From compositional semantics to Bayesian pragmatics via logical inference

Julian Grove, Jean-Philippe Bernardy, and Stergios Chatzikyriakidis

CLASP, University of Gothenburg

NALOMA II, June 16, 2021

Outline

We are here

An inference task: anaphora resolution

(1) Emacs is waiting for the command. It is prepared.

(1) Emacs is waiting for the command. It is prepared.

- One pronoun
- Two possible antecedents
- $2^1 = 2$ possible interpretations to consider

An inference task: anaphora resolution

(1) Emacs is waiting for the command. It is prepared.

- One pronoun
- Two possible antecedents
- $2^1 = 2$ possible interpretations to consider

$\rightsquigarrow$ Emacs is prepared.

An inference task: anaphora resolution

(1) Emacs is waiting for the command. It is prepared.

- One pronoun
- Two possible antecedents
- $2^1 = 2$ possible interpretations to consider

$\rightsquigarrow$ Emacs is prepared.

(2) Ashley is waiting for Amy. She sees her.

An inference task: anaphora resolution

(1) Emacs is waiting for the command. It is prepared.

- One pronoun
- Two possible antecedents
- $2^1 = 2$ possible interpretations to consider

$\rightsquigarrow$ Emacs is prepared.

(2) Ashley is waiting for Amy. She sees her.

- Two pronouns
- Two possible antecedents
- $2^2 = 4$ possible interpretations to consider

An inference task: anaphora resolution

(1) Emacs is waiting for the command. It is prepared.

- One pronoun
- Two possible antecedents
- $2^1 = 2$ possible interpretations to consider

$\leadsto$ Emacs is prepared.

(2) Ashley is waiting for Amy. She sees her.

- Two pronouns
- Two possible antecedents
- $2^2 = 4$ possible interpretations to consider

$\leadsto$ Ashley sees Amy. ???

Two approaches to the problem

Meanings can be characterised logically.

Two approaches to the problem

Meanings can be characterised logically.

- Sentence meanings correspond to logical formulae.

Meanings can be characterised logically.

- Sentence meanings correspond to logical formulae.
- Determined compositionally, through functional application.

Two approaches to the problem

Meanings can be characterised logically.

- Sentence meanings correspond to logical formulae.
- Determined compositionally, through functional application.
  - E.g., in terms of the simply typed λ-calculus (Montague, 1973).

Meanings can be characterised logically.

- Sentence meanings correspond to logical formulae.
- Determined compositionally, through functional application.
  - E.g., in terms of the simply typed λ-calculus (Montague, 1973).
- Can serve as the basis for systems of inference, i.e., by computing entailment using theorem provers and proof assistants (Bekki, 2014; Mineshima et al., 2015; Bernardy and Chatzikyriakidis, 2019).

Two approaches to the problem

Meanings can be characterised logically.

- Sentence meanings correspond to logical formulae.
- Determined compositionally, through functional application.
  - E.g., in terms of the simply typed λ-calculus (Montague, 1973).
- Can serve as the basis for systems of inference, i.e., by computing entailment using theorem provers and proof assistants (Bekki, 2014; Mineshima et al., 2015; Bernardy and Chatzikyriakidis, 2019).
- Such a system's behavior is controlled, predictable, and well-understood.

Two approaches to the problem

Meanings can be characterised logically.

- Sentence meanings correspond to logical formulae.
- Determined compositionally, through functional application.
  - E.g., in terms of the simply typed λ-calculus (Montague, 1973).
- Can serve as the basis for systems of inference, i.e., by computing entailment using theorem provers and proof assistants (Bekki, 2014; Mineshima et al., 2015; Bernardy and Chatzikyriakidis, 2019).
- Such a system's behavior is controlled, predictable, and well-understood.
- Much manual intervention needed; inflexible in the face of gradience and uncertainty.

Two approaches to the problem

Meanings can be characterised statistically or probabilistically.

Two approaches to the problem

Meanings can be characterised statistically or probabilistically.

- In terms of linguistic contexts (distributional semantics).

Two approaches to the problem

Meanings can be characterised statistically or probabilistically.

- In terms of linguistic contexts (distributional semantics).
- In terms of probability distributions over possible worlds or situations.

Two approaches to the problem

Meanings can be characterised statistically or probabilistically.

