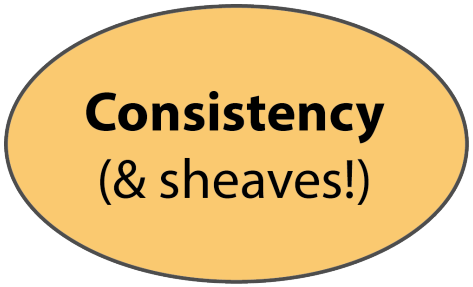
Causality and Signalling in Garden-Path Sentences

Samson Abramsky on Logic and Structure in Computer Science and Beyond

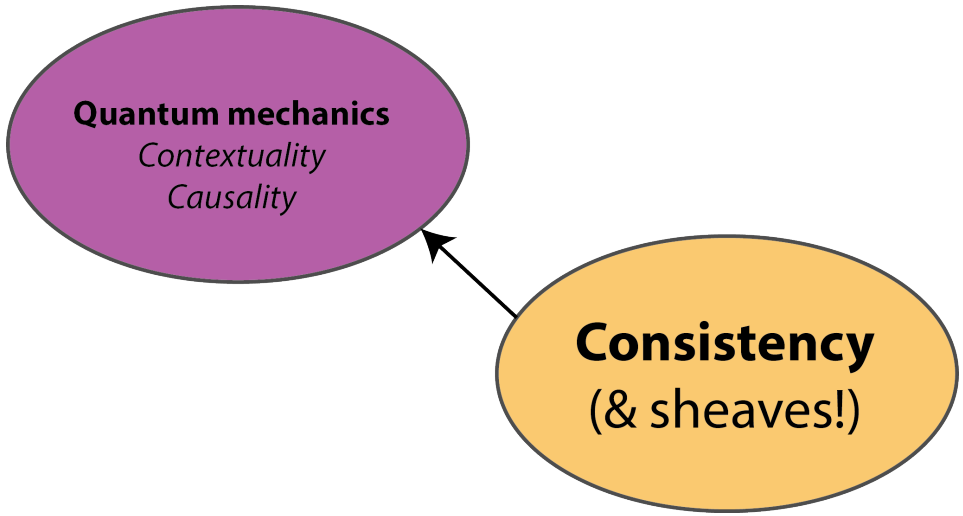
Daphne Wang & Mehrnoosh Sadrzadeh

20th September 2023

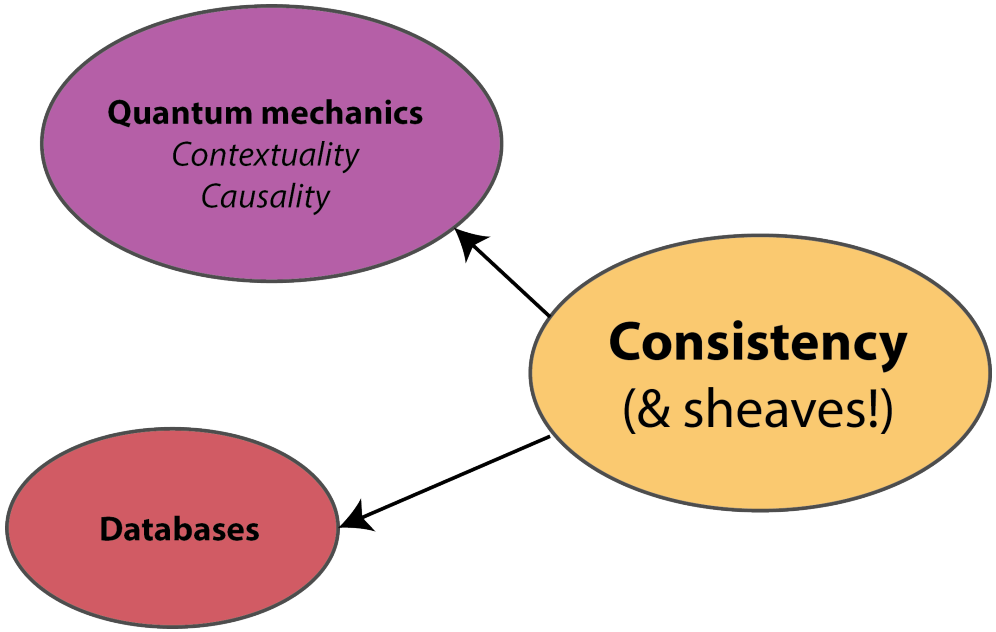
Consistency everywhere



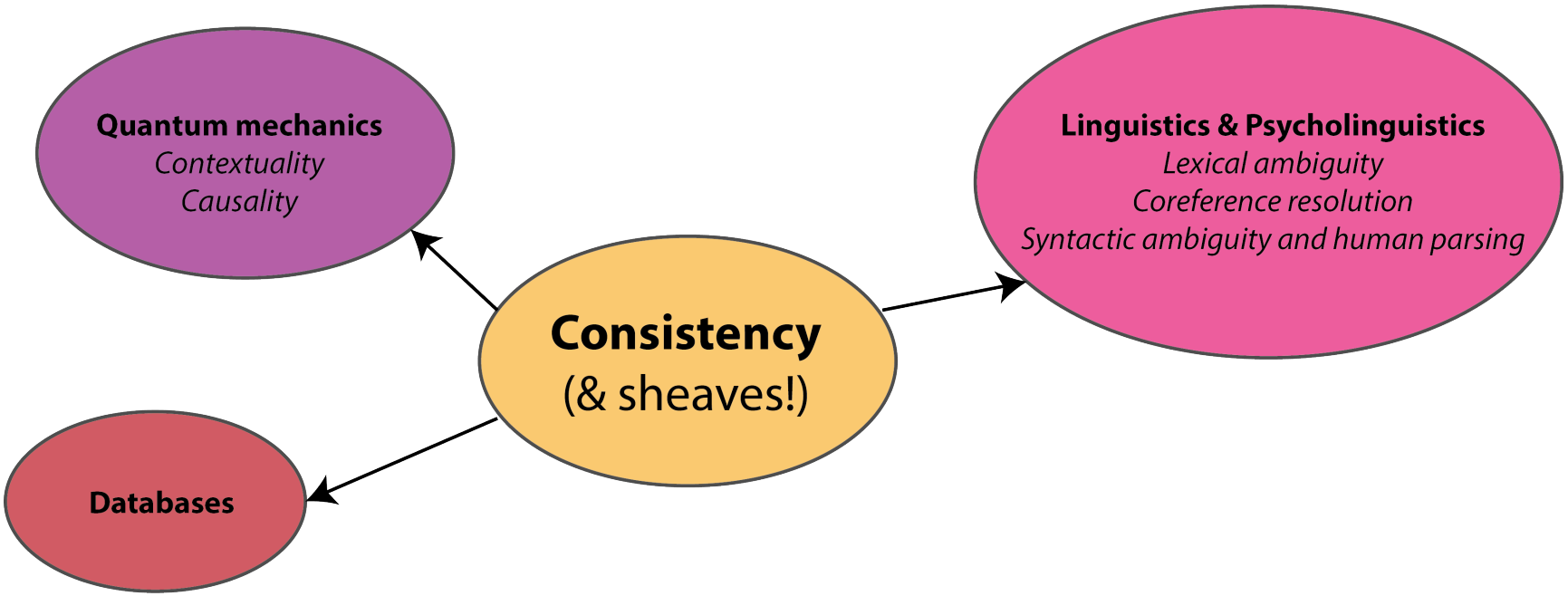
Consistency everywhere



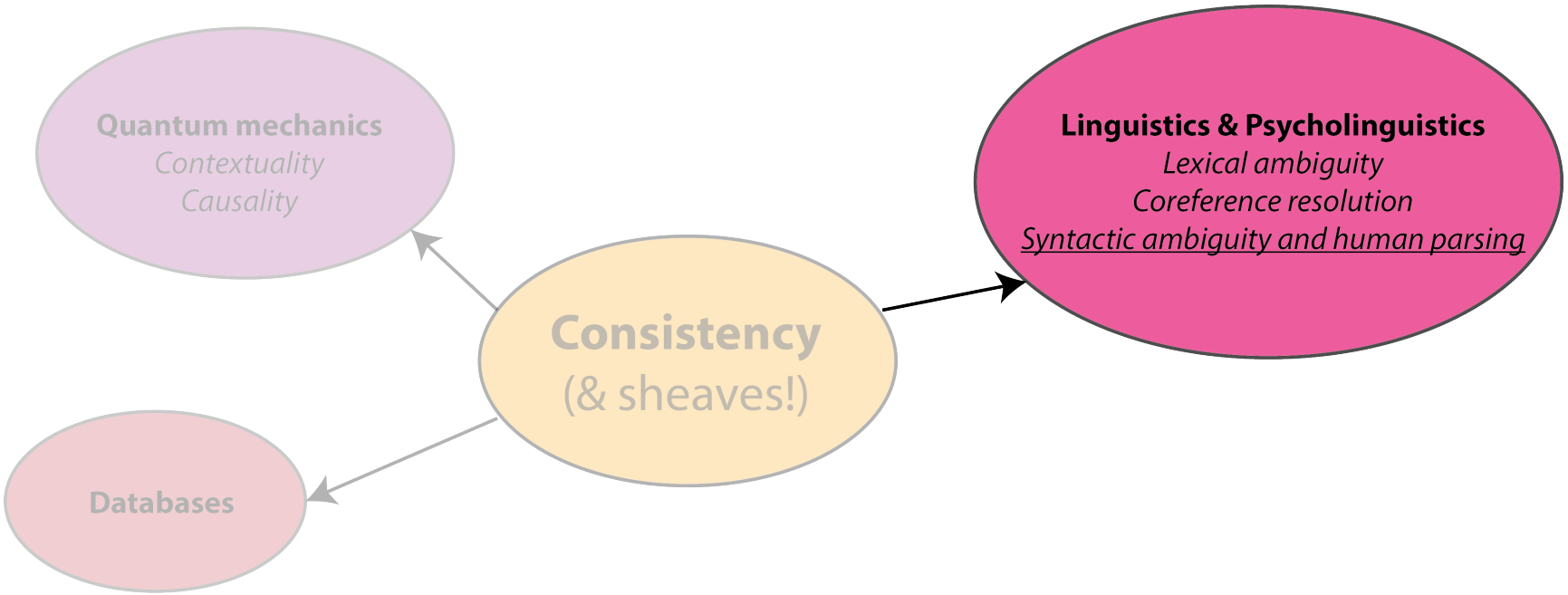
Consistency everywhere



Consistency everywhere



Consistency everywhere



Garden-path sentences and human parsing

Garden-path sentences

Definition

A *garden-path* sentence is a sentence which is grammatically correct and unambiguous, but which forces the reader to initially adopt a wrong syntactic parse.

Garden-path sentences

Definition

A *garden-path* sentence is a sentence which is grammatically correct and unambiguous, but which forces the reader to initially adopt a wrong syntactic parse.

Examples

The employees understood the contract would change

Garden-path sentences

Definition

A *garden-path* sentence is a sentence which is grammatically correct and unambiguous, but which forces the reader to initially adopt a wrong syntactic parse.

Examples

The employees understood the contract would change

Garden-path sentences

Definition

A *garden-path* sentence is a sentence which is grammatically correct and unambiguous, but which forces the reader to initially adopt a wrong syntactic parse.

Examples

The employees understood the contract would change

Because the employees negotiated the contract would change

Garden-path sentences

Definition

A *garden-path* sentence is a sentence which is grammatically correct and unambiguous, but which forces the reader to initially adopt a wrong syntactic parse.

Examples

The employees understood the contract would change

Because the employees negotiated the contract would change

Garden-path sentences

Definition

