Natural Logic and Vehicles of Inference

Celebration Event in Honor of Johan van Benthem

Thomas Icard
Stanford University

September 27, 2014

Natural Logic (van Benthem, 1987):

> "using linguistic constructs directly as a vehicle of inference"

Natural Logic (van Benthem, 1987):

> "using linguistic constructs directly as a vehicle of inference"
>
> rather than focusing on a single "worst-case vehicle of interpretation," we should recognize "a variety of such interpretations, whose structure is itself a topic for semantic research."

Classic example: Monotonicity Calculus

Every American likes jazz
———
Every Tennessean likes jazz
———
Every Tennessean likes some form of music

Classic example: Monotonicity Calculus

$$\dfrac{\dfrac{\text{Every American likes jazz}}{\text{Every Tennessean likes jazz}}}{\text{Every Tennessean likes some form of music}}$$

$$\downarrow \text{Every} \uparrow$$

Classic example: Monotonicity Calculus

$$\dfrac{\dfrac{\text{No American likes jazz}}{\text{No Tennessean likes jazz}}}{\text{No Tennessean likes West Coast jazz}}$$

$\downarrow$ No $\downarrow$

Classic example: Monotonicity Calculus

$\downarrow$ Every $\uparrow$ $\qquad\qquad$ $\downarrow$ No $\downarrow$

Classic example: Monotonicity Calculus

$\downarrow$ Every $\uparrow$ $\qquad$ $\downarrow$ No $\downarrow$

$\uparrow$ Some $\uparrow$ $\qquad$ $\uparrow$ Not_Every $\downarrow$

Classic example: Monotonicity Calculus

$\downarrow$ Every $\uparrow$	$\downarrow$ No $\downarrow$
$\uparrow$ Some $\uparrow$	$\uparrow$ Not_Every $\downarrow$
$*$ Most $\uparrow$	$*$ Few $\downarrow$

$$\frac{\text{Frank failed to complete his taxes}}{\text{Frank failed to complete his taxes on time}}$$

$$\frac{\text{Frank failed to complete his taxes}}{\text{Frank failed to complete his taxes on time}}$$

$$\frac{\text{Every Tennessean who failed to complete his taxes on time likes jazz}}{\text{Every Tennessean who failed to complete his taxes likes jazz}}$$

$$\dfrac{\dfrac{\text{Every [American]}^{-}\text{ likes [jazz]}^{+}}{\text{Every [Tennessean]}^{-}\text{ likes [jazz]}^{+}}}{\text{Every [Tennessean]}^{-}\text{ likes [some form of music]}^{+}}$$

$$\dfrac{\text{Frank failed to [complete his taxes]}^{-}}{\text{Frank failed to [complete his taxes on time]}^{-}}$$

$$\dfrac{\text{Every Tennessean who failed to [complete his taxes on time]}^{+}\text{ likes jazz}}{\text{Every Tennessean who failed to [complete his taxes]}^{+}\text{ likes jazz}}$$

$$\frac{\text{All Americans like } [\text{jazz}]^{+}}{\text{All Americans likes } [\text{some form of music}]^{+}}$$

$$\frac{\text{Most Americans likes } [\text{jazz}]^{+}}{\text{Most Americans likes } [\text{some form of music}]^{+}}$$

All Americans like [jazz]$^{+}$

All Americans likes [some form of music]$^{+}$

Most Americans likes [jazz]$^{+}$

Most Americans likes [some form of music]$^{+}$

- From a logical point of view, such patterns cross-cut the standard "order hierarchy."

$$\frac{\text{All Americans like [jazz]}^{+}}{\text{All Americans likes [some form of music]}^{+}}$$

$$\frac{\text{Most Americans likes [jazz]}^{+}}{\text{Most Americans likes [some form of music]}^{+}}$$

- From a logical point of view, such patterns cross-cut the standard "order hierarchy."
- Empirically, these two argument patterns are indeed seen to be equally easy for people (Oaksford & Chater, 2001).

Two parts to this talk:

Two parts to this talk:

1. Natural logic as opening up new research questions in pure logic: “logic for its own sake”

Two parts to this talk:

1. Natural logic as opening up new research questions in pure logic: “logic for its own sake”

2. Deep links between logic and other areas of inquiry, in this case language and cognition.

Two parts to this talk:

1. Natural logic as opening up new research questions in pure logic: "logic for its own sake"

2. Deep links between logic and other areas of inquiry, in this case language and cognition.

This dual influence is characteristic of Johan's work.

Part 1: Logical Issues

- Early work on Monotonicty Calculus by van Benthem (1986) and Sánchez-Valencia (1991) showed how to mark types with monotonicity information and develop simple proof systems.