- In terms of linguistic contexts (distributional semantics).
- In terms of probability distributions over possible worlds or situations.
- Often non-compositional (in Montague's sense).

Two approaches to the problem

Meanings can be characterised statistically or probabilistically.

- In terms of linguistic contexts (distributional semantics).
- In terms of probability distributions over possible worlds or situations.
- Often non-compositional (in Montague's sense).
- Can serve as the basis for explicit theories of pragmatics, e.g., Rational Speech Act (RSA) models (Lassiter and Goodman, 2013; Goodman and Stuhlmüller, 2013; Lassiter and Goodman, 2017; Emerson, 2020).

Two approaches to the problem

Meanings can be characterised statistically or probabilistically.

- In terms of linguistic contexts (distributional semantics).
- In terms of probability distributions over possible worlds or situations.
- Often non-compositional (in Montague's sense).
- Can serve as the basis for explicit theories of pragmatics, e.g., Rational Speech Act (RSA) models (Lassiter and Goodman, 2013; Goodman and Stuhlmüller, 2013; Lassiter and Goodman, 2017; Emerson, 2020).
- Often much less manual supervision required (such systems can be *learned*).

Two approaches to the problem

Meanings can be characterised statistically or probabilistically.

- In terms of linguistic contexts (distributional semantics).
- In terms of probability distributions over possible worlds or situations.
- Often non-compositional (in Montague's sense).
- Can serve as the basis for explicit theories of pragmatics, e.g., Rational Speech Act (RSA) models (Lassiter and Goodman, 2013; Goodman and Stuhlmüller, 2013; Lassiter and Goodman, 2017; Emerson, 2020).
- Often much less manual supervision required (such systems can be *learned*).
- Flexible in the face of gradience and uncertainty.

Plan: a best-of-both-worlds approach

Our plan is to offer a combined approach.

Plan: a best-of-both-worlds approach

Our plan is to offer a combined approach.

- Sentence meanings correspond to *probability distributions* over FOL formulae.

Plan: a best-of-both-worlds approach

Our plan is to offer a combined approach.

- Sentence meanings correspond to *probability distributions* over FOL formulae.
- Probability distributions are computed compositionally, using standard Montagovian tools.

Plan: a best-of-both-worlds approach

Our plan is to offer a combined approach.

- Sentence meanings correspond to *probability distributions* over FOL formulae.
- Probability distributions are computed compositionally, using standard Montagovian tools.
- Can be used to capture gradient patterns of inference (by computing expected values of probability distributions).

Plan: a best-of-both-worlds approach

Our plan is to offer a combined approach.

- Sentence meanings correspond to *probability distributions* over FOL formulae.
- Probability distributions are computed compositionally, using standard Montagovian tools.
- Can be used to capture gradient patterns of inference (by computing expected values of probability distributions).
- Can be used to capture entailment (via theorem proving).

Plan: a best-of-both-worlds approach

Our plan is to offer a combined approach.

- Sentence meanings correspond to *probability distributions* over FOL formulae.
- Probability distributions are computed compositionally, using standard Montagovian tools.
- Can be used to capture gradient patterns of inference (by computing expected values of probability distributions).
- Can be used to capture entailment (via theorem proving).

Plan: a best-of-both-worlds approach

Our plan is to offer a combined approach.

- Sentence meanings correspond to *probability distributions* over FOL formulae.
- Probability distributions are computed compositionally, using standard Montagovian tools.
- Can be used to capture gradient patterns of inference (by computing expected values of probability distributions).
- Can be used to capture entailment (via theorem proving).

The trick is to use *probabilistic programs*. These allow us to view the logical semantics and the probabilistic semantics as two modular aspects of the same computation.

Example: RSA and anaphora resolution

We illustrate our approach by building an RSA model of anaphora resolution.

Example: RSA and anaphora resolution

We illustrate our approach by building an RSA model of anaphora resolution.

- The speaker utters a sentence with pronouns having possible antecedents.

Example: RSA and anaphora resolution

We illustrate our approach by building an RSA model of anaphora resolution.

- The speaker utters a sentence with pronouns having possible antecedents.
- The listener computes a posterior over interpretations of the pronouns (and thus the utterance).