A *garden-path* sentence is a sentence which is grammatically correct and unambiguous, but which forces the reader to initially adopt a wrong syntactic parse.

Examples

NP/S The employees understood the contract would change

Because the employees negotiated the contract would change

Garden-path sentences

Definition

A *garden-path* sentence is a sentence which is grammatically correct and unambiguous, but which forces the reader to initially adopt a wrong syntactic parse.

Examples

NP/S

The employees understood the contract would change

Because the employees negotiated the contract would change

Garden-path sentences

Definition

A *garden-path* sentence is a sentence which is grammatically correct and unambiguous, but which forces the reader to initially adopt a wrong syntactic parse.

Examples

NP/S The employees understood the contract would change

Because the employees negotiated the contract would change

Garden-path sentences

Definition

A *garden-path* sentence is a sentence which is grammatically correct and unambiguous, but which forces the reader to initially adopt a wrong syntactic parse.

Examples

NP/S The employees understood the contract would change

NP/Z Because the employees negotiated the contract would change

Garden-path sentences

Definition

A *garden-path* sentence is a sentence which is grammatically correct and unambiguous, but which forces the reader to initially adopt a wrong syntactic parse.

Examples

NP/S The employees understood the contract would change

NP/Z Because the employees negotiated the contract would change

Garden-path sentences

Definition

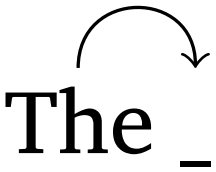
A *garden-path* sentence is a sentence which is grammatically correct and unambiguous, but which forces the reader to initially adopt a wrong syntactic parse.

Examples

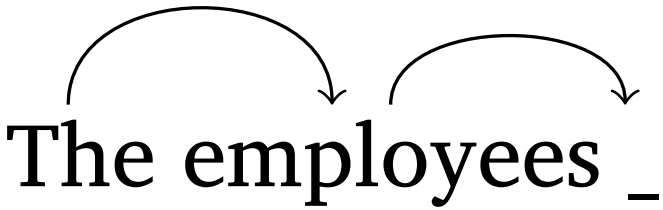
NP/S The employees understood the contract would change