- Early work on Monotonicty Calculus by van Benthem (1986) and Sánchez-Valencia (1991) showed how to mark types with monotonicity information and develop simple proof systems.

- Soundness of the marking system, and hence of the proof system, was established. However, completeness remained open.

"If we only take some natural subset of the rules, would the system be complete for some 'rougher semantics', closer to mental models that we use?"

(van Benthem 2011, CSLI Handout)

"If we only take some natural subset of the rules, would the system be complete for some 'rougher semantics', closer to mental models that we use?"

(van Benthem 2011, CSLI Handout)

In recent work, Larry Moss and I have taken on this question, couching it in a broader study of monotonicity reasoning in general (2013, 2014).

Simple Arithmetic Example

Which is bigger, $-(7 + 2^{-3})$ or $-(7 + 2^{-4})$?

Simple Arithmetic Example

Which is bigger, $-(7+2^{-3})$ or $-(7+2^{-4})$?

$$\dfrac{\dfrac{\dfrac{\dfrac{3 < 4}{-4 < -3}\ (-x \text{ is antitone})}{2^{-4} < 2^{-3}}\ (2^x \text{ is monotone})}{7+2^{-4} < 7+2^{-3}}\ (7+x \text{ is monotone})}{-(7+2^{-3}) < -(7+2^{-4})}\ (-x \text{ is antitone})$$

Types and Type Domains

Definition
$\mathcal{M} = \{+, -, \cdot\}$.

Types and Type Domains

Definition

$\mathcal{M} = \{+, -, \cdot\}$.

Definition

Let $\mathcal{B}$ be a set of base types. The full set of types $\mathcal{T}$ is defined as the smallest superset of $\mathcal{B}$, such that whenever $\sigma, \tau \in \mathcal{T}$, so is $\sigma \xrightarrow{m} \tau$, for each $m \in \mathcal{M}$.

Definition

A *structure* is a system $\mathcal{S} = \{\mathbb{D}_\tau\}_{\tau \in \mathcal{T}}$ of preorders.

Base types $\beta \in \mathcal{B}$ have no requirement on $\mathbb{D}_\beta$.

Definition

A *structure* is a system $\mathcal{S} = \{\mathbb{D}_\tau\}_{\tau \in \mathcal{T}}$ of preorders.

Base types $\beta \in \mathcal{B}$ have no requirement on $\mathbb{D}_\beta$.

For functional types $\sigma \xrightarrow{m} \tau$, we require:

1. $D_{\sigma \xrightarrow{+} \tau}$ = monotone functions from $\mathbb{D}_\sigma$ to $\mathbb{D}_\tau$.
2. $D_{\sigma \xrightarrow{-} \tau}$ = antitone functions from $\mathbb{D}_\sigma$ to $\mathbb{D}_\tau$.
3. $D_{\sigma \xrightarrow{\cdot} \tau}$ = all functions from $\mathbb{D}_\sigma$ to $\mathbb{D}_\tau$.

Definition

A *structure* is a system $\mathcal{S} = \{\mathbb{D}_\tau\}_{\tau \in \mathcal{T}}$ of preorders.

Base types $\beta \in \mathcal{B}$ have no requirement on $\mathbb{D}_\beta$.

For functional types $\sigma \xrightarrow{m} \tau$, we require:

1. $D_{\sigma \xrightarrow{+} \tau}$ = monotone functions from $\mathbb{D}_\sigma$ to $\mathbb{D}_\tau$.
2. $D_{\sigma \xrightarrow{-} \tau}$ = antitone functions from $\mathbb{D}_\sigma$ to $\mathbb{D}_\tau$.
3. $D_{\sigma \xrightarrow{\cdot} \tau}$ = all functions from $\mathbb{D}_\sigma$ to $\mathbb{D}_\tau$.

The ordering $\leq_{\sigma \xrightarrow{m} \tau}$ on $D_{\sigma \xrightarrow{m} \tau}$ is given pointwise:

$f \leq_{\sigma \xrightarrow{m} \tau} g$ if and only if $f(a) \leq_\tau g(a)$ for $a \in D_\sigma$.

There is a natural relation $\preceq$ between types, such that whenever $\sigma \preceq \tau$, there is a canonical embedding from $\mathbb{D}_\sigma$ into $\mathbb{D}_\tau$.

