not **nest**
Complex, crossing dependencies provide the best bias – with no recursion.
Does crossing-type context sensitive structure always help?

- Mix nesting and cross: mostly nesting, with 1%, or 10% of parentheses not following the structure.
Slightly breaking constituent structure makes better language learners
Are these cognitive biases unique to language?
What does a language learner need to start from?

- Importence of bootstrapping meaning, discourse, reference, information structure
Use language models to address two linguistic questions:

What makes language acquisition possible?

Method: structural injection before LM training

How do speakers represent syntactic information?

Method: subjecthood representation analysis
Linguistic property: subjecthood

- Who does what to who, being the subject vs the object
- Subjecthood is relevant in basically every utterance, in every language
- How do we represent this relation?
How is subjecthood represented in language models?

- Mapping out grammatical role in neural networks

The subject
verbs the
object
Use LMs to look at subjecthood:

1) Across different languages

2) Interacting with semantics
Subjecthood representation in different languages

The subject verbs the object

Neural network (1,000s of dimensions)
Subjecthood representation in different languages

Does the English mapping still work?

(1,000s of dimensions)
Subjecthood is cross-lingual

- Subject-object geometry is similar in-language (red) and out-of-language (black)
Typology: how languages treat **intransitives**

Transitive: The *dog* chased the *cat*

Intransitive: The *glass* broke

**Nominative** languages

**Ergative** languages
Subjecthood representation of intransitives

The subject verbs the object

Neural network (1,000s of dimensions)
Subjecthood representation of intransitives

Neural network (1,000s of dimensions)

The intransitive verbs
What is the behavior of those universal classifiers on S nouns?

- Nominative languages
  - S is labelled O

- Ergative, split-ergative languages
  - S is labelled A
How does this robust cross-lingual representation work?

- **Proposal**: through integrating the grammatical relationship of subjecthood with cross-lingual meaning representation
Subjecthood is complicated, influenced by meaning

Intransitives
The glass broke
Isabel broke the glass

Discrete
S
NP
A A N V Adv
Colorless green ideas sleep furiously

Animacy
He ran all day
The fridge ran all day

Passive voice
The cat jumped on to the perch
The perch was jumped on to by the cat

Case

Volitionality
Mary punched Sam
Mary liked Sam
Mary forgot Sam

[Comrie 1989 Language Universals and Linguistic Typology]
[Hopper and Thompson 1980 Transitivity in Grammar and Discourse]
Is subjecthood a discrete category?

Discrete

Prototype

Animacy,
Passive voice,
Volitionality,
Agency,
Case,

Colorless green ideas sleep furiously
Classifier probabilities show **animacy** effects, even when controlling for syntactic role

Animacy

**He** ran all day

The **fridge** ran all day
Classifier probabilities show passive voice effects

Passive voice

The cat jumped on to the perch

The perch was jumped on to by the cat
● **Prototype effects in LMs:** Many factors play into making something a subject

We also look at the effect of **case**. Also working on: discourse, information structure (given/new)
But is it all just prototypes?

Discrete vs. Prototype

Colorless green ideas sleep furiously

Animacy, Passive voice, Volitionality, Agency, Case, ...
What if we test the same sentences (*with the same prototype effects*) but we **swap the labels**?

The **chef** chopped the **onion**,  

The **onion** chopped the **chef**

Do they have different classifications?
Yes – Representation differences that are caused **only** by syntactic word order
Both grammatical subjecthood and prototype effects

Discrete

Prototype

Animacy,
Passive voice,
Volitionality,
Agency,
Case,
The high-dimensional space of LMs provides a model for a complex notion of subjecthood

- Both grammatical and functional aspects, in one representation model
Use language models to address two linguistic questions:

What makes language acquisition possible?

Complex, crossing dependencies bootstrap language learning

How do speakers represent syntactic information?

Combine discrete grammatical rules with functional semantics
Language models and language

- Language models are a flexible testbed for thinking about human language
- We can control their training, and inspect their internal representations
- LMs provide tangible models for expanding linguistic theory

Thanks!