We are here

1. Overview
2. Our framework
3. Anaphora resolution
4. Conclusions

Evaluating truth against background knowledge

Inference requires a characterisation of background knowledge, for which we employ the notion of a *world state*.

Evaluating truth against background knowledge

Inference requires a characterisation of background knowledge, for which we employ the notion of a *world state*.

- That is, a set of FOL formulae; e.g., $\{\text{height}_{\text{ashley}} \geq 1.63m\}$.

Evaluating truth against background knowledge

Inference requires a characterisation of background knowledge, for which we employ the notion of a *world state*.

- That is, a set of FOL formulae; e.g., $\{\mathrm{height}_{\mathrm{ashley}} \geq 1.63m\}$.

Evaluating truth against background knowledge

Inference requires a characterisation of background knowledge, for which we employ the notion of a *world state*.

- That is, a set of FOL formulae; e.g., $\{\mathrm{height}_{\mathrm{ashley}} \geq 1.63m\}$.

Probabilistic background knowledge can be represented as a random variable Γ representing a world state.

Evaluating truth against background knowledge

Inference requires a characterisation of background knowledge, for which we employ the notion of a *world state*.

- That is, a set of FOL formulae; e.g., $\{\text{height}_{\text{ashley}} \geq 1.63m\}$.

Probabilistic background knowledge can be represented as a random variable Γ representing a world state.

- E.g., $\Gamma = \{\text{height}_{\text{ashley}} \geq \theta\}$, where $\theta \sim \mathcal{N}(1.63m, 0.06m)$.

Evaluating truth against background knowledge

Inference requires a characterisation of background knowledge, for which we employ the notion of a *world state*.

- That is, a set of FOL formulae; e.g., $\{\text{height}_{\text{ashley}} \geq 1.63m\}$.

Probabilistic background knowledge can be represented as a random variable Γ representing a world state.

- E.g., $\Gamma = \{\text{height}_{\text{ashley}} \geq \theta\}$, where $\theta \sim \mathcal{N}(1.63m, 0.06m)$.

Evaluating truth against background knowledge

Inference requires a characterisation of background knowledge, for which we employ the notion of a *world state*.

- That is, a set of FOL formulae; e.g., $\{\text{height}_{\text{ashley}} \geq 1.63m\}$.

Probabilistic background knowledge can be represented as a random variable Γ representing a world state.

- E.g., $\Gamma = \{\text{height}_{\text{ashley}} \geq \theta\}$, where $\theta \sim \mathcal{N}(1.63m, 0.06m)$.

We can check the probability that some formula ϕ is *entailed* by or is *compatible with* some unknown world-state with a known distribution.

Evaluating truth against background knowledge

Inference requires a characterisation of background knowledge, for which we employ the notion of a *world state*.

- That is, a set of FOL formulae; e.g., $\{\text{height}_{\text{ashley}} \geq 1.63m\}$.

Probabilistic background knowledge can be represented as a random variable Γ representing a world state.

- E.g., $\Gamma = \{\text{height}_{\text{ashley}} \geq \theta\}$, where $\theta \sim \mathcal{N}(1.63m, 0.06m)$.

We can check the probability that some formula ϕ is *entailed* by or is *compatible with* some unknown world-state with a known distribution.

- ϕ is entailed by Γ: $\mathbb{E}_\Gamma[\Gamma \vdash \phi]$.

Evaluating truth against background knowledge

Inference requires a characterisation of background knowledge, for which we employ the notion of a *world state*.

- That is, a set of FOL formulae; e.g., $\{\text{height}_{\text{ashley}} \geq 1.63m\}$.

Probabilistic background knowledge can be represented as a random variable Γ representing a world state.

- E.g., $\Gamma = \{\text{height}_{\text{ashley}} \geq \theta\}$, where $\theta \sim \mathcal{N}(1.63m, 0.06m)$.

We can check the probability that some formula ϕ is *entailed* by or is *compatible with* some unknown world-state with a known distribution.

- ϕ is entailed by Γ: $\mathbb{E}_\Gamma[\Gamma \vdash \phi]$.
- ϕ is compatible with Γ: $\mathbb{E}_\Gamma[\Gamma, \phi \nvdash \bot]$.

Evaluating truth against background knowledge

Inference requires a characterisation of background knowledge, for which we employ the notion of a *world state*.