Definition

Given a set $\mathtt{Con}$ of constants and a function $\mathtt{type} : \mathtt{Con} \to \mathcal{T}$, the set T of typed terms $t : \tau$ is defined recursively, as follows:

Definition

Given a set $\texttt{Con}$ of constants and a function $\texttt{type} : \texttt{Con} \to \mathcal{T}$, the set T of typed terms $t : \tau$ is defined recursively, as follows:

1. If $c \in \texttt{Con}$, then $c : \mathsf{type}(c)$ is a typed term.
2. If $t : \sigma \xrightarrow{m} \tau$ and $u : \rho$ are typed terms and $\rho \preceq \sigma$, then $t(u) : \tau$ is a typed term.

Monotonicity Calculus

$$(\textsc{Refl})\ \frac{}{t \leq t} \qquad (\textsc{Trans})\ \frac{t \leq u \quad u \leq v}{t \leq v}$$

$$(\textsc{Mono})\ \frac{u \leq v}{t_{\uparrow}(u) \leq t_{\uparrow}(v)} \qquad (\textsc{Anti})\ \frac{v \leq u}{t_{\downarrow}(u) \leq t_{\downarrow}(v)}$$

$$(\textsc{Point})\ \frac{s \leq t}{s(u) \leq t(u)}$$

Monotonicity Calculus

$$(\textsc{Refl})\ \frac{}{t \leq t} \qquad (\textsc{Trans})\ \frac{t \leq u \quad u \leq v}{t \leq v}$$

$$(\textsc{Mono})\ \frac{u \leq v}{t_{\uparrow}(u) \leq t_{\uparrow}(v)} \qquad (\textsc{Anti})\ \frac{v \leq u}{t_{\downarrow}(u) \leq t_{\downarrow}(v)}$$

$$(\textsc{Point})\ \frac{s \leq t}{s(u) \leq t(u)}$$

Theorem (Icard & Moss, 2013)

This calculus is sound and strongly complete.

- We can also include variables and λ-abstraction.

- We can also include variables and λ-abstraction.
- In this setting, even completeness of the typing calculus becomes an interesting question.

- We can also include variables and λ-abstraction.
- In this setting, even completeness of the typing calculus becomes an interesting question.
- As Johan long ago observed, this requires a kind of preservation result.

Lemma (Icard & Moss)

1. $\lambda x.t$ is semantically monotone iff all free occurrences of x in t are in positive position.
2. $\lambda x.t$ is semantically antitone iff all free occurrences of x in t are in negative position.

$$(\textsc{Refl})\ \frac{}{t \leq t} \qquad (\textsc{Trans})\ \frac{t \leq u \quad u \leq v}{t \leq v}$$

$$(\textsc{Mono})\ \frac{u \leq v}{t_{\uparrow}(u) \leq t_{\uparrow}(v)} \qquad (\textsc{Anti})\ \frac{v \leq u}{t_{\downarrow}(u) \leq t_{\downarrow}(v)}$$

$$(\textsc{Point})\ \frac{s \leq t}{s(u) \leq t(u)}$$

$$(\textsc{Refl})\ \frac{}{t \leq t} \qquad (\textsc{Trans})\ \frac{t \leq u \quad u \leq v}{t \leq v}$$

$$(\textsc{Mono})\ \frac{u \leq v}{t_{\uparrow}(u) \leq t_{\uparrow}(v)} \quad (\textsc{Anti})\ \frac{v \leq u}{t_{\downarrow}(u) \leq t_{\downarrow}(v)}$$

$$(\textsc{Point})\ \frac{s \leq t}{s(u) \leq t(u)}$$

$$(\alpha)\ \frac{}{\lambda x.t \leq \lambda y.t^{x}_{y}} \qquad (\beta)\ \frac{}{(\lambda x.t)s \equiv t^{x}_{s}}$$

$$(\xi)\ \frac{t \leq s}{\lambda x.t \leq \lambda x.s}$$

Theorem (Icard & Moss, 2014)

This calculus is sound and strongly complete.

N.B. It is undecidable, if assumptions are allowed!

- This is an analogue of a result of H. Friedman (1975) on simply typed λ-calculus.

- This is an analogue of a result of H. Friedman (1975) on simply typed λ-calculus.
- Similar results were obtained by van Benthem (1991) on the so-called Boolean λ-calculus.

- This is an analogue of a result of H. Friedman (1975) on simply typed λ-calculus.
- Similar results were obtained by van Benthem (1991) on the so-called Boolean λ-calculus.
- In this setting—with a distinguished truth-value type—several open questions remain.

Preservation Question

- Suppose t is of truth value type. When is $\lambda x.t$ semantically monotone or antitone?