NP/Z Because the employees negotiated the contract would change

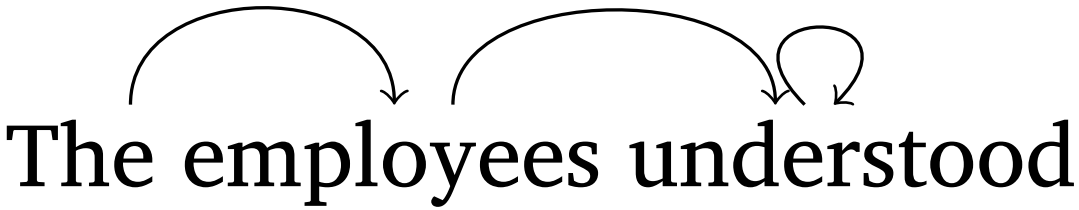
The psycholinguistics theories



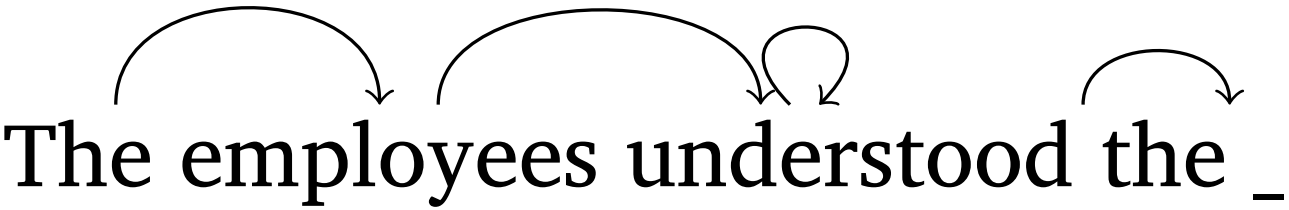
The psycholinguistics theories



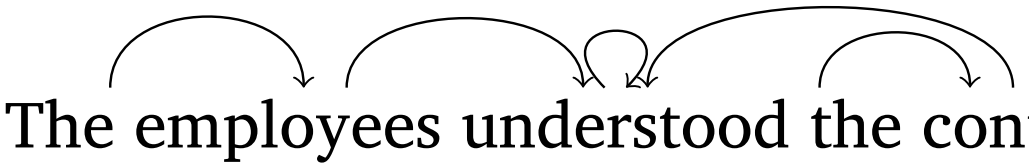
The psycholinguistics theories

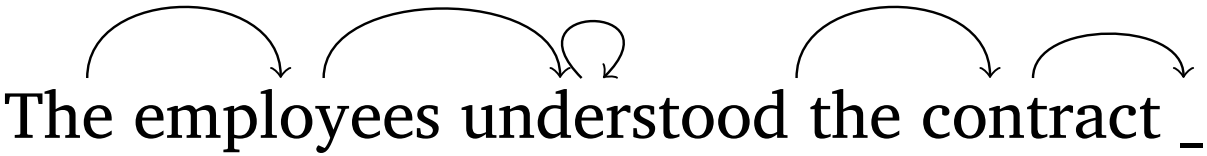


The psycholinguistics theories



The psycholinguistics theories



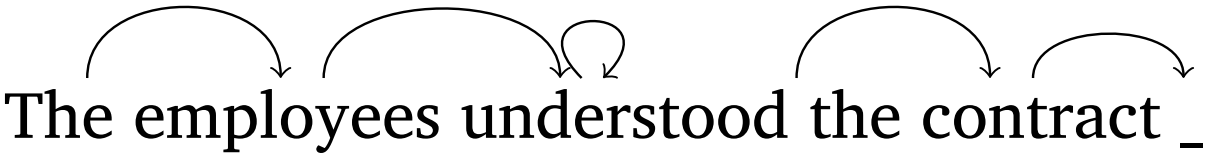


The psycholinguistics theories

very likely

⇓

easy to access



The psycholinguistics theories

very likely

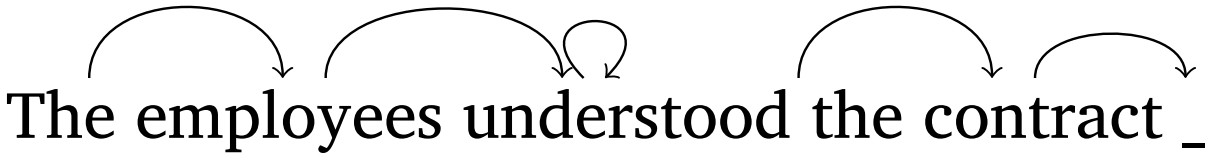
⇓

easy to access

unlikely

⇓

hard to access



The psycholinguistics theories

very likely

⇓

easy to access

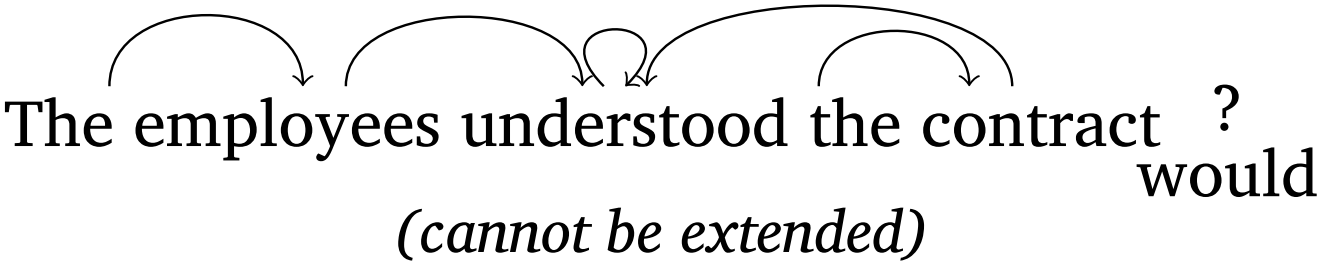
unlikely

⇓

hard to access

The psycholinguistics theories

impossible



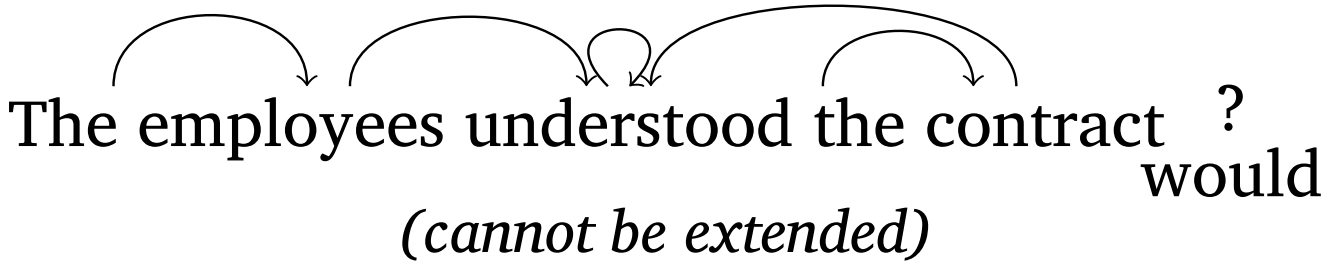
unlikely

⇓

hard to access

The psycholinguistics theories

impossible

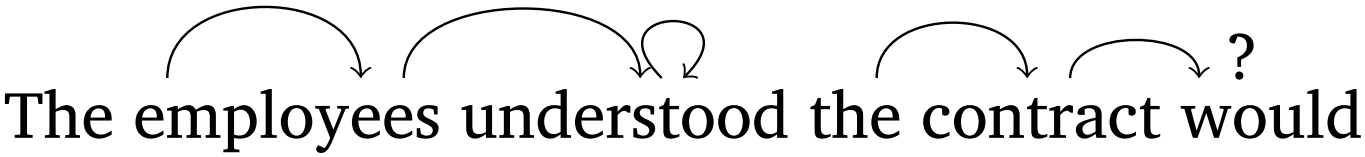


(cannot be extended)

unlikely

⇓

hard to access



The psycholinguistics theories

impossible

(cannot be extended)

unlikely

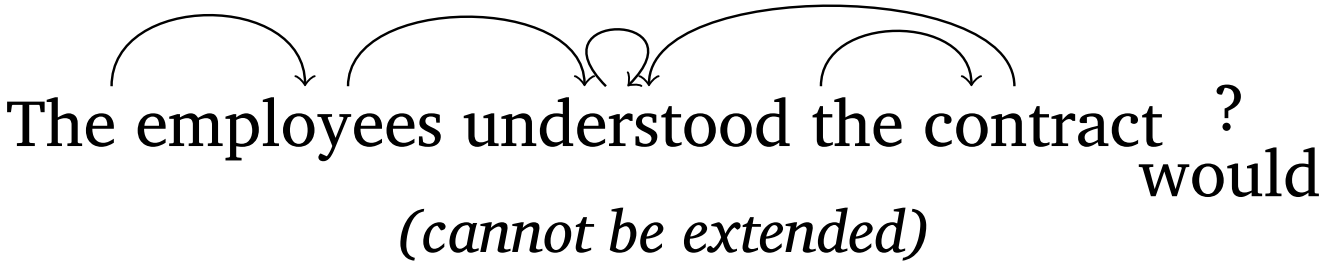
⇓

hard to access

(can be extended)

The psycholinguistics theories

impossible



(cannot be extended)

has to be adopted

(can be extended)

The psycholinguistics theories

impossible

(cannot be extended)

has to be adopted

⇧

Reanalysis

(can be extended)

Computational models of human behaviour

Predictions in human behaviours

- Words that are predictable (in context) are read more easily

Predictions in human behaviours

- Words that are predictable (in context) are read more easily

Example:

1. The boat passed easily under the …
2. Rita slowly walked down the shaky …

Predictions in human behaviours

- Words that are predictable (in context) are read more easily

Example:

1. The boat passed easily under the **bridge**
2. Rita slowly walked down the shaky **bridge**

Predictions in human behaviours

- Words that are predictable (in context) are read more easily

Example:

1. The boat passed easily under the **bridge**
2. Rita slowly walked down the shaky **bridge**

("*bridge*" in 1. read faster than "*bridge*" in 2.)