- That is, a set of FOL formulae; e.g., $\{\text{height}_{\text{ashley}} \geq 1.63m\}$.

Probabilistic background knowledge can be represented as a random variable Γ representing a world state.

- E.g., $\Gamma = \{\text{height}_{\text{ashley}} \geq \theta\}$, where $\theta \sim \mathcal{N}(1.63m, 0.06m)$.

We can check the probability that some formula ϕ is *entailed* by or is *compatible with* some unknown world-state with a known distribution.

- ϕ is entailed by Γ: $\mathbb{E}_\Gamma[\Gamma \vdash \phi]$.
- ϕ is compatible with Γ: $\mathbb{E}_\Gamma[\Gamma, \phi \nvdash \bot]$.

Evaluating truth against background knowledge

Inference requires a characterisation of background knowledge, for which we employ the notion of a *world state*.

- That is, a set of FOL formulae; e.g., $\{\text{height}_{\text{ashley}} \geq 1.63m\}$.

Probabilistic background knowledge can be represented as a random variable Γ representing a world state.

- E.g., $\Gamma = \{\text{height}_{\text{ashley}} \geq \theta\}$, where $\theta \sim \mathcal{N}(1.63m, 0.06m)$.

We can check the probability that some formula ϕ is *entailed* by or is *compatible with* some unknown world-state with a known distribution.

- ϕ is entailed by Γ: $\mathbb{E}_\Gamma[\Gamma \vdash \phi]$.
- ϕ is compatible with Γ: $\mathbb{E}_\Gamma[\Gamma, \phi \nvdash \bot]$.

We compute the $(\vdash)$ relation using a standard FOL tableau theorem prover (limited to depth 10).

Our model is defined by the following relations:

u: utterance
ϕ: inferred proposition
$C(u)$: utterance cost
θ: set of linguistic parameters, which are drawn from a probability distribution coming from the Montague semantics

Our model is defined by the following relations:

$$P_{L_1}(\phi \mid u) \propto P_{S_1}(u \mid \phi) \times P(\phi) \qquad \text{(The pragmatic listener } L_1\text{)}$$

u: utterance
ϕ: inferred proposition
$C(u)$: utterance cost
θ: set of linguistic parameters, which are drawn from a probability distribution coming from the Montague semantics

The RSA model

Our model is defined by the following relations:

$$P_{L_1}(\phi \mid u) \propto P_{S_1}(u \mid \phi) \times P(\phi) \qquad \text{(The pragmatic listener } L_1\text{)}$$

$$P_{S_1}(u \mid \phi) \propto (P_{L_0}(\phi \mid u)/C(u))^{\alpha} \qquad \text{(The pragmatic speaker } S_1\text{)}$$

u: utterance
ϕ: inferred proposition
$C(u)$: utterance cost
θ: set of linguistic parameters, which are drawn from a probability distribution coming from the Montague semantics

The RSA model

Our model is defined by the following relations:

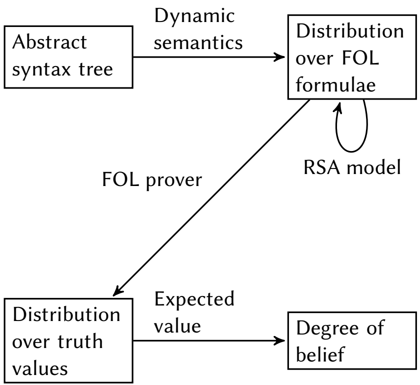
$$P_{L_1}(\phi \mid u) \propto P_{S_1}(u \mid \phi) \times P(\phi) \qquad \text{(The pragmatic listener } L_1\text{)}$$

$$P_{S_1}(u \mid \phi) \propto (P_{L_0}(\phi \mid u)/C(u))^\alpha \qquad \text{(The pragmatic speaker } S_1\text{)}$$

