



THE ROYAL
SOCIETY

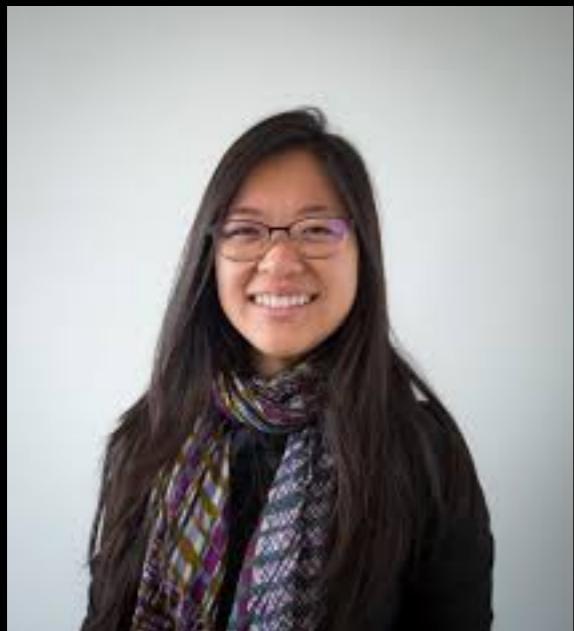


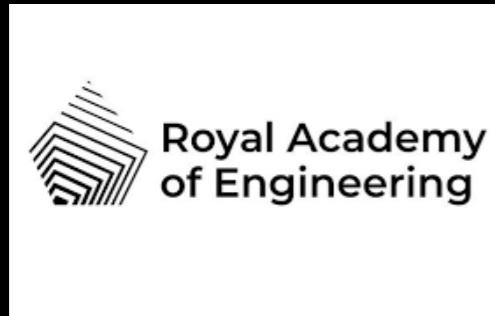
How can LLMs learn from Humans?

Wang, Sadrzadeh, Chow, Breheny (UCL)

+ Stajonovic (DeepMind)

UCL PNLoS, RAErg RC





How can LLMs become more human?

Mehrnoosh Sadrzadeh, Professor of CS and RAEng Research Chair
Joint work with Wang, Chow, Breheny (UCL) + Stajonovic (UCL and DeepMind)



LLMs

- LMs have come a long way, leading to LLMs.
- I feel really tired, what should I do? When tired, you must rest, lie down and close your eyes, take a gentle walk, put your feet up.

- BERT: When really tired [...]
- Chat GPT:

If you're feeling really tired, here are some quick, effective ways to recharge:

- 1. Take a Short Nap (10–20 minutes)**
 - A quick power nap can boost energy without leaving you groggy. Try not to exceed 20 minutes so you avoid deeper sleep stages, which can make waking up more difficult.
- 2. Stay Hydrated**
 - Sometimes fatigue is linked to dehydration. Drink a glass of water to see if that helps perk you up.
- 3. Get Fresh Air**
 - Stepping outside or opening a window for fresh air and sunlight can instantly improve energy. The natural light can help reset your internal clock and boost mood.
- 4. Stretch or Move Around**

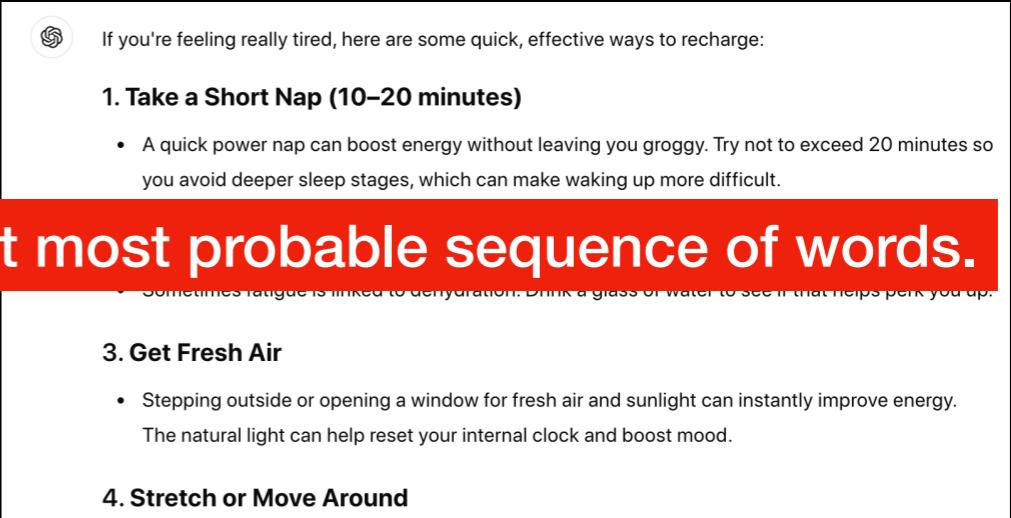
LLMs

- LMs have come a long way, leading to LLMs.
- I feel really tired, what should I do? When tired, you must rest, lie down and close your eyes, take a gentle walk, put your feet up.

- BERT: When really tired [...]

Their predictions are lexical, over the next most probable sequence of words.

- Chat GPT:



If you're feeling really tired, here are some quick, effective ways to recharge:

- 1. Take a Short Nap (10–20 minutes)**
 - A quick power nap can boost energy without leaving you groggy. Try not to exceed 20 minutes so you avoid deeper sleep stages, which can make waking up more difficult.
- Sometimes fatigue is linked to dehydration. Drink a glass of water to see if that helps perk you up.
- 3. Get Fresh Air**
 - Stepping outside or opening a window for fresh air and sunlight can instantly improve energy. The natural light can help reset your internal clock and boost mood.
- 4. Stretch or Move Around**

Human Language Understanding

- Psycholinguistic experiments: humans language understanding is based on prediction:
 - Structural
 - Syntactic, semantic, pragmatic, phonologic, morphologic,
 - A famous test: Garden path sentence

Garden Path

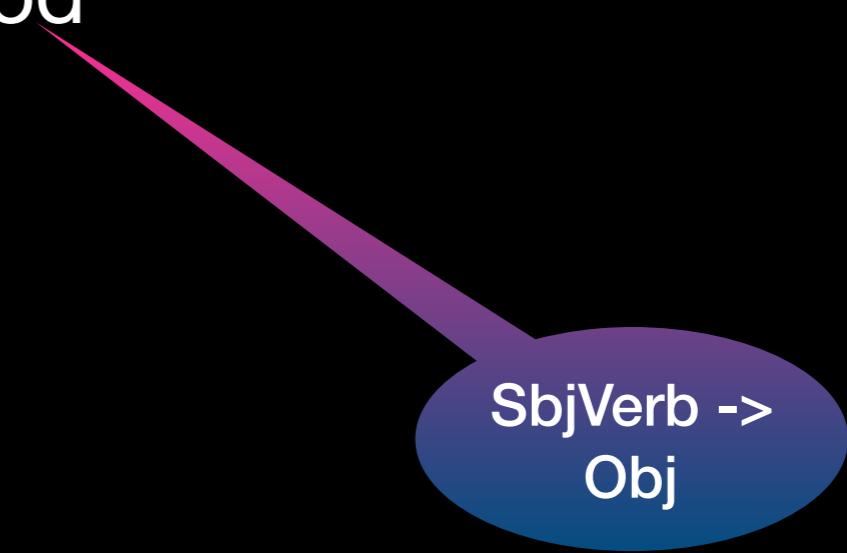
The faithful employees understood [NP] the technical contract would be changed very soon.

Garden Path

The faithful employees understood

Garden Path

The faithful employees understood



SbjVerb ->
Obj

Garden Path

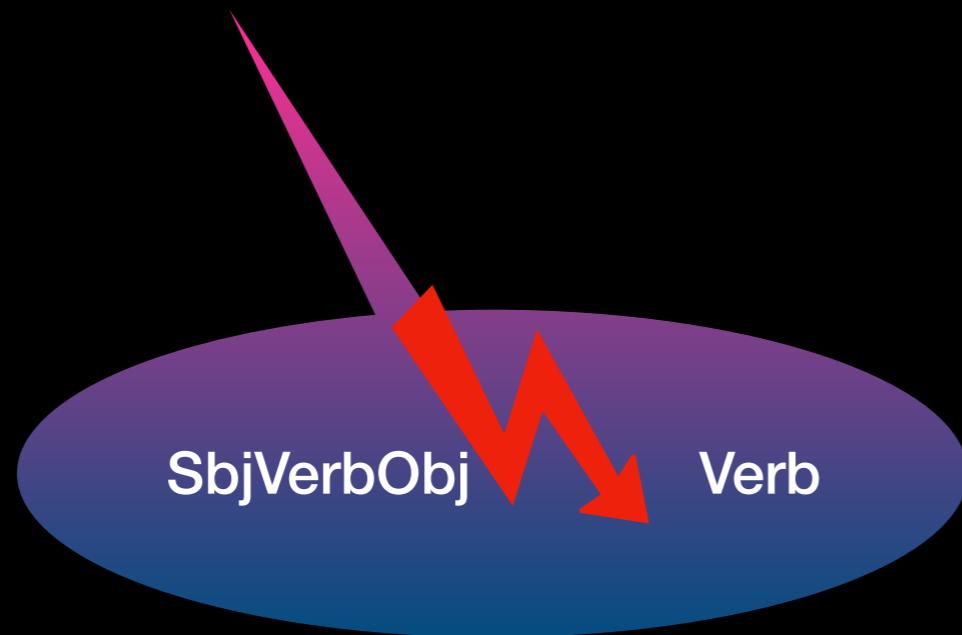
The faithful employees understood the technical *contract*



SbjVerbObj -> Adv

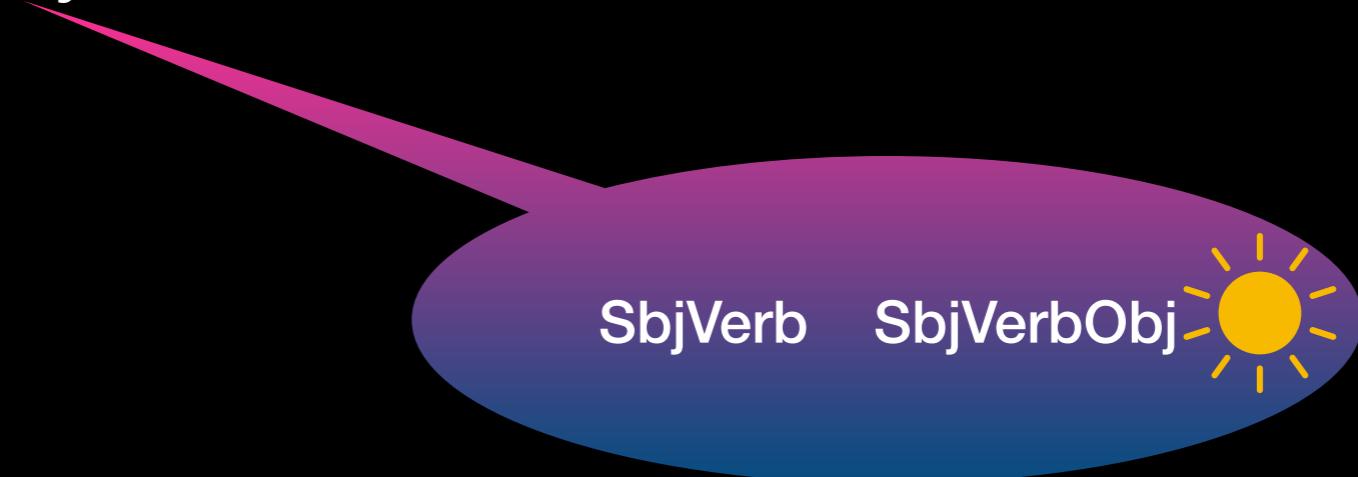
Garden Path

The faithful employees understood the technical contract would be.



Garden Path

The faithful employees understood the technical *contract* would be changed very soon.



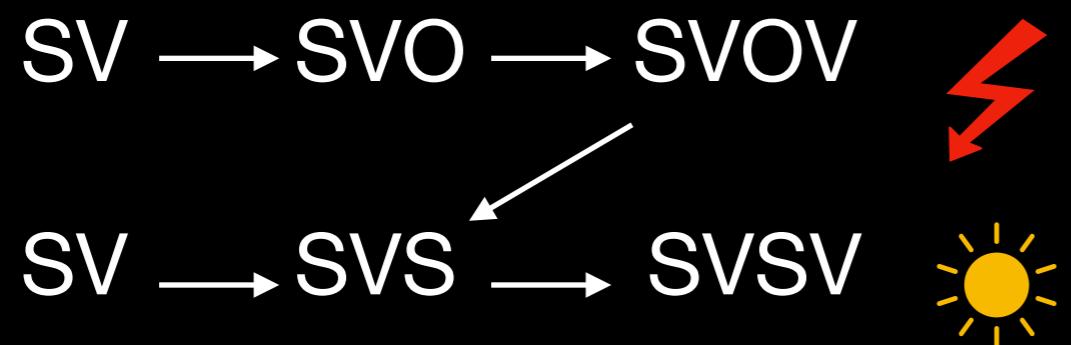
SbjVerb

SbjVerbObj



Garden Path Sentences

As the woman edited the magazine entertained the guests.



Pinned by Bever 1970, arguing for existence of

primordial relational structures in human sentence processing.



Related Work for Humans

- Psycholinguistic experiments with Eye Tracking to measure human reading time.
 - Pickering and Traxler, 1998, “Recovery from Garden Path sentences”, classified GP, including:
 1. NP/S: a sentential complement S is mistaken for a noun phrase NP.
 - *The faithful employees understood [NP] the technical contract would be changed very soon.*
 2. NP/Z: a main verb is expected to have an NP as object, but it does not (Z for zero objects).
 - *Because the employees negotiated [NP] the technical report would be changed very soon.*
 - Sturt, Pickering Crocker, 1999, “Structural change and reanalysis in language comprehension”
 - 64 GP + controls
 - Grodner et al, 2003, “Reanalysis in human sentence comprehension”
 - 80 GP + controls

Related Work for LLMs

- **Surprisal was related to human reading times.**
 - Hale, 2001, 2003, 2006: the information theoretic measure of surprisal should correlate with humane cognitive load.
 - Levy 2008, Smith&Levy 2013: experimental evidence but only for naturalistic data.
 - Wilcox et al 2023: 5 different language families, above.



Problems with SP



- Schijndel and Linzen 2018, 2021, Arehalli et al 2022, Huang et al 2023:
 - SP does not correlate well with human reading times in Garden Path sentences
 - It cannot distinguish between different structural types of GP, e.g. easy versus hard garden path.

Our solution

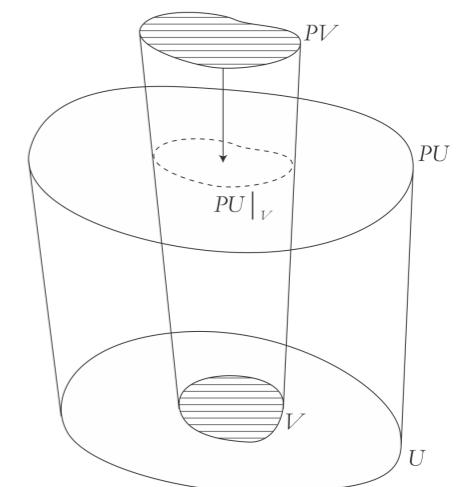
- A framework that combines statistics and structure.
- An event presheaves $P(X, \prec)$ over a preorder (X, \prec) and a set of events $s: U \rightarrow O$ consists of:

$U \longrightarrow P(U)$

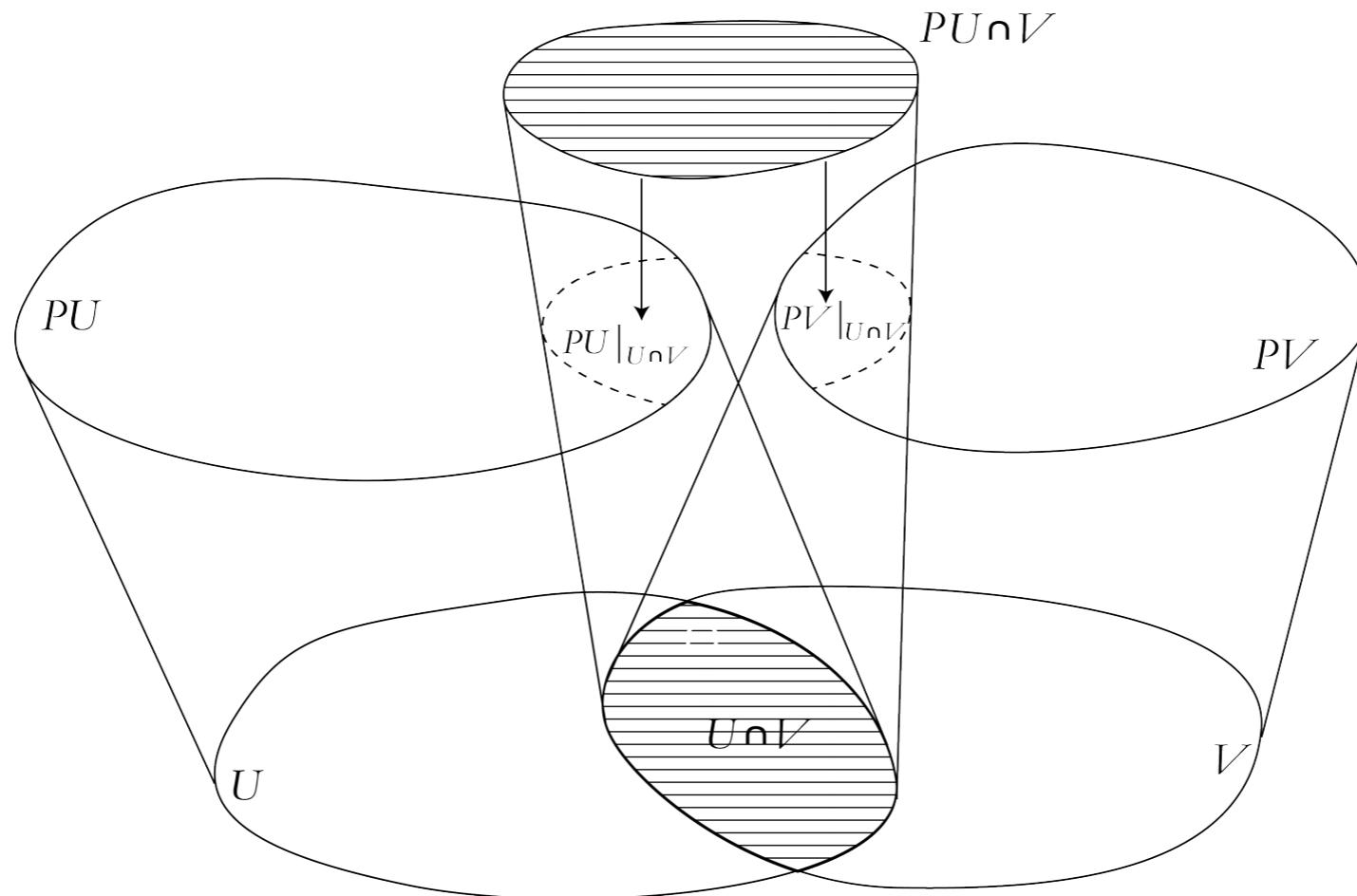
sends an element U of X to its “**data**”
each data point is called a “**section**”

$V \prec U \longrightarrow P(V \prec U)$

acts on the preorder by *restricting* P_U to P_V , i.e. the data on U to the data on V .



Glueing and Compatibility



There is a *glueing* between PU and PV whenever there is an intersection between U and V , so is one between their restrictions $PU|_{U \cap V}$ and $PV|_{U \cap V}$.

Probabilities

When the data is probabilistic, the presheaf map is composed with the distribution map.

The probabilistic presheaf assigns probability distributions to elements U of X :

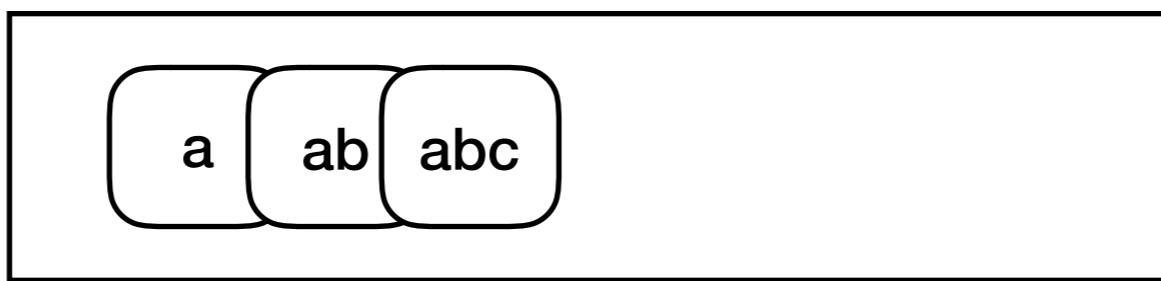
$$\mathcal{D}P_U = \{d \mid d : U \rightarrow \mathbf{R}^+\}$$

whenever $V < U$, $\mathcal{D}P$ computes the marginals of probabilities of elements of U when restricted to V

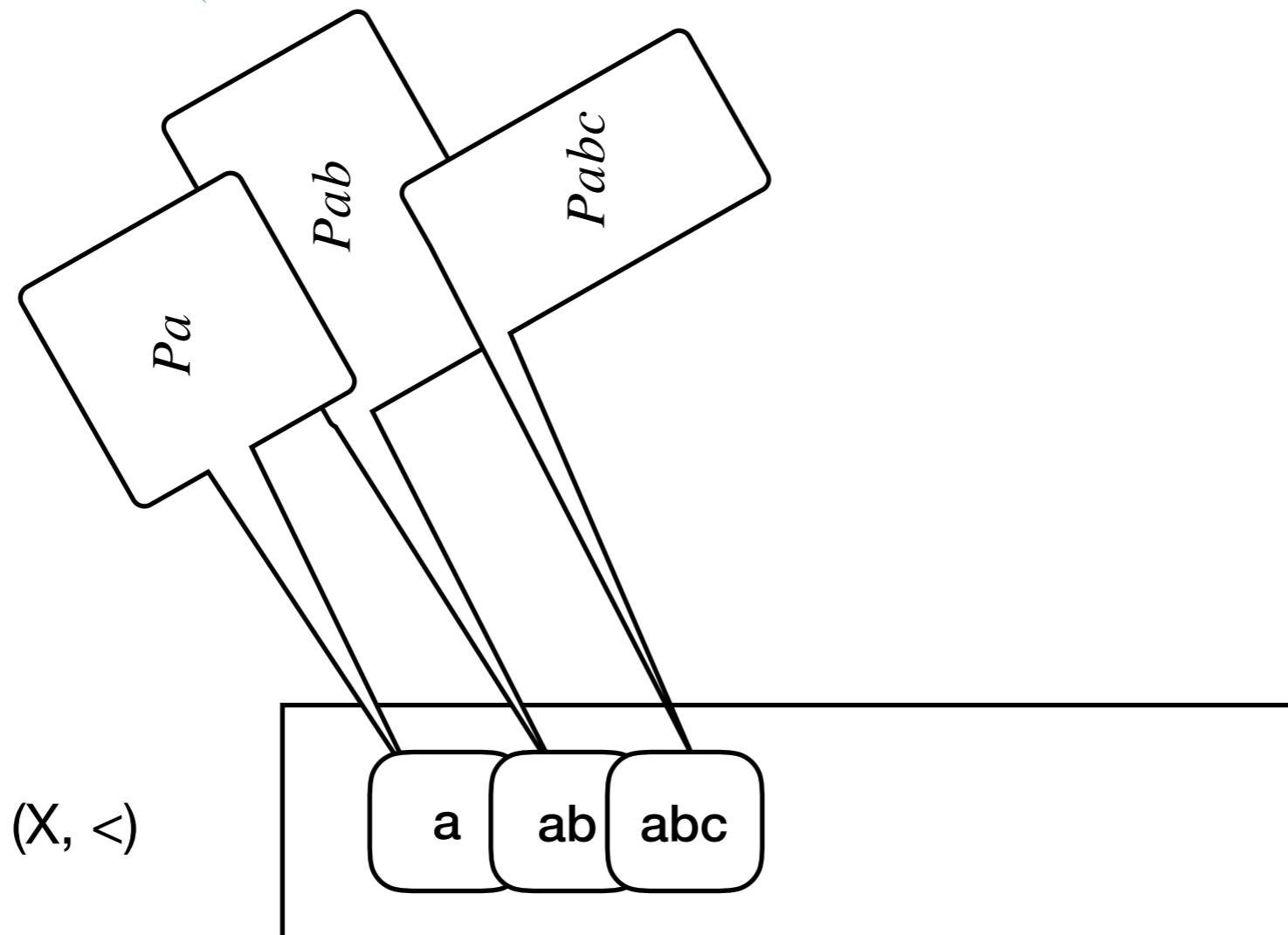
$$d_V(v) = \sum_{u \in V} d_U(u)$$

We call these presheaves, ***probabilistic event presheaves***.

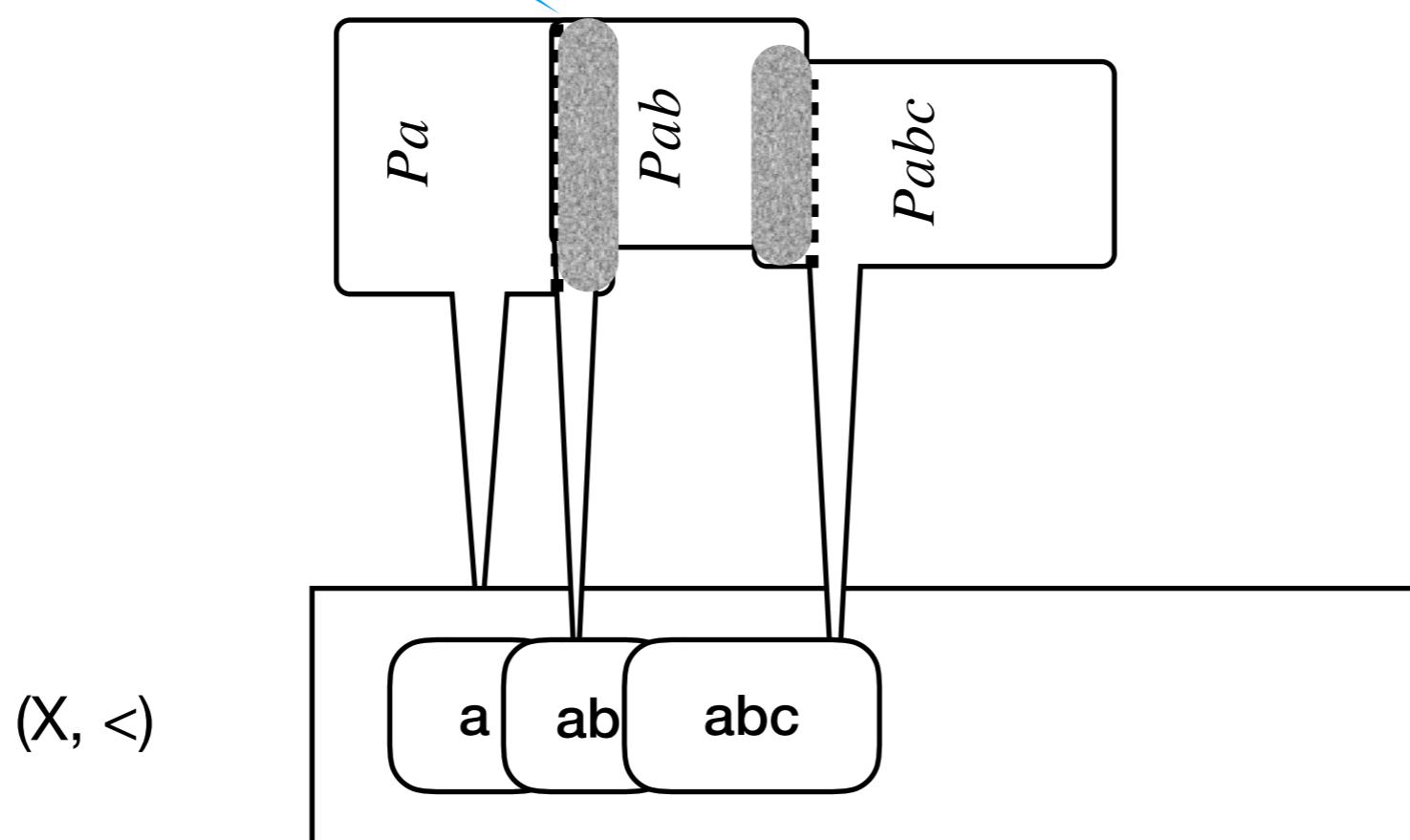
$(X, <)$



Dependency Graphs

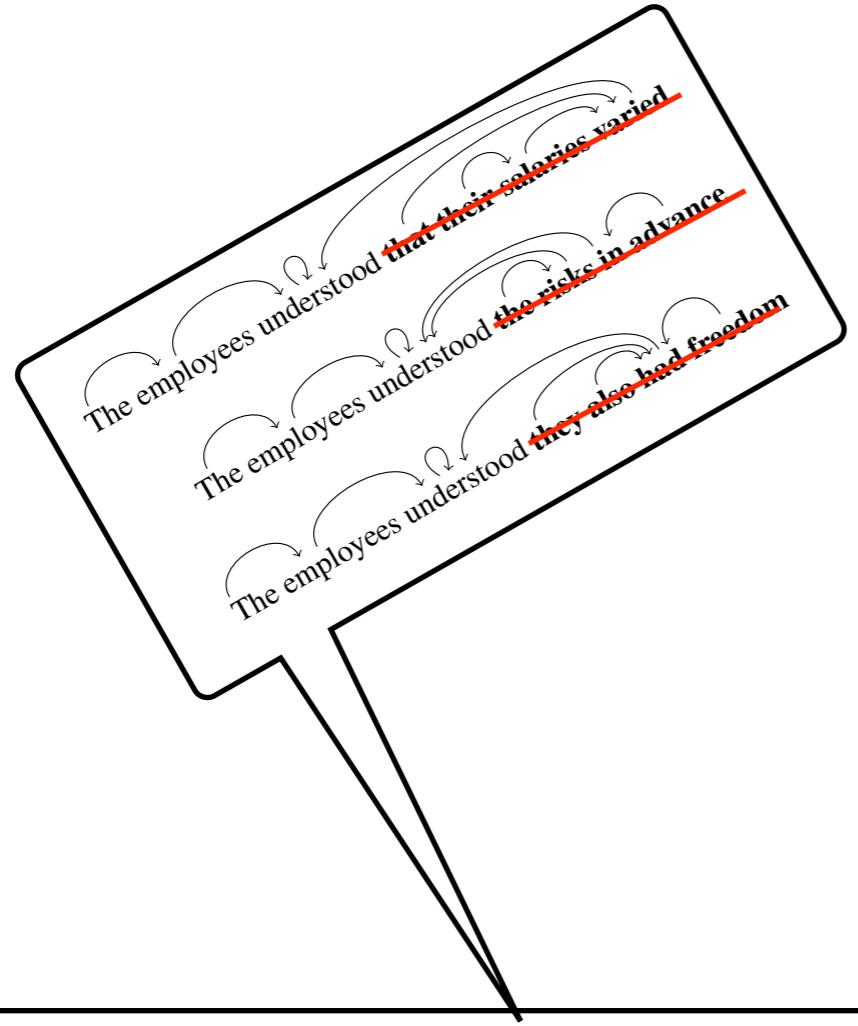


Dependency
graph of the
common prefix



Human Sentence Processing

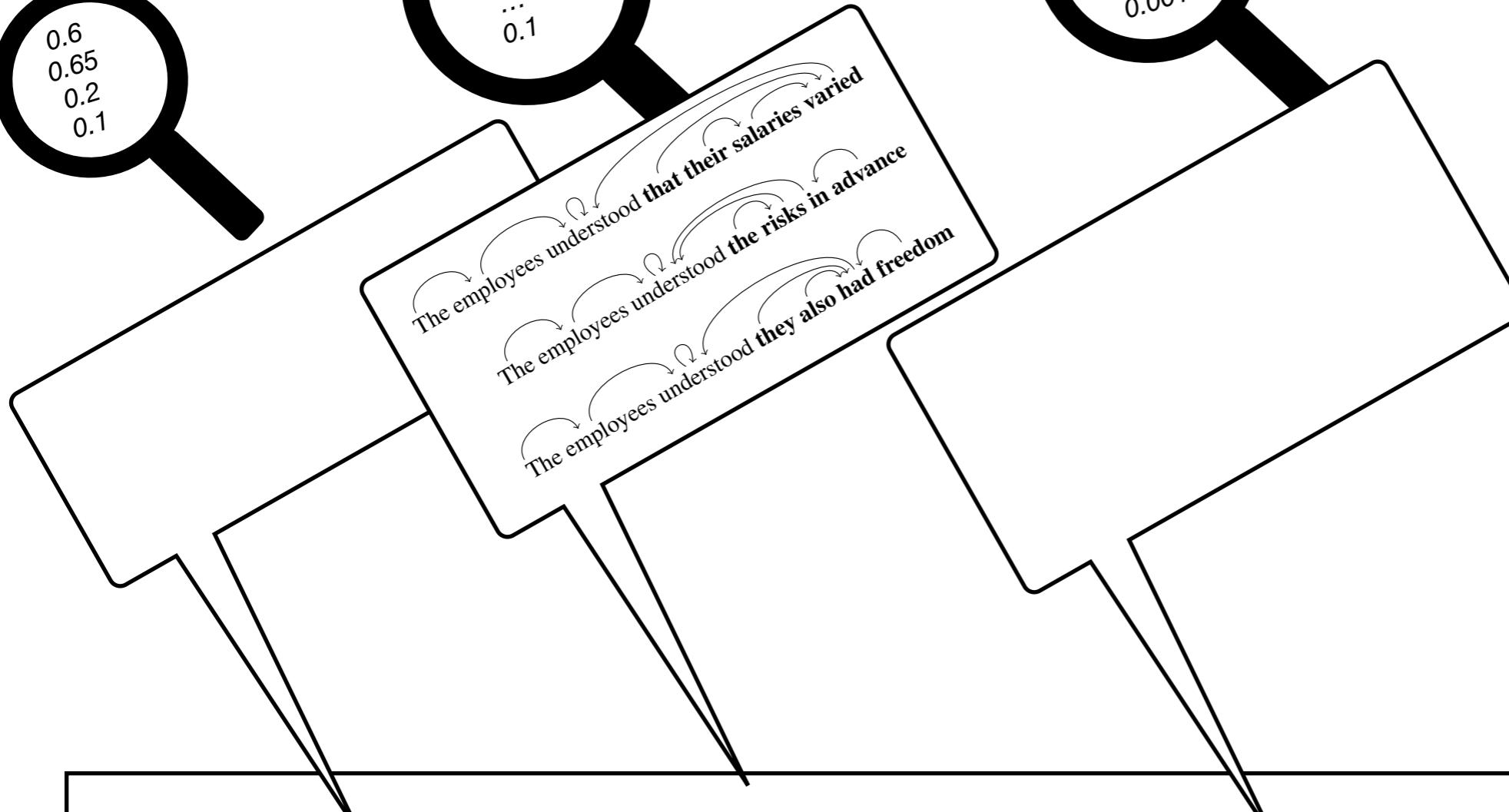
- A vocabulary:
 $\{\text{the, employees, understood, contract, change, would}\}$
- The set of phrases over this vocabulary:
 $\{\text{The, The employees, The employees understood, The employees understood the contract, ...}\}$
- The prefix relation defines the preorder:
The <The employees <The employees understood< The employees understood the contract
- The predictions over subphrases (*completions*) is the data of presheaf.
- Comprehending the subphrases by assigning *interpretations* are the events: **syntactic**, semantics, pragmatic, all of this together.



The employees

The employees understood

employees understood the report would change.



Method

The employees understood the ...



The employees understood the need for...
The employees understood the seriousness of...
The employees understood the security risks ...

The employees understood the need for ...



spaCy

The employees understood the need for ...



The employees understood the w_5 ...

$\mathcal{D}P$

Probabilities

$$d(\text{The employees understood [...] [...] [...] [...] }) = 0.80$$

$$d(\text{The employees understood [...] [...] [...] [...] }) = 0.15$$

$$d(\text{The employees understood [...] [...] [...] [...] }) = 0.05$$



Incompatibility Fraction

- The degree of *cognitive dissonance* caused by correction is computed by taking the divergence of the probabilities of interpretations of the larger subphrase (the 2nd one) restricted to the smaller one (the 1st one).

$$d_{m_1 m_2} |_{m_1} (o) = \sum_{o' \in O} d_{m_1 m_2}(oo')$$

- Different divergence can be used here:

$$\textbf{IF-min} : 1 - \sum_o \min(d_{m_1}(o), d_{m_1 m_2} |_{m_1}(o))$$

$$\textbf{IF-KL} : \text{KL}(d_{m_1} || d_{m_1 m_2} |_{m_1}) \quad KL(d_{m_1} || d_{m_1 m_2 | m_1}) = \sum_o d_{m_1}(o) \log \frac{d_{m_1}(o)}{d_{m_1 m_2 | m_1}(o)}$$

$$\textbf{IF-JS} : \text{JS}(d_{m_1} || d_{m_1 m_2} |_{m_1})$$

$$\textbf{Surprisal:} \quad SP(w_n | w_1 \dots w_{n-1})$$

Experiments

- Completions from GPT-2 and their syntactic structures from spaCy.
 - Probabilities were obtained by grouping same structures and sampling from the GPT-2 model.
 - 4 Psycholinguistic Datasets:
 - 2 from Pickering and Traxler 1998
 - 2 from Sturt and Pickering 1999
 - 48 Easy
 - 48 Hard
- Dataset 1. (i) GP. The dog catcher worried the terrier which fell wouldn't fit into the box.
- Dataset 2. (i) GP. After the judge decided the verdict of the trial caught the old man's attention.
- +
- Dataset 1. (ii) DisAmb. The dog catcher worried that the terrier which fell wouldn't fit into the box.
- Dataset 2. (ii) DisAmb. After the judge decided, the verdict of the trial caught the old man's attention.

Results

Pickering and Traxler

	Equation	ρ	p-value
IF-min	$0.0018 \times \mathbf{IF}_{\min} - 0.0776$	0.595	0.00032
IF-JS	$0.0016 \times \mathbf{IF}_{JS} - 0.1333$	0.568	0.00068
IF-KL	$0.0066 \times \mathbf{IF}_{KL} - 0.4238$	0.445	0.0106
SP	$0.7361 \times \mathbf{SP} + 268.8467$	0.356	0.045

Results

Pickering and Traxler

	All		Hard (NP/Z) GP		Easy (NP/S) GP	
Method	GPE	SE	GPE	SE	GPE	SE
IF-min	39.47	0.17	53.94	2.72	24.99	2.74
IF-JS	39.69	0.43	52.22	2.40	27.16	2.31
IF-KL	52.81	3.64	62.20	4	43.42	3.30
Surprisal	0.35	0.16	0.72	0.32	-0.02	0.05
Human First Pass	39.5		46.5		32.5	
Human Total	185.5		215.5		155.5	

Results

Sturt and Pickering 1999

	IF_{min}	IF_{JS}	SP
ρ	0.8744	0.8805	0.5536
p -value	1.99×10^{-4}	1.57×10^{-4}	0.062

	IF_{min}	IF_{JS}	SP	Human
NP/S	96.07 ± 79	163.60 ± 119.29	2.05 ± 42	87
NP/Z	137.48 ± 76	211.97 ± 99	-27.69 ± 51	400
p -value	0.0396	0.0873	0.0148	

Papers



ACL Anthology

News

FAQ

Corrections

Submissions

Github

Search...

How can large language models become more human?

Daphne Wang, Mehrnoosh Sadrzadeh, Miloš Stanojević, Wing-Yee Chow, Richard Breheny

Abstract

Psycholinguistic experiments reveal that efficiency of human language use is founded on predictions at both syntactic and lexical levels. Previous models of human prediction exploiting LLMs have used an information theoretic measure called *surprisal*, with success on naturalistic text in a wide variety of languages, but under-performance on challenging text such as garden path sentences. This paper introduces a novel framework that combines the lexical predictions of an LLM with the syntactic structures provided by a dependency parser. The framework gives rise to an *Incompatibility Fraction*. When tested on two ga

THE ROYAL SOCIETY
PUBLISHING

All Journals

Brought to you by
UCLA



Sign in

Institutional Access



Search



Home

Content

Information for

About us

Sign up for alerts RSS feeds Propose an issue

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES

Open Access

Check for updates

View PDF

Tools Share

Cite this article

Section

Abstract

1. Introduction

2. Background

Research articles

Causality and signalling of garden-path sentences

Daphne Wang and Mehrnoosh Sadrzadeh

Published: 29 January 2024 | <https://doi.org/10.1098/rsta.2023.0013>

Abstract

Sheaves are mathematical objects that describe the globally compatible data associated with open sets of a topological space. Original examples of sheaves were continuous functions; later they also became powerful tools in algebraic geometry, as well as logic and set theory. More recently, sheaves have been applied to the theory of contextuality in quantum mechanics. Whenever the local data are not necessarily compatible, sheaves are replaced by the simpler setting of presheaves. In previous work, we used presheaves to model lexically ambiguous phrases in natural language and identified the

Details References Related Figures

This Issue



18 March 2024

Volume 382, Issue 2268

Papers



ACL Anthology

News FAQ Corrections Submissions GitHub

Search...

How can large language models become more human?

Daphne Wang, Mehrnoosh Sadrzadeh, Miloš Stanojević, Wing-Yee Chow, Richard Breheny

Access | Search

Abstract

Psycholinguistic lexical levels. *surprisal*, with garden path sentences syntactic structures on two garden paths and outperformed



ACL Anthology

News FAQ Corrections Submissions GitHub

Search

On the Quantum-like Contextuality of Ambiguous Phrases

Daphne Wang, Mehrnoosh Sadrzadeh, Samson Abramsky, Victor Cervantes

Abstract

Language is contextual as meanings of words are dependent on their contexts. Contextuality is, concomitantly, a well-defined concept in quantum mechanics where it is considered a major resource for quantum computations. We investigate whether natural language exhibits any of the quantum mechanics' contextual features. We show that meaning combinations in ambiguous phrases can be modelled in the sheaf-theoretic framework for quantum contextuality, where they can become probabilistically contextual. Using the framework of Contextuality-by-Default (CbD), we explore the probabilistic variants of these and show that CbD-contextuality is also possible.

Anthology ID: 2021.sem space-1.5

Volume: Proceedings of the 2021 Workshop on Semantic Spaces at the Intersection of NLP, Physics, and Cognitive Science (SemSpace)

Month: June

Year: 2021

Address: Groningen, The Netherlands

Editors: Martha Lewis, Mehrnoosh Sadrzadeh

Venue: SemSpace

SIG: SIGSEM

Publisher: Association for Computational Linguistics

Note: -

tic and led such as th the n tested den path,

of contextuality compatible, sheaves used d identified the

18 March 2024
Volume 382, Issue 2268

References Related Figures

This Issue



Papers



ACL Anthology

News

FAQ

Corrections

Submissions

Github

Search...

How can large language models become more human?

Daphne Wang, Mehrnoosh Sadrzadeh, Miloš Stanojević, Wing-Yee Chow, Richard Breheny

Access | Search

Abstract

Psycholinguistic lexical levels. *surprisal*, with



ACL Anthology

News

FAQ

Corrections

Submissions

Github

Search

On the Quantum-like Contextuality of Ambiguous Phrases

tic and
led
t such as

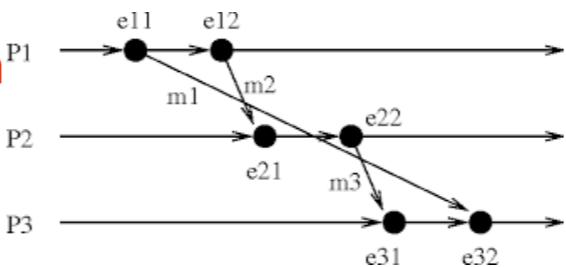
References Related Figures

The 4th workshop Quantum Contextuality in Quantum Mechanics and Beyond (QCQMB)

VIRTUAL MEETING: May 17-21, 11:50 am - 7 pm UTC/GMT (7:50 am - 3 pm EDT), 2021



Future Work

- **Summary:** Combining Psycholinguistics theories with LLMs makes them more natural.
- Better Syntax: incremental probabilities of parsers rather than the extravaganza!
- Beyond syntax: a theory of interpretation/comprehension, e.g. adding **semantics and pragmatics**
- SP vs IF, SP with IF
- Large NLP datasets: Huang et. al 2023.
- More reliable human reading data: FMRI, OVP
- **Reanalysis using causal orders or Bayesian**
- Connections to **QM** (Kin Ian Lo QTM 2024, QPL 2023, Tilen Stokin-Lambik)