Computational models based on surprisal

Definition

The surprisal of a word w_n in the context $w_1 \dots w_{n-1}$ is given by:

$$S(w_n) = -\log_2 \left(P\left[w_n \mid w_1 \dots w_{n-1}\right]\right)$$

Computational models based on surprisal

Definition

The surprisal of a word w_n in the context $w_1 \dots w_{n-1}$ is given by:

$$S(w_n) = -\log_2 \left(P\left[w_n \mid w_1 \dots w_{n-1}\right]\right)$$

Example: in 1. the surprisal of **bridge** is:
$-\log_2 \left(P\left[\text{bridge} \mid \text{The boat passed easily under the}\right]\right)$

Computational models based on surprisal

Definition

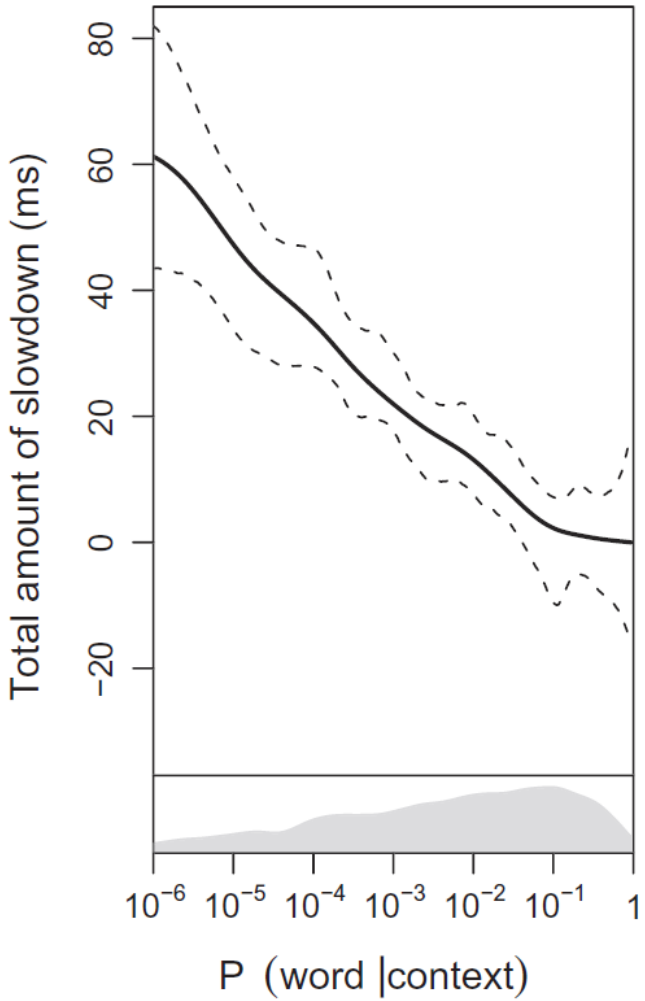
The surprisal of a word w_n in the context $w_1 \ldots w_{n-1}$ is given by:

$$S(w_n) = -\log_2 \left(P\left[w_n \mid w_1 \ldots w_{n-1}\right]\right)$$

Example: in 1. the surprisal of **bridge** is:
$-\log_2 \left(P\left[\text{bridge} \mid \text{The boat passed easily under the}\right]\right)$

- Works well in naturalistic sources (e.g. newspapers, novels, etc.)

Computational models based on surprisal



Computational models based on surprisal

Definition

The surprisal of a word w_n in the context $w_1 \ldots w_{n-1}$ is given by:

$$S(w_n) = -\log_2 \left(P\left[w_n \mid w_1 \ldots w_{n-1}\right]\right)$$

Example: in 1. the surprisal of **bridge** is:
$-\log_2 \left(P\left[\text{bridge} \mid \text{The boat passed easily under the}\right]\right)$

- Works well in naturalistic sources (e.g. newspapers, novels, etc.)
- Fails for rarer constructions

Computational models based on surprisal

Definition

The surprisal of a word w_n in the context $w_1 \ldots w_{n-1}$ is given by:

$$S(w_n) = -\log_2 \left(P\left[w_n \mid w_1 \ldots w_{n-1}\right]\right)$$

Example: in 1. the surprisal of **bridge** is:
$-\log_2 \left(P\left[\text{bridge} \mid \text{The boat passed easily under the}\right]\right)$

- Works well in naturalistic sources (e.g. newspapers, novels, etc.)
- Fails for rarer constructions
  e.g. **garden-path sentences**

Issues with surprisal

Failures of surprisal

Issues with surprisal

Failures of surprisal

- Under-estimates the slowdown

Issues with surprisal

Failures of surprisal

- Under-estimates the slowdown

	Prediction (ms)	Observed(ms)
NP/S	24	87
NP/Z	30	400

Issues with surprisal

Failures of surprisal

- Under-estimates the slowdown

	Prediction (ms)	Observed(ms)
NP/S	24	87
NP/Z	30	400

- Cannot see the difference between NP/S and NP/Z

Issues with surprisal

Failures of surprisal

- Under-estimates the slowdown

	Prediction (ms)	Observed(ms)
NP/S	24	87
NP/Z	30	400

- Cannot see the difference between NP/S and NP/Z (sometimes even predicts opposite trends)

Issues with surprisal

Failures of surprisal

- Under-estimates the slowdown

	Prediction (ms)	Observed(ms)
NP/S	24	87
NP/Z	30	400

- Cannot see the difference between NP/S and NP/Z (sometimes even predicts opposite trends)

Possible reasons

Issues with surprisal

Failures of surprisal

- Under-estimates the slowdown

	Prediction (ms)	Observed(ms)
NP/S	24	87
NP/Z	30	400

- Cannot see the difference between NP/S and NP/Z (sometimes even predicts opposite trends)

Possible reasons

- Grammatical structure is implicit

Issues with surprisal

Failures of surprisal

- Under-estimates the slowdown

	Prediction (ms)	Observed(ms)
NP/S	24	87
NP/Z	30	400

- Cannot see the difference between NP/S and NP/Z (sometimes even predicts opposite trends)

Possible reasons

- Grammatical structure is implicit
  (Do large language models know about grammar?)

Issues with surprisal

Failures of surprisal

- Under-estimates the slowdown

	Prediction (ms)	Observed(ms)
NP/S	24	87
NP/Z	30	400

- Cannot see the difference between NP/S and NP/Z (sometimes even predicts opposite trends)

Possible reasons

- Grammatical structure is implicit
  (Do large language models know about grammar?)
- Consistent with parallelism, but that is also implicit

Issues with surprisal

Failures of surprisal

- Under-estimates the slowdown

	Prediction (ms)	Observed(ms)
NP/S	24	87
NP/Z	30	400

- Cannot see the difference between NP/S and NP/Z (sometimes even predicts opposite trends)

Possible reasons

- Grammatical structure is implicit
  (Do large language models know about grammar?)
- Consistent with parallelism, but that is also implicit
- Cannot deal with reanalysis

The presheaf model

Definition of the model

$\mathscr{C}_1$ $\qquad$ The $\xrightarrow{\leq}$ The employees

Definition of the model

$\mathscr{C}_2$ The employees $\xrightarrow{\leq}$ The employees understood

Definition of the model

$\mathscr{C}_3$ The employees understood $\overset{\leq}{\longrightarrow}$ The employees understood the

Definition of the model

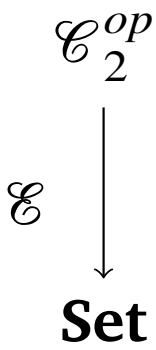
$\mathscr{C}_2$ The employees $\xrightarrow{\leq}$ The employees understood