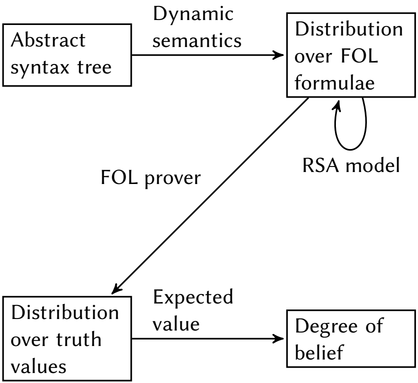
$$P_{L_0}(\phi \mid u) = \mathbb{E}_{\theta,\Gamma}[\Gamma, \phi, [\![u]\!]^\theta \nvdash \bot] \qquad \text{(The literal listener } L_0\text{)}$$

u: utterance
ϕ: inferred proposition
$C(u)$: utterance cost
θ: set of linguistic parameters, which are drawn from a probability distribution coming from the Montague semantics

Bird's-eye view

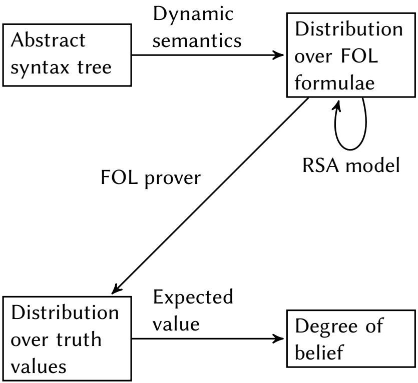


Bird's-eye view



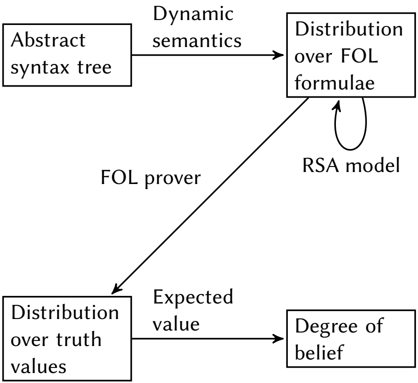
- Syntax ⇒ FOL

Bird's-eye view



- Syntax ⇒ FOL
- FOL ⇒ Booleans

Bird's-eye view



- Syntax $\Rightarrow$ FOL
- FOL $\Rightarrow$ Booleans
- Booleans $\Rightarrow$ Expected values

We are here

(1) Emacs is waiting for the command. It is prepared.

- One pronoun
- Two possible antecedents
- $2^1 = 2$ possible interpretations ϕ for L_1 to consider

(2) Ashley is waiting for Amy. She sees her.

- Two pronouns
- Two possible antecedents
- $2^2 = 4$ possible interpretations for L_1 to consider

Background knowledge

Γ, a probabilistic program that returns world states

Background knowledge

Γ, a probabilistic program that returns world states

- We choose, ahead of time, a set of formulae Ψ which may be used to construct a world-state.

Background knowledge

Γ, a probabilistic program that returns world states

- We choose, ahead of time, a set of formulae Ψ which may be used to construct a world-state.
- For any formula $\psi \in \Psi$, either $\psi \in \Gamma$ or $\neg\psi \in \Gamma$, as according to some Bernoulli distribution.

Background knowledge

Γ, a probabilistic program that returns world states

- We choose, ahead of time, a set of formulae Ψ which may be used to construct a world-state.
- For any formula $\psi \in \Psi$, either $\psi \in \Gamma$ or $\neg\psi \in \Gamma$, as according to some Bernoulli distribution.

Background knowledge

Γ, a probabilistic program that returns world states

- We choose, ahead of time, a set of formulae Ψ which may be used to construct a world-state.
- For any formula $\psi \in \Psi$, either $\psi \in \Gamma$ or $\neg\psi \in \Gamma$, as according to some Bernoulli distribution.

All of the following formulae are given prior probability 0.5:

- $\exists x : \mathrm{wait\_for}(\mathrm{emacs}, x)$
- $\exists x : \mathrm{wait\_for}(\mathrm{the\_command}, x)$
- $\exists x : \mathrm{wait\_for}(\mathrm{ashley}, x)$
- $\exists x : \mathrm{wait\_for}(\mathrm{amy}, x)$
- $\mathrm{prepared}(\mathrm{emacs})$
- $\mathrm{prepared}(\mathrm{the\_command})$
- $\exists x : \mathrm{see}(\mathrm{ashley}, x)$
- $\exists x : \mathrm{see}(\mathrm{amy}, x)$

Background knowledge: animacy

We distinguish examples (1) and (2) in terms of the animacy entailed for the subjects of the antecedent sentences.

We distinguish examples (1) and (2) in terms of the animacy entailed for the subjects of the antecedent sentences.