Preservation Question

- Suppose t is of truth value type. When is $\lambda x.t$ semantically monotone or antitone?
- For the so called Lambek fragment of λ-calculus, van Benthem (1991) showed that this can be characterized in terms of positive and negative occurrences.

Preservation Question

- Suppose t is of truth value type. When is $\lambda x.t$ semantically monotone or antitone?
- For the so called Lambek fragment of λ-calculus, van Benthem (1991) showed that this can be characterized in terms of positive and negative occurrences.
- For the general case this is open.

Further Logical Issues

- Systems with "internalized polarity" marking, and connections with negative polarity items (Dowty 1994; Bernardi 2002; Moss 2012; etc.).

Further Logical Issues

- Systems with "internalized polarity" marking, and connections with negative polarity items (Dowty 1994; Bernardi 2002; Moss 2012; etc.).

- Beyond monotonicity, including other logical relations and classes of functions (MacCartney & Manning 2008, 2009; Icard 2012, 2014).

Part 2: Application to Language and Cognition

The Monotonicity Calculus has been influential in computational linguistics and psycholinguistics:

- As MacCartney & Manning (2007) showed, "surface level" monotonicity reasoning can improve performance on the Recognizing Textual Entailment challenge.

The Monotonicity Calculus has been influential in computational linguistics and psycholinguistics:

- As MacCartney & Manning (2007) showed, "surface level" monotonicity reasoning can improve performance on the Recognizing Textual Entailment challenge.
- Suggestive results by Geurts (2003) and Geurts & van der Slik (2005) provide evidence that monotonicity plays an important role in language processing.

"Even before disambiguation has taken place, *some* consequences can usually be drawn already. Inference is not an all-or-nothing matter."

(van Benthem 1987)

At least two reasons why natural logic is helpful:

1. It would be computationally expensive to default always to the most complex, “worst case” interpretation.

At least two reasons why natural logic is helpful:

1. It would be computationally expensive to default always to the most complex, "worst case" interpretation.

$$\frac{\text{Everyone talked with at least four people from different departments}}{\text{Everyone talked with at least one other person}}$$

At least two reasons why natural logic is helpful:

1. It would be computationally expensive to default always to the most complex, "worst case" interpretation.

$$\frac{\text{Everyone talked with at least four people from different departments}}{\text{Everyone talked with at least one other person}}$$

2. There may be uncertainty about the language itself, whereas certain inferences do not even require disambiguation.

At least two reasons why natural logic is helpful:

1. It would be computationally expensive to default always to the most complex, "worst case" interpretation.

$$\frac{\text{Everyone talked with at least four people from different departments}}{\text{Everyone talked with at least one other person}}$$

2. There may be uncertainty about the language itself, whereas certain inferences do not even require disambiguation.

$$\frac{\text{Most who know a foreign language learned it at home}}{\text{Most who know a foreign language learned it at home or at school}}$$

But . . . as van Benthem (1987) points out, one must be careful with surface reasoning:

$$\frac{\text{Everyone with a garden water it}}{\text{Everyone with a garden-statue water it}}$$

Question: What does it mean to "use linguistic constructs directly as a vehicle of inference"?

Question: What does it mean to "use linguistic constructs directly as a vehicle of inference"?

Two possibilities:

1. Formal proofs / pattern matching (cf. Rips 1994, etc.). N.B.: natural logic is often associated with proof theoretic semantics.

Question: What does it mean to "use linguistic constructs directly as a vehicle of inference"?

Two possibilities:

1. Formal proofs / pattern matching (cf. Rips 1994, etc.). N.B.: natural logic is often associated with proof theoretic semantics.

2. Simplified, and suitably abstracted, model checking (cf. Johnson-Laird 1983), established by monotonicity (and other) patterns.

- Imagine constructing only "partial" mental models, forgetting some aspects of meaning.

- Imagine constructing only "partial" mental models, forgetting some aspects of meaning.
- Suppose we randomly sample situations in which a premise $t(u)$ is true, and v for which either $u \leq v$ or $v \leq u$, in order to determine whether $t(v)$ follows.

- Imagine constructing only "partial" mental models, forgetting some aspects of meaning.
- Suppose we randomly sample situations in which a premise $t(u)$ is true, and v for which either $u \leq v$ or $v \leq u$, in order to determine whether $t(v)$ follows.
- How likely are we to be correct?

- Imagine constructing only "partial" mental models, forgetting some aspects of meaning.
- Suppose we randomly sample situations in which a premise $t(u)$ is true, and v for which either $u \leq v$ or $v \leq u$, in order to determine whether $t(v)$ follows.
- How likely are we to be correct?
  - If it does follow, we will always be correct.