Definition of the model

$\mathscr{C}_2^{op}$

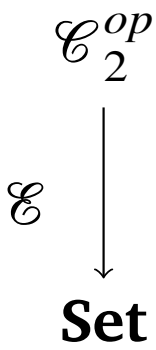
$$\text{The employees} \xleftarrow{\leq} \text{The employees understood}$$

Definition of the model



The employees $\xleftarrow{\leq}$ The employees understood

Definition of the model



$$\text{The employees} \xleftarrow{\leq} \text{The employees understood}$$

$$\mathscr{E}(\text{The employees understood}) = \{ \text{The employees understood}\ ,\ \text{The employees understood}\ \_\ ,\ \text{The employees understood}\ ,\ \ldots \}$$

Definition of the model

$$\begin{array}{ccc} \mathscr{C}_2^{op} & & \text{The employees} \xleftarrow{\leq} \text{The employees understood} \\ \mathscr{E} \downarrow & & \\ \textbf{Set} & & \end{array}$$

spaCy

$$\mathscr{E}(\text{The employees understood}) = \{ \text{The employees understood}\ ,\ \text{The employees understood}\ \_\ ,\ \text{The employees understood}\ ,\ \ldots \}$$

Definition of the model

$$\mathscr{C}_2^{op} \xrightarrow{\mathscr{E}} \textbf{Set}$$

$$\text{The employees} \xleftarrow{\leq} \text{The employees understood}$$

$$\mathscr{E}(\text{The employees understood}) = \{\ \overset{\frown}{\text{The employees understood}}\ ,\ \overset{\frown}{\text{The employees understood}}\ \_\ ,\ \overset{\frown}{\text{The employees understood}}\ , \ldots \}$$

$$\mathscr{E}(\text{The employees understood} \to \text{The employees}) : \mathscr{E}(\text{The employees understood}) \to \mathscr{E}(\text{The employees})$$

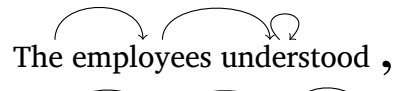
Definition of the model

$\mathscr{C}_2^{op}$

$\mathscr{E}$ ↓

Set

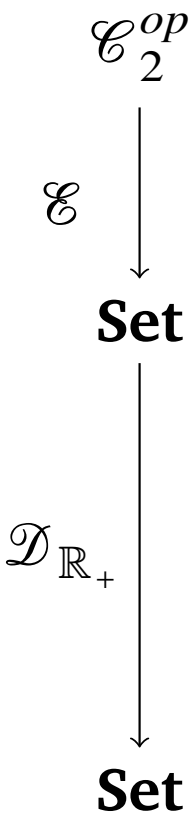
The employees $\xleftarrow{\leq}$ The employees understood

$\mathscr{E}$(The employees understood) = {  ,  ,  , … }

$\mathscr{E}$(The employees understood→The employees) : $\mathscr{E}$(The employees understood) → $\mathscr{E}$(The employees)

e.g.  ↦ 

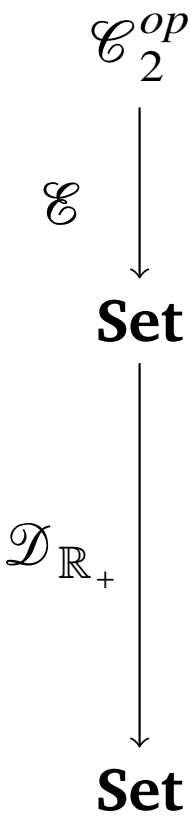
Definition of the model



$$\text{The employees} \xleftarrow{\leq} \text{The employees understood}$$

$$\mathscr{E}(\text{The employees understood}) = \{ \text{The employees understood}\ ,\ \text{The employees understood}\ \_\ ,\ \text{The employees understood}\ , \ldots \}$$

Definition of the model



The employees $\xleftarrow{\leq}$ The employees understood

$\mathscr{E}$(The employees understood) = { The employees understood , The employees understood _ , The employees understood , … }

$\mathscr{D}_{\mathbb{R}_+}(A) := \{\text{probability distributions over } A\}$

Definition of the model

$\mathscr{C}_2^{op}$

$\mathscr{E}$ ↓

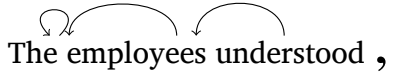
Set

$\mathscr{D}_{\mathbb{R}_+}$ ↓

Set

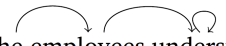
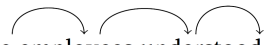
The employees $\xleftarrow{\leq}$ The employees understood

$\mathscr{E}$(The employees understood) = { 



$\mathscr{D}_{\mathbb{R}_+}(A) := \{\text{probability distributions over } A\}$

e.g.

Parse	Probability
 The employees understood	0.75
 The employees understood _	0.20
 The employees understood	0.05

$\in \mathscr{D}_{\mathbb{R}_+}\mathscr{E}$(The employees understood)

Definition of the model

$\mathscr{C}_2^{op}$

$\mathscr{E}$ ↓

Set

$\mathscr{D}_{\mathbb{R}_+}$ ↓

Set

The employees $\xleftarrow{\leq}$ The employees understood

$\mathscr{E}$(The employees understood) = { The employees understood , The employees understood _ , The employees understood , … }

$\mathscr{D}_{\mathbb{R}_+}(A) := \{\text{probability distributions over } A\}$

e.g.

Parse	Probability
The employees understood	0.75
The employees understood _	0.20
The employees understood	0.05

$\in \mathscr{D}_{\mathbb{R}_+}\mathscr{E}$(The employees understood)

Definition of the model

$\mathscr{C}_2^{op}$

$\mathscr{E}$ ↓

Set

$\mathscr{D}_{\mathbb{R}_+}$ ↓

Set

The employees $\xleftarrow{\leq}$ The employees understood

$\mathscr{E}$(The employees understood) = { The employees understood , The employees understood _ , The employees understood , … }

$\mathscr{D}_{\mathbb{R}_+}(A) := \{\text{probability distributions over } A\}$

$\mathscr{D}_{\mathbb{R}_+}(A \xrightarrow{f} B) :: d \mapsto d|_B \text{ s.t. } d|_B(b) = \sum_{f(a)=b} d(a)$

The signalling fraction

Consistency (General case)

A B

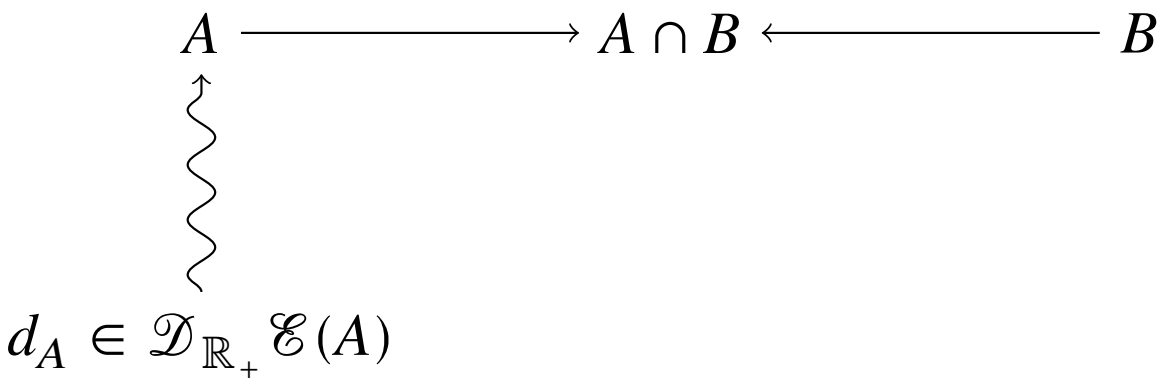
The signalling fraction

Consistency (General case)