(1) Emacs is waiting for the command. It is prepared.

(2) Ashley is waiting for Amy. She sees her.

Background knowledge: animacy

We distinguish examples (1) and (2) in terms of the animacy entailed for the subjects of the antecedent sentences.

(1) Emacs is waiting for the command. It is prepared.

(2) Ashley is waiting for Amy. She sees her.

$animate(emacs)$	0.2
$animate(the\_command)$	0.2
$animate(ashley)$	0.9
$animate(amy)$	0.9

Background knowledge: animacy

We distinguish examples (1) and (2) in terms of the animacy entailed for the subjects of the antecedent sentences.

(1) Emacs is waiting for the command. It is prepared.

(2) Ashley is waiting for Amy. She sees her.

$animate(emacs)$	0.2
$animate(the\_command)$	0.2
$animate(ashley)$	0.9
$animate(amy)$	0.9

We also ensure that each world state satisfies the following lexical entailments:

Background knowledge: animacy

We distinguish examples (1) and (2) in terms of the animacy entailed for the subjects of the antecedent sentences.

(1) Emacs is waiting for the command. It is prepared.

(2) Ashley is waiting for Amy. She sees her.

$animate(emacs)$	0.2
$animate(the\_command)$	0.2
$animate(ashley)$	0.9
$animate(amy)$	0.9

We also ensure that each world state satisfies the following lexical entailments:

- $\forall x : (\exists y : \text{wait\_for}(x, y)) \rightarrow \text{animate}(x)$

Background knowledge: animacy

We distinguish examples (1) and (2) in terms of the animacy entailed for the subjects of the antecedent sentences.

(1) Emacs is waiting for the command. It is prepared.

(2) Ashley is waiting for Amy. She sees her.

$animate(emacs)$	0.2
$animate(the\_command)$	0.2
$animate(ashley)$	0.9
$animate(amy)$	0.9

We also ensure that each world state satisfies the following lexical entailments:

- $\forall x : (\exists y : \text{wait\_for}(x, y)) \to \text{animate}(x)$
- $\forall x : \text{prepared}(x) \to \text{animate}(x)$

Background knowledge: animacy

We distinguish examples (1) and (2) in terms of the animacy entailed for the subjects of the antecedent sentences.

(1) Emacs is waiting for the command. It is prepared.

(2) Ashley is waiting for Amy. She sees her.

$animate(emacs)$	0.2
$animate(the\_command)$	0.2
$animate(ashley)$	0.9
$animate(amy)$	0.9

We also ensure that each world state satisfies the following lexical entailments:

- $\forall x : (\exists y : \text{wait\_for}(x, y)) \rightarrow \text{animate}(x)$
- $\forall x : \text{prepared}(x) \rightarrow \text{animate}(x)$
- $\forall x : (\exists y : \text{see}(x, y)) \rightarrow \text{animate}(x)$

Example (1) model and results

(1) Emacs is waiting for the command. It is prepared.

$$P_{L_1}(\phi \mid (1)) \propto P_{S_1}((1) \mid \phi) \times \mathbb{E}_{\Gamma}[\Gamma, \phi \nvdash \bot]$$

$$P_{S_1}(u \mid \phi) \propto (P_{L_0}(\phi \mid u)/e^{(npCost*\#_{\text{NPs}}(u))+(pnCost*\#_{\text{pronouns}}(u))})^{\alpha}$$

$$P_{L_0}(\phi \mid u) = \mathbb{E}_{\theta,\Gamma}[\Gamma, \phi, [\![u]\!]^{\theta} \nvdash \bot]$$

Alternative u's for P_{S_1} are gotten by substituting pronouns by their corresponding NPs, given the interpretation ϕ.

(1) Emacs is waiting for the command. It is prepared.

$$P_{L_1}(\phi \mid (1)) \propto P_{S_1}((1) \mid \phi) \times \mathbb{E}_{\Gamma}[\Gamma, \phi \nvdash \bot]$$

$$P_{S_1}(u \mid \phi) \propto (P_{L_0}(\phi \mid u) / e^{(npCost * \#_{\text{NPs}}(u)) + (pnCost * \#_{\text{pronouns}}(u))})^{\alpha}$$

$$P_{L_0}(\phi \mid u) = \mathbb{E}_{\theta,\Gamma}[\Gamma, \phi, [\![u]\!]^{\theta} \nvdash \bot]$$

Alternative u's for P_{S_1} are gotten by substituting pronouns by their corresponding NPs, given the interpretation ϕ.