- Imagine constructing only “partial” mental models, forgetting some aspects of meaning.
- Suppose we randomly sample situations in which a premise $t(u)$ is true, and v for which either $u \leq v$ or $v \leq u$, in order to determine whether $t(v)$ follows.
- How likely are we to be correct?
  - If it does follow, we will always be correct.
  - If not, it depends on what the terms are ...

$$\frac{\text{Some}(A, C) \qquad B \subseteq C}{\text{Some}(A, B)}$$

$$\frac{\text{Some}(A, C) \qquad B \subseteq C}{\text{Some}(A, B)} \qquad\qquad \frac{\text{All}(C, B) \qquad C \subseteq A}{\text{All}(A, B)}$$

$$\vdots$$

$$\frac{\text{Some}(A, C) \qquad B \subseteq C}{\text{Some}(A, B)} \qquad\qquad \frac{\text{All}(C, B) \qquad C \subseteq A}{\text{All}(A, B)}$$

$$\vdots$$

How "dense" are the counterexamples to these inference patterns among "small" models?

How much model searching do we need to do, to ensure the probability of mistaking an invalid pattern as valid is below, say, 0.08?

How much model searching do we need to do, to ensure the probability of mistaking an invalid pattern as valid is below, say, 0.08?

- Some: 3 random models of size 3

How much model searching do we need to do, to ensure the probability of mistaking an invalid pattern as valid is below, say, 0.08?

- Some: 3 random models of size 3
- All: 4 random models of size 4

How much model searching do we need to do, to ensure the probability of mistaking an invalid pattern as valid is below, say, 0.08?

- Some: 3 random models of size 3
- All: 4 random models of size 4
- All*: 3 random models of size 3

How much model searching do we need to do, to ensure the probability of mistaking an invalid pattern as valid is below, say, 0.08?

- Some: 3 random models of size 3
- All: 4 random models of size 4
- All*: 3 random models of size 3
- Most: 7 random models of size 16

How much model searching do we need to do, to ensure the probability of mistaking an invalid pattern as valid is below, say, 0.08?

- Some: 3 random models of size 3
- All: 4 random models of size 4
- All*: 3 random models of size 3
- Most: 7 random models of size 16
- Most*: 3 random models of size 5

Why is it so hard to find small countermodels for inferences of the following form?

$$\frac{\text{Most}(C, B) \qquad A \subseteq C}{\text{Most}(A, B)}$$

Why is it so hard to find small countermodels for inferences of the following form?

$$\frac{\text{Most}(C, B) \qquad A \subseteq C}{\text{Most}(A, B)}$$

Because often enough, this is a plausible inference!

Why is it so hard to find small countermodels for inferences of the following form?

$$\frac{\text{Most}(C, B) \qquad A \subseteq C}{\text{Most}(A, B)}$$

Because often enough, this is a plausible inference!

(Cf. fact that NPIs appear in restrictor of *Most*.)

Moral of this quick exercise:

- For the classic parts of Monotonicity Calculus, random model checking can be very effective.

Moral of this quick exercise:

- For the classic parts of Monotonicity Calculus, random model checking can be very effective.

- This also allows going beyond logical validity, to assess (quickly and efficiently) merely plausible conclusions.

Moral of this quick exercise:

- For the classic parts of Monotonicity Calculus, random model checking can be very effective.
- This also allows going beyond logical validity, to assess (quickly and efficiently) merely plausible conclusions.
- Natural Logic, and especially Johan's own discussion of it, points the way toward potentially useful coarsenings for this purpose:

Moral of this quick exercise:

- For the classic parts of Monotonicity Calculus, random model checking can be very effective.
- This also allows going beyond logical validity, to assess (quickly and efficiently) merely plausible conclusions.
- Natural Logic, and especially Johan's own discussion of it, points the way toward potentially useful coarsenings for this purpose: perhaps these are the vehicles of inference.

Recall two quotations:

“a variety of interpretations, whose structure is itself a topic for semantic research”

“some ‘rougher semantics’, closer to mental models that we use . . . ”

Conclusion

- Natural Logic gives rise to fascinating questions of purely logical interest.

Conclusion

- Natural Logic gives rise to fascinating questions of purely logical interest.

- It has already been influential in the study of language and cognition.

Conclusion

- Natural Logic gives rise to fascinating questions of purely logical interest.
- It has already been influential in the study of language and cognition.
- I believe it holds yet further promise, as we think harder about how these vehicles of inference should be characterized, which in turn opens up new interfaces with cognitive science, e.g., *via* probabilistic computation.

Thanks for listening!

And thank you, Johan, for all the inspiring ideas (and much else besides)!