$$A \longrightarrow A \cap B \longleftarrow B$$

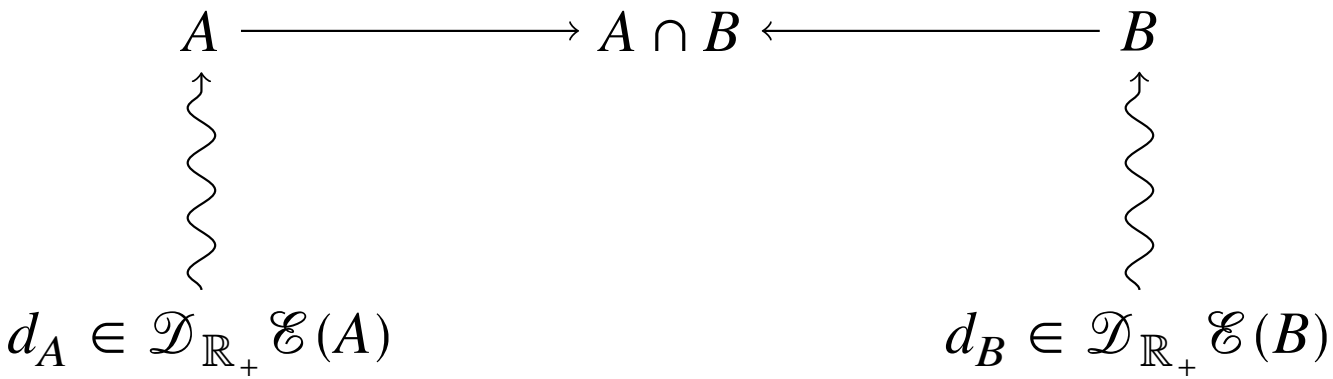
The signalling fraction

Consistency (General case)



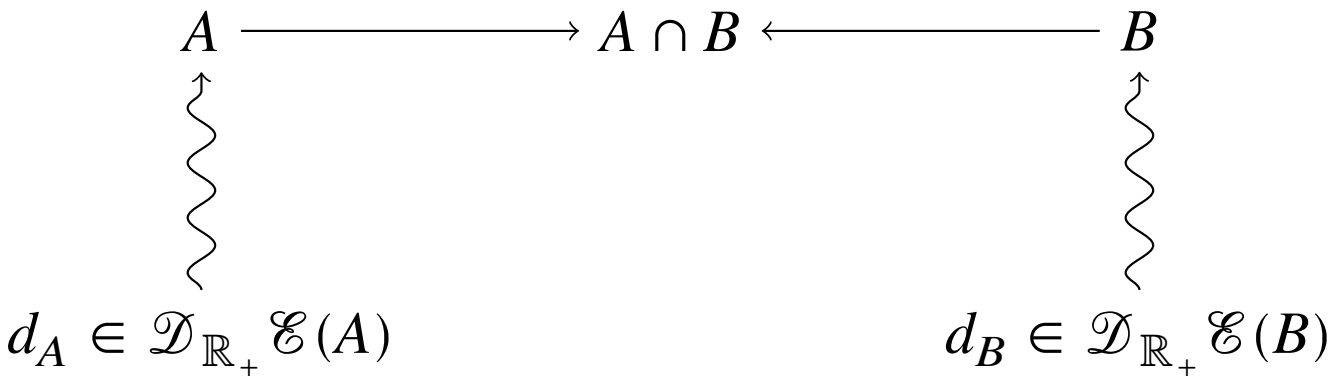
The signalling fraction

Consistency (General case)



The signalling fraction

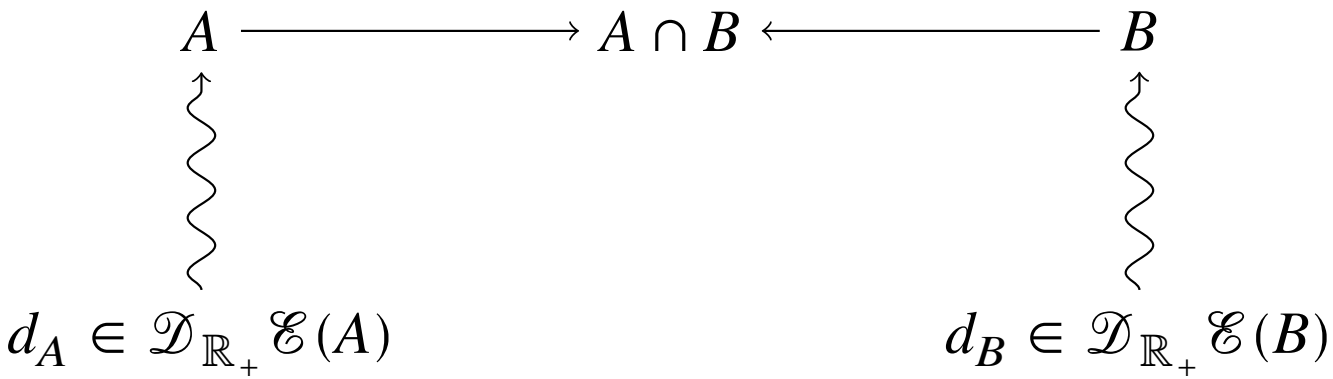
Consistency (General case)



$$d = \{d_A, d_B\} \text{ consistent} \iff d_A|_{A\cap B} = d_B|_{A\cap B}$$

The signalling fraction

Consistency (General case)



$d = \{d_A, d_B\}$ consistent $\iff$ $d_A|_{A\cap B} = d_B|_{A\cap B}$
(sheaf condition)

The signalling fraction

The Signalling Fraction (SF) (General case)

The signalling fraction

The Signalling Fraction (SF) (General case)

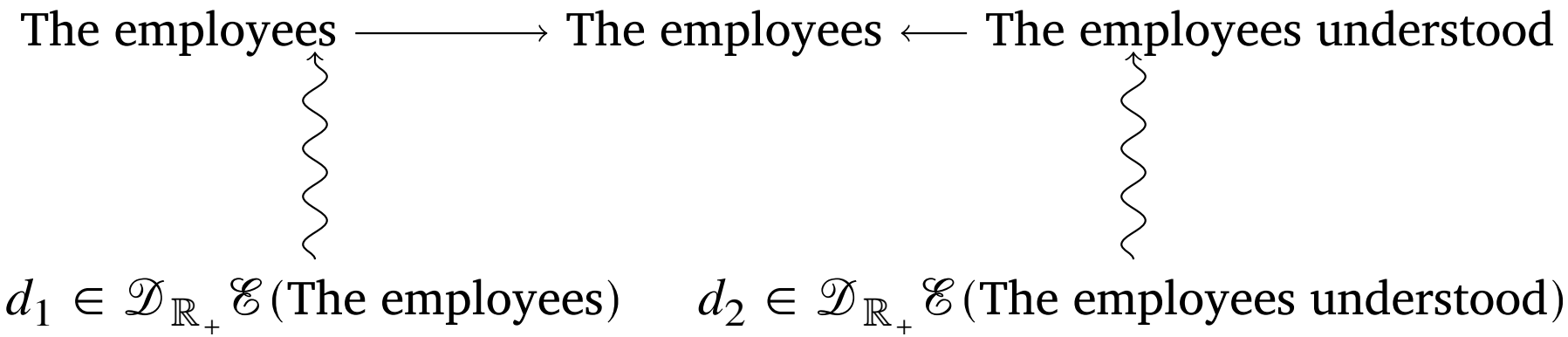
Definition

Given $d = \{d_1, d_2\}$, we define SF to the minimal $\lambda \in [0, 1]$ s.t. $\exists d_{NS}$ consistent and d' (not necessarily consistent) s.t.:

$$d = (1 - \lambda) \cdot d_{NS} + \lambda \cdot d'$$

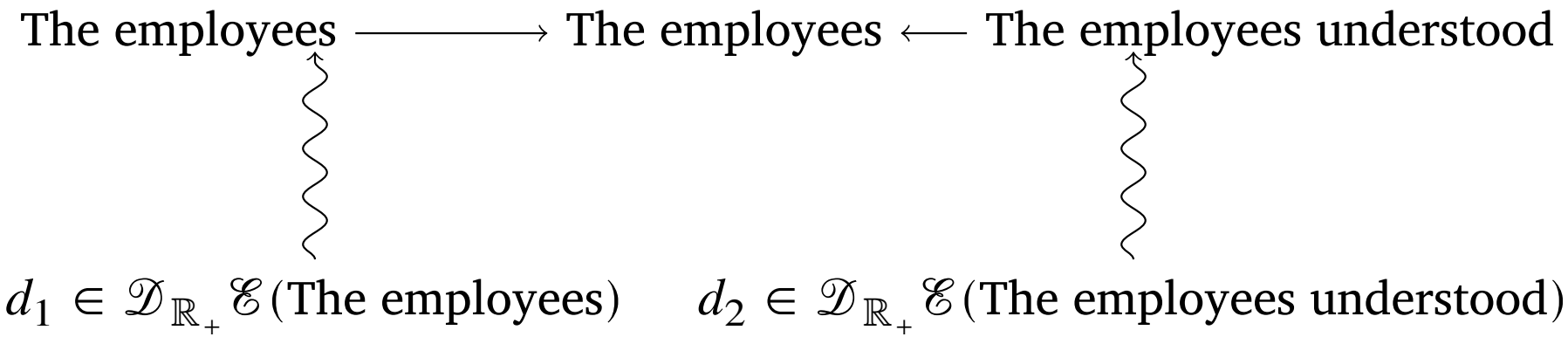
The signalling fraction

The Signalling Fraction (SF) (Example)



The signalling fraction

The Signalling Fraction (SF) (Example)



SF is the minimal λ s.t.:

$$d = (1 - \lambda) \cdot d_{NS} + \lambda \cdot d'$$

where: $d_{NS,\text{The employees understood}}\big|_{\text{The employees}} = d_{NS,\text{The employees}}$

The signalling fraction

Interpretation of SF

Low SF (mostly consistent)

The signalling fraction

Interpretation of SF

Low SF (mostly consistent) $\Longrightarrow$ Low need for reanalysis

The signalling fraction

Interpretation of SF

Low SF (mostly consistent) $\Longrightarrow$ Low need for reanalysis

High SF (highly inconsistent)

The signalling fraction

Interpretation of SF

Low SF (mostly consistent) $\Longrightarrow$ Low need for reanalysis

High SF (highly inconsistent) $\Longrightarrow$ High need for reanalysis

The signalling fraction

Interpretation of SF

Low SF (mostly consistent) $\Longrightarrow$ Low need for reanalysis

High SF (highly inconsistent) $\Longrightarrow$ High need for reanalysis

$\Longrightarrow$ Should have a correlation between reading difficulty (reading times) and SF

Comparison with surprisal

	Surprisal	SF

Comparison with surprisal

	Surprisal	SF
Forward-looking/Predictive		

Comparison with surprisal

	Surprisal	SF
Forward-looking/Predictive	Yes	Yes

Comparison with surprisal

	Surprisal	SF
Forward-looking/Predictive	Yes	Yes
Uses grammatical structure		

Comparison with surprisal

	Surprisal	SF
Forward-looking/Predictive	Yes	Yes
Uses grammatical structure	?	Yes

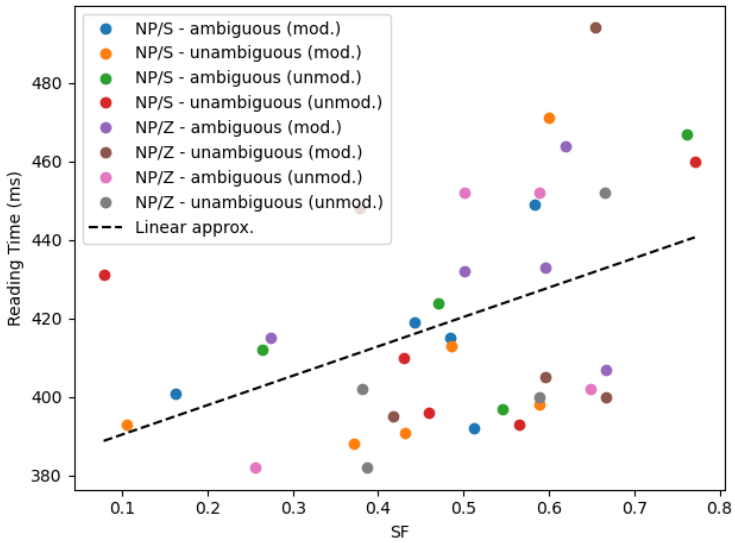
Comparison with surprisal

	Surprisal	SF
Forward-looking/Predictive	Yes	Yes
Uses grammatical structure	?	Yes
Parallel model		

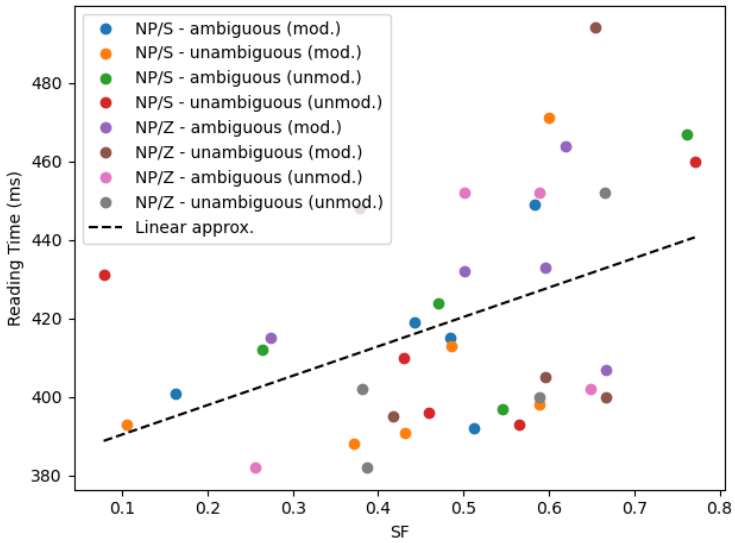
Comparison with surprisal

	Surprisal	SF
Forward-looking/Predictive	Yes	Yes
Uses grammatical structure	?	Yes
Parallel model	?	Yes

Empirical Results - Correlation

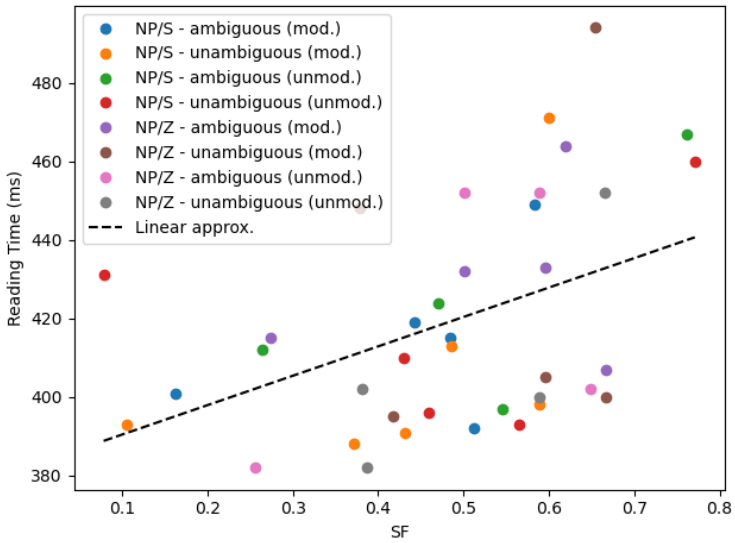


Empirical Results - Correlation



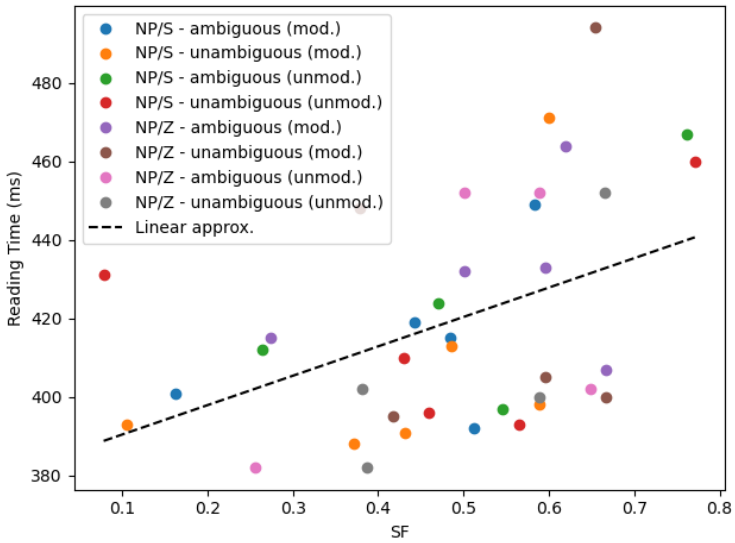
- Pearson's ρ: 0.78

Empirical Results - Correlation



- Pearson's ρ: 0.78
- p-value: 0.0004

Empirical Results - Correlation



- Pearson's ρ: 0.78
- p-value: 0.0004
- $RT \simeq 75\text{SF} + 383\ ms$

Empirical Results - Predictions

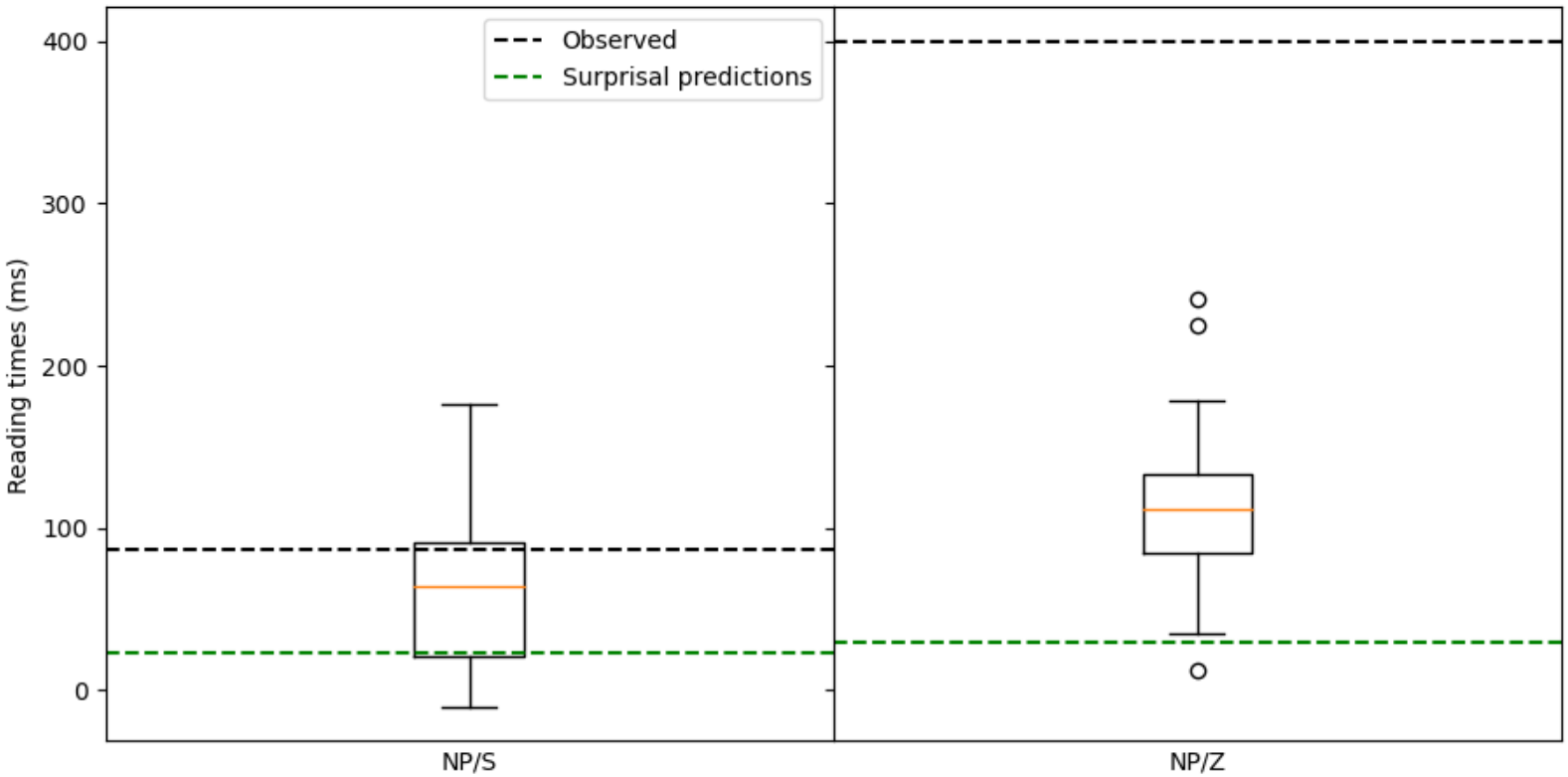


Figure: Garden-path effect predictions from SF

Empirical Results - Levels of difficulty

		BERT model		
		distilbert	bert-base	bert-large
spaCy model	en_core_web_sm	0.03	0.01	0.09
	en_core_web_lg	0.02	0.04	0.24
	en_core_web_trf	0.39	0.0001	0.01

Figure: p-values associated with the t-test evaluating whether the garden-path effect predictions obtained from SF for NP/S and NP/Z are sampled from the same distribution.

Conlusion

Summary of the results

- Created a presheaf model of the human parsing process, which is close to the theories of psycholinguistics

Summary of the results

- Created a presheaf model of the human parsing process, which is close to the theories of psycholinguistics
- We used a measure of “sheafness” SF to quantify the difficulty of parsing words in a sentence

Summary of the results

- Created a presheaf model of the human parsing process, which is close to the theories of psycholinguistics
- We used a measure of “sheafness” SF to quantify the difficulty of parsing words in a sentence
- We obtained good correlations between SF and reading times

Summary of the results

- Created a presheaf model of the human parsing process, which is close to the theories of psycholinguistics
- We used a measure of “sheafness” SF to quantify the difficulty of parsing words in a sentence
- We obtained good correlations between SF and reading times
- We managed to obtain statistically differences between predictions from garden-path sentences which have different levels of difficulty

Summary of the results

- Created a presheaf model of the human parsing process, which is close to the theories of psycholinguistics
- We used a measure of “sheafness” SF to quantify the difficulty of parsing words in a sentence
- We obtained good correlations between SF and reading times
- We managed to obtain statistically differences between predictions from garden-path sentences which have different levels of difficulty
- We compared our results with the state-of-the-art methods from computational linguistics, and obtained more accurate predictions

Future work

- Add more structure to the presheaf (e.g. include semantic information)

Future work

- Add more structure to the presheaf (e.g. include semantic information)
- Combine approaches using SF and surprisal

Future work

- Add more structure to the presheaf (e.g. include semantic information)
- Combine approaches using SF and surprisal
- Produce models of the reanalysis process

Future work

- Add more structure to the presheaf (e.g. include semantic information)
- Combine approaches using SF and surprisal
- Produce models of the reanalysis process
- Use this model to investigate other linguistic phenomena (e.g. memory effects)

Future work

- Add more structure to the presheaf (e.g. include semantic information)
- Combine approaches using SF and surprisal
- Produce models of the reanalysis process
- Use this model to investigate other linguistic phenomena (e.g. memory effects)
- Contextuality with $k < \infty$-lookback?

Thank you!

To appear in Philosophical Transactions of Royal Society A as “Causality and Signalling in Garden-Path Sentences”