Results:

α	*pnCost*	*npCost*	*Emacs* bias
0.5	1	2	86.9%
4.0	1	2	98.6%

Example (2) model and results

(2) Ashley is waiting for Amy. She sees her.

$$P_{L_1}(\phi \mid (1)) \propto P_{S_1}((1) \mid \phi) \times \mathbb{E}_{\Gamma}[\Gamma, \phi \nvdash \bot]$$

$$P_{S_1}(u \mid \phi) \propto (P_{L_0}(\phi \mid u)/e^{(npCost * \#_{\text{NPs}}(u)) + (pnCost * \#_{\text{pronouns}}(u))})^{\alpha}$$

$$P_{L_0}(\phi \mid u) = \mathbb{E}_{\theta,\Gamma}[\Gamma, \phi, [\![u]\!]^{\theta} \nvdash \bot]$$

Alternative u's for P_{S_1} are gotten by substituting pronouns by their corresponding NPs, given the interpretation ϕ.

Example (2) model and results

(2) Ashley is waiting for Amy. She sees her.

$$P_{L_1}(\phi \mid (1)) \propto P_{S_1}((1) \mid \phi) \times \mathbb{E}_\Gamma[\Gamma, \phi \nvdash \bot]$$

$$P_{S_1}(u \mid \phi) \propto (P_{L_0}(\phi \mid u)/e^{(npCost*\#_{\text{NPs}}(u))+(pnCost*\#_{\text{pronouns}}(u))})^\alpha$$

$$P_{L_0}(\phi \mid u) = \mathbb{E}_{\theta,\Gamma}[\Gamma, \phi, [\![u]\!]^\theta \nvdash \bot]$$

Alternative u's for P_{S_1} are gotten by substituting pronouns by their corresponding NPs, given the interpretation ϕ.

Results:

α	*pnCost*	*npCost*	*Ashley* bias for *she*	*Ashley* bias for *her*
0.5	1	2	52.9%	50%
4.0	1	2	54.2%	50%

We are here

Conclusions

About the RSA model:

Conclusions

About the RSA model:

- The posterior distribution inferred by L_1 is highly dependent on background knowledge (animacy priors plus lexical entailments).

Conclusions

About the RSA model:

- The posterior distribution inferred by L_1 is highly dependent on background knowledge (animacy priors plus lexical entailments).
  - A pronoun which is entailed to be animate will seek out animacy in its antecedent, as example (1) showed.

About the RSA model:

- The posterior distribution inferred by L_1 is highly dependent on background knowledge (animacy priors plus lexical entailments).
  - A pronoun which is entailed to be animate will seek out animacy in its antecedent, as example (1) showed.
  - The model is less certain when both possible antecedents are animate (example (2)).

Conclusions

About the RSA model:

- The posterior distribution inferred by L_1 is highly dependent on background knowledge (animacy priors plus lexical entailments).
  - A pronoun which is entailed to be animate will seek out animacy in its antecedent, as example (1) showed.
  - The model is less certain when both possible antecedents are animate (example (2)).
- The model thus seems to achieve a kind of abductive inference, i.e., by computing the posterior that is most compatible with background knowledge.

Conclusions

About the RSA model:

- The posterior distribution inferred by L_1 is highly dependent on background knowledge (animacy priors plus lexical entailments).
  - A pronoun which is entailed to be animate will seek out animacy in its antecedent, as example (1) showed.
  - The model is less certain when both possible antecedents are animate (example (2)).
- The model thus seems to achieve a kind of abductive inference, i.e., by computing the posterior that is most compatible with background knowledge.

Conclusions

About the RSA model:

- The posterior distribution inferred by L_1 is highly dependent on background knowledge (animacy priors plus lexical entailments).
  - A pronoun which is entailed to be animate will seek out animacy in its antecedent, as example (1) showed.
  - The model is less certain when both possible antecedents are animate (example (2)).
- The model thus seems to achieve a kind of abductive inference, i.e., by computing the posterior that is most compatible with background knowledge.

More generally:

Conclusions

About the RSA model:

- The posterior distribution inferred by L_1 is highly dependent on background knowledge (animacy priors plus lexical entailments).
  - A pronoun which is entailed to be animate will seek out animacy in its antecedent, as example (1) showed.
  - The model is less certain when both possible antecedents are animate (example (2)).
- The model thus seems to achieve a kind of abductive inference, i.e., by computing the posterior that is most compatible with background knowledge.

More generally:

- Logical entailment can serve as the basis for probabilistic inference via probabilistic programs.

Conclusions

About the RSA model:

- The posterior distribution inferred by L_1 is highly dependent on background knowledge (animacy priors plus lexical entailments).
  - A pronoun which is entailed to be animate will seek out animacy in its antecedent, as example (1) showed.
  - The model is less certain when both possible antecedents are animate (example (2)).
- The model thus seems to achieve a kind of abductive inference, i.e., by computing the posterior that is most compatible with background knowledge.

More generally:

- Logical entailment can serve as the basis for probabilistic inference via probabilistic programs.
- Such programs can be built compositionally, using standard Montagovian tools.

References

Bekki, Daisuke. 2014. Representing Anaphora with Dependent Types. In *Logical Aspects of Computational Linguistics*, ed. Nicholas Asher and Sergei Soloviev, Lecture Notes in Computer Science, 14–29. Berlin, Heidelberg: Springer.

Bernardy, Jean-Philippe, and Stergios Chatzikyriakidis. 2019. A Wide-Coverage Symbolic Natural Language Inference System. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, 298–303. Turku, Finland: Linköping University Electronic Press. https://www.aclweb.org/anthology/W19-6131.

Emerson, Guy. 2020. Linguists Who Use Probabilistic Models Love Them: Quantification in Functional Distributional Semantics. In *Proceedings of the Probability and Meaning Conference (PaM 2020)*, 41–52. Gothenburg: Association for Computational Linguistics. https://www.aclweb.org/anthology/2020.pam-1.6.

Goodman, Noah D., and Andreas Stuhlmüller. 2013. Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics in Cognitive Science* 5:173–184. https://onlinelibrary.wiley.com/doi/abs/10.1111/tops.12007.

Lassiter, Daniel, and Noah D. Goodman. 2013. Context, scale structure, and statistics in the interpretation of positive-form adjectives. *Semantics and Linguistic Theory* 23:587–610. https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/2658, number: 0.

Lassiter, Daniel, and Noah D. Goodman. 2017. Adjectival vagueness in a Bayesian model of interpretation. *Synthese* 194:3801–3836. https://doi.org/10.1007/s11229-015-0786-1.

Mineshima, Koji, Pascual Martínez-Gómez, Yusuke Miyao, and Daisuke Bekki. 2015. Higher-order logical inference with compositional semantics. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2055–2061. Lisbon, Portugal: Association for Computational Linguistics. https://www.aclweb.org/anthology/D15-1244.

Montague, Richard. 1973. The Proper Treatment of Quantification in Ordinary English. In *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, ed. K. J. J. Hintikka, J. M. E. Moravcsik, and P. Suppes, Synthese Library, 221–242. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-2506-5_10.

How can one compute probabilistic truth/entailment?

We compute probability distributions over logical formulae, world-states, and truth values using *probabilistic programs*.

Probabilistic programs

A probabilistic program that returns a value of type α is a function of type $(\alpha \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$: it consumes a function from values of type α to reals, in order to return a real.

- Example: a program that returns values from some list l with a uniform distribution is $\lambda f.\mathtt{sum}(\mathtt{map} f l)/(\mathtt{length} l)$.
  - Given a function f, it returns its mean across l.
  - If α is $\mathbb{R}$, feeding this program the identity function results in an expected value.

Probabilistic programs can be composed!

If:

- $m : ((\alpha \longrightarrow \beta) \longrightarrow \mathbb{R}) \longrightarrow \mathbb{R}$
- $n : (\alpha \longrightarrow \mathbb{R}) \longrightarrow \mathbb{R}$

Then:

- $\lambda k.m(\lambda f.n(\lambda x.k(fx))) : (\beta \longrightarrow \mathbb{R}) \longrightarrow \mathbb{R}$

(This is applicative composition in the continuation